2809288901

UNIVERSITY OF LONDON THESIS

Degree P hD   Year 2007   Name of Author HULME
Ashley Thomas

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 It is an unpublished typescript and the copyright is held by the author.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.

LOAN
Theses may not be lent to individuals, but the University 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: The Theses Section, University of London Library, Senate House, Malet Street, London WC1E 7HU.

REPRODUCTION
University of London theses may not be reproduced without explicit written permission from the University of London 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 University 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 _____

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

Combined Experimental and Computational Studies of the Polymorphism of Small Organic Molecules

Ashley Thomas Hulme

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

University College London

December 2006

Department of Chemistry

University College London

20 Gordon Street

London

WC1H 0AJ

United Kingdom

UMI Number: U593336

UMI

Dissertation Publishing

ProQuest

I .   ... confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

for Hana

Abstract

Polymorphism is the ability of a molecule to adopt more than one crystalline form and the control of polymorphism is of importance to the fine chemical industry. Complimentary computational crystal structure prediction and experimental crystallisation techniques have been used to investigate the polymorphism of four organic molecules, none of which were previously known to be polymorphic. For each molecule computational crystal structure prediction produced possible crystal structures, which could correspond to new polymorphs. Manual crystallisation techniques were employed in three instances, and an automated crystallisation platform was used in the fourth, to discover new polymorphs. The crystal structures of all new polymorphs and solvates were fully determined, where possible, by single crystal X-ray diffraction for comparison to the predicted structures.

The 5-fluorouracil crystallisation screen produced one new polymorph which corresponded to a low energy predicted structure. For 5-fluorocytosine, where no anhydrous forms had previously been determined, two new polymorphs were discovered, one of which was predicted by the computational results. The study on 3-azabicyclo[3.3.1]nonane-2,4-dione aimed to find a new hydrogen bond dimer-based polymorph inspired by the results of earlier prediction studies. The crystallisation screen produced one new polymorph which was structurally related to the previously reported chain-based structure, along with a high temperature plastic phase. Four polymorphs of 4-hydroxycoumarin were discovered, of which two were fully characterised by single crystal X-ray diffraction and two were identified by powder X-ray diffraction. Many of the newly discovered solvates of these molecules had their hydrogen bonding rationalised in terms of the hydrogen bonded motifs found in the predicted structures of the parent molecule.

The viability of computationally predicting monohydrate structures was investigated, using 5-azauracil monohydrate as a test system. This proved a success, with the known crystal structure found by the computational method to be energetically competitive with the other hypothetical structures.

Table of contents

Figures

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Schemes

Tables

Chapter 5

Chapter 6

Chapter 7

Abbreviations and definitions

5FU	5-Fluorouracil
5FC	5-Fluorocytosine
API	Active Pharmaceutical Ingredient
BQT	3-Azabicyclo[3.3.1]nonane-2,4-dione
CCDC	Cambridge Crystallographic Database Centre
CSD	Cambridge Structural Database
CSP	Crystal Structure Prediction
CSP1999	CCDC international blind test of structure prediction (1999)
CSP2001	CCDC international blind test of structure prediction (2001)
CSP2004	CCDC international blind test of structure prediction (2004)
Drug Substance	*see* API
Drug Product	API formulated into the drug delivery vehicle (tablet, suspension etc.)
DSC	Differential Scanning Calorimetry
DMA	Distributed Multipole Analysis
ESP	Electrostatic potential
EV	Solvent evaporation crystallisation method
FDA	Food and Drug Administration – oversees pharmaceutical regulation in the United States
Parent	The neutral API (free acid or base) from which a salt is derived
Salt	Reaction of the parent API in an acid/base reaction to become a charged species, with a constituent counter-ion from the reaction
STA	Simultaneous Thermal Analysis, usually a combination of DSC and TGA
SXRD	Single crystal X-Ray Diffraction
TGA	Thermogravimetric Analysis
VD	Vapour diffusion crystallisation method
XRPD	X-Ray Powder Diffraction

Chapter 1 Introduction

1.1 General introduction to polymorphism and its study

The origin of the word polymorphism derives from the Greek, poly = many and morph = forms. Its most common usages in science are predominantly biological in context, such as 'genetic polymorphism' which refers to minor differences in the same gene (alleles) that can manifest themselves in physical differences such as eye colour, hair type or diseases. The use of the word polymorph in the organic solid state is analogous – with the molecule of interest the 'gene', the different crystalline forms it can adopt the 'alleles' and the differences between polymorphs manifesting themselves as different physical and chemical properties.

Polymorphism can have important implications for the development of new fine chemical products such as pharmaceuticals, pigments and dyes, high energy materials and food ingredients.[1] The arrangement of the molecules in a crystal structure determines its physical properties,[2] and different polymorphs can exhibit varying physical, chemical and mechanical properties such as solubility and bioavailability; heat capacity and enthalpy of melting; melting and sublimation temperatures; rates of solid state reactions and chemical stability.[3] Lack of control of the identity and quantity of polymorphs can potentially lead to inconsistency in the performance of the product. This could include drug substances failing to achieve the desired therapeutic activity, unreliable explosion characteristics for high energy materials, inconsistency in the colour of a dye or pigment compound or inferior oraleptic characteristics in chocolate due to polymorphism of a substituent.[1] To obtain reliable results, the polymorphism of the compound must be controlled, and to achieve this, full knowledge of the solid forms of the compound and their inter-relationships is required.

Polymorphism is only one aspect of the full range of solid forms in which an organic molecule can potentially participate (scheme 1.1) and the definition of each of the terms shown in scheme 1.1 will provide a framework for the study of the organic solid state.

Parent Molecule

Molecular Adducts | Chiral | Amorphous

Solvates (inc. Hydrates) | Co-crystals | Salts | Racemate | Optically Resolved | Polyamorphic Forms

Desolvated Solvates

Polymorphs

Conformational | Packing
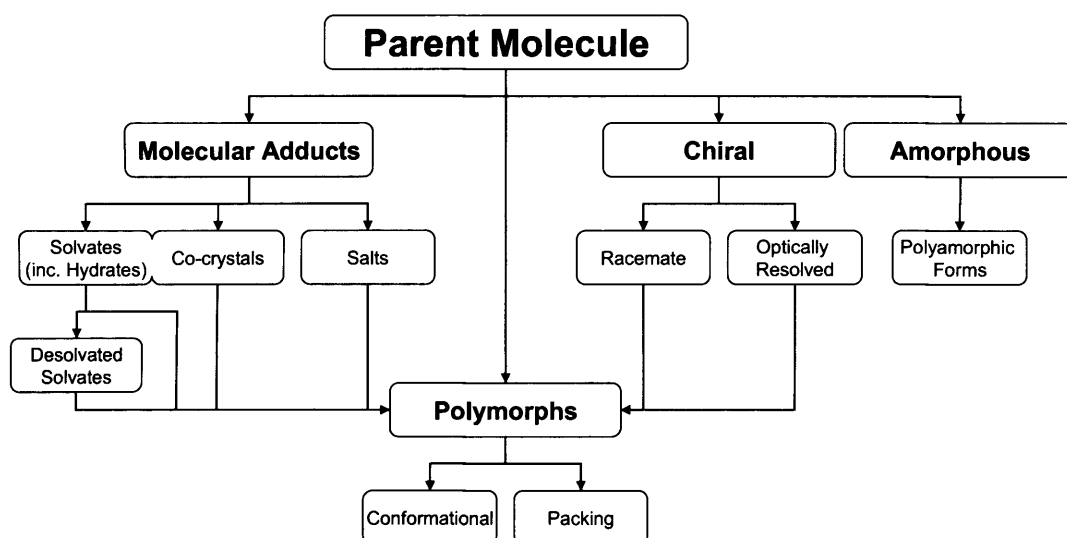
Scheme 1.1: Possible solid state forms in which the organic molecule of interest, the parent molecule, can potentially participate. Not all combinations are illustrated

1.2 Definitions

1.2.1 Polymorphism

Organic solid state definitions of the term polymorphism in the scientific literature are dependent on the degree of inclusiveness that each author desired. The most permissive definition of polymorphism includes all of the solid state forms in which the neutral parent molecule is present, including non-crystalline forms (amorphous forms) and those in which the parent molecule forms a crystalline molecular adduct with another molecular species (solvates and co-crystals).[4;5] The least permissive definition of polymorphism is that the only difference between polymorphs is that they exhibit different crystal structures due to differences in the molecular packing arrangement, which can include changes to the conformation of the molecules therein.[6] Implicit within this definition are the similarities between polymorphs: polymorphs are crystalline and have the same chemical composition, while the molecules comprising the crystal exhibit the same structural- and stereo- isomerism. McCrone[7] has provided a sound working definition, that has stood the test of time: "two polymorphs will be different in crystal structure but identical in the liquid and vapour states". This definition succinctly excludes amorphous forms and molecular adducts. For the purposes of this work, the strictest definition above will be employed:

The only difference between polymorphs is that they exhibit different crystal structures due to differences in the molecular packing arrangement, which can include changes to the conformation of the molecules therein.

Polymorphism in elements, such as that of the diamond, graphite and fullerene forms of carbon is usually termed allotropy. The only sub-categorisation of polymorphism is with regard to molecular conformational flexibility. Rigid molecules, without the capability of changing conformation, can only differ in the relative orientation, or packing, of the molecules to give alternative crystal structures, giving rise to the term 'packing polymorphism'. Conformational polymorphs, in addition to packing differences, will have significant differences between the conformations of the constituent molecules in the polymorphs, which requires the molecule to have at least one torsion angle around which it is relatively free to rotate. Thus where there are significant molecular conformational differences, the polymorphs are classed as conformational polymorphs. The differentiation between these terms is necessarily artificial, as even for polymorphs of rigid molecules the conformations will differ slightly.[8]

1.2.2 Amorphous forms

Amorphous forms differ from crystalline forms by the degree of order inherent in the form. While crystalline forms have short range interactions and long range order, amorphous forms have only the short range interactions with a disordered arrangement of the molecules,[9] and can be envisaged as supercooled liquids. There have been suggestions of different modifications of the amorphous state for some compounds, though the occurrence is thought to be rare, leading to the designation 'polyamorphism' – a direct analogy to polymorphism in the crystalline state.[10] For amorphous forms there exists a temperature at which the amorphous material converts from a glass-like state to a more rubber-like state, because of increased molecular mobility above this temperature, known as the glass transition temperature (T_g).

1.2.3 Molecular adducts – solvates, co-crystals and salts

Datta and Grant have defined a crystal as a molecular adduct "… when its lattice consists of more than one chemical component",[6] and it is used here as a general term to encompass solvates, co-crystals and salts.

Solvates are crystal structures in which both the parent molecule and the solvent of crystallisation are present and have been sub-classified according to the role of the solvent in the crystal structure.[11] In isolated lattice site solvates, each solvent molecule does not have close contacts with any other solvent molecules – the solvent molecules are separated by molecules of the parent species. If there is no strong interaction between the solvent and the non-solvent molecules in the crystal, then these solvates can be referred to as clathrates, and the solvent can be viewed as filling voids in the lattice. In channel solvates the solvent is contained within channels in the crystal structure. Desolvation can occur where the channels meet the surface of the crystal. The loss of solvent from the end of a channel causes a thermodynamic gradient along the channel causing solvent migration and further desolvation. In some channel solvates the solvent content in the channels can vary with the vapour pressure of the solvent, and the solvent can be included in non-stoichiometric quantities. The third solvate sub-category contains ion associated solvates. These solvates can occur for salts, and the coordination can be very strong, leading to difficulties in desolvating such solvates.

Solvate crystals once removed from the mother liquor of the crystallisation are often unstable with respect to desolvation, or alternatively the desolvation can be effected by exposing the solvate crystals to elevated temperature. There are three possible products of desolvation: the now unsolvated lattice can collapse to a previously known unsolvated crystalline form, to an amorphous state, or the unsolvated compound can retain the crystal structure of the original solvate – a desolvated solvate.[12;13] Such phases are of lower density than conventional polymorphs and tend to be more disordered.[14] Lattice collapses associated with desolvation can be advantageous in pharmaceutical development in a particle engineering context because it can give micrometer-size particles with a small size distribution.

In scheme 1.1, hydrates have been deliberately included as a subset of solvates, because while there is nothing inherently special about water as a solvent in solvates, the prevalence of hydrates[15] requires their explicit inclusion, and indeed the importance

of hydrates is highlighted by the work presented in chapter 7 which contains some of the first crystal structure predictions for a monohydrate.

Co-crystals and solvates belong to a broad continuum, in which the unionised parent compound is present in the crystal, along with one or more other molecular components. The difference between solvates and co-crystals has been defined by Almarsson and Zaworotko: "The primary difference is the physical state of the isolated pure components: if one component is liquid at room temperature, then the crystals are referred to as solvates; if both components are solids at room temperature, the products are referred to as co-crystals"[16] This definition is not universally accepted, with Dunitz defining 'co-crystal' as a term that encompasses solvates as a sub-category: "[a co-crystal is] ... a crystal containing two or more components together. Thus *co-crystal* encompasses molecular compounds, molecular complexes, solvates, inclusion compounds, channel compounds, clathrates ..."[17]. While the Almarsson and Zaworotko delineation is artificial, in most circumstances it provides a clear distinction and will be adhered to in this work.

Salts are formed when the parent compound containing acid or base functional groups is ionised in an acid-base reaction and crystallises with the counter-ion from the acid-base reaction. The chemical composition of the parent molecule is now changed, and because of this and the charged nature of the molecular components of the resultant crystal, these crystalline products are salts rather than co-crystals.

1.2.4 Chiral forms

For a chiral compound in solution, comprised of R and S enantiomers, two crystallisation outcomes are possible. Both enantiomers can crystallise together to yield one crystalline form containing both R and S enantiomers, to give a racemic crystalline product. Alternatively each enantiomer can crystallise separately, giving crystals of pure R and crystals of pure S, the process of optical resolution. In each of these three forms, the racemate, pure R and pure S, the steroisomerism is different and this inequality between their constituent molecules precludes them from being designated polymorphs.

Chiral molecules which exhibit dynamic isomerism and rapid equilibration in the liquid state have been used as an argument, in conjunction with McCrone's definition of polymorphism that the two different crystal forms become equivalent in

the liquid state, to assign crystals of the same molecule but with different chirality (pure R, pure S and RS) as polymorphs.[18-20] At the same time it has been acknowledged that for such systems where the interconversion is slow, then such crystal forms do not qualify as polymorphically related. The two optically resolved crystal forms and the racemate, even when they show rapid equilibration upon melting or dissolution, are not polymorphs in the strictest sense because of the molecular stereo-isomeric inequivalence in the crystalline state. For such crystal structures the qualification 'chiral polymorph' can be used to connote the same molecular connectivity but stereo-isomeric inequivalence.

1.2.5 Phenomena related to polymorphism

The terms defined in sections 1.2.1–1.2.4 provides a framework within which to classify the different solid state forms of a compound, however there are related phenomena that require further interpretation. Highlighted below are some phenomena whose classification is ambiguous using the conceptual framework outline above.

Molecules that can tautomerise can potentially crystallise into different crystal structures containing the different tautomeric forms.[19] Strictly tautomerism is a form of structural isomerism, and crystal structures containing different isomers are not polymorphs. Similar confusion can surround molecules that can crystallise in either neutral or zwitterionic forms. In both of these cases the isomerism is caused by a facile proton transfer and, according to McCrone's definition of polymorphism, such forms would be identical in the liquid state soon after melting or dissolution due to rapid equilibration. So while neither of these phenomena can be assigned as true polymorphism they can be defined with qualification as tautomeric polymorphism and zwitterionic polymorphism respectively.

Polytypism is a feature of some layer structures, in which the repeat stacking of the layers is not well defined. In a set of polytypes the constitution of the fundamental layer will be identical but the individual polytypes will exhibit different stacking of the layers. In such cases the two crystallographic axes that lie in the plane of the layer are well defined and constant between the different polytypes, while the third axis length will vary depending on the stacking present in each particular polytype.

The term 'pseudopolymorph' is a historical term that has been used for various phenomena similar to polymorphism, usually, but not exclusively, referring to solvates.[19] Both Seddon[21] and Bernstein[22] have argued against the use of this term because there is ambiguity in both the use of the prefix pseudo- and the precise definition of the term pseudopolymorph. The term adds little to the language of the organic solid state that is not covered more precisely by other terms, and for this reason it will be avoided.

1.3 The prevalence of polymorphism

The prevalence of polymorphism in the organic solid state is a factor in determining its potential importance to the fine chemicals industry. McCrone stated that "every compound has different polymorphic forms, and that, in general, the number of forms known for a particular compound is proportional to the time and money spent in research on that compound",[7] suggesting that the more a molecule is investigated for polymorphism, the more instances of polymorphism will be found.

An accurate quantification of the proportion of organic molecules that exhibit polymorphism based on existing data is difficult because, to arrive at an accurate statistic, only the relatively small set of systems for which polymorphism has been thoroughly investigated can be considered. An analysis of the Cambridge Structural Database[23] (CSD) for polymorphs in 1999 yielded only 321 polymorphic systems in the subset of organic molecules.[24] Of this 321, 291 were dimorphic, 27 were trimorphic and three had four polymorphs. By 2004, the prevalence of multi-polymorphic instances had increased to 50 trimorphic systems, 5 tetramorphic, one pentamorphic and one hexamorphic system.[25] Such numbers are a very small percentage of the structures present in the CSD because the vast majority of crystal structures present in the CSD were determined only to elucidate the structure of the molecule in the crystal – in such cases the first crystal structure obtained was published, with no further investigation of the solid state diversity. For such molecules upon which no solid state investigation has been done, no comment can be made about their possible polymorphism. The Innsbruck School of Pharmacognoscy have investigated the polymorphism of a large number of organic molecules over the past 50 years using hot stage microscopy as their principal characterisation technique and conclude that approximately one third of organic

molecules can form polymorphs, with a further third capable of forming solvates.[26] If this figure is approximately accurate, the potential for polymorphism (and solvate formation) for any individual organic molecule is significant and will warrant investigation if it is to be used as a fine chemical.

1.4 Polymorphism in the pharmaceutical industry

The pharmaceutical industry in particular focuses significant resources in discovering and controlling the solid state forms of newly developed active pharmaceutical ingredients (APIs). It is critical to discover all solid state forms to select the form for drug product development that has optimal physical and chemical characteristics, and to specify that single form to the regulatory authorities who will only licence that one form for use in the final drug product, and to maximise intellectual property protection. In this context the solid state diversity of two widely used pharmaceuticals, 5-fluorouracil and 5-fluorocytosine, is investigated (chapters 3 & 4).

1.4.1 Optimisation of the solid state form

The optimal form of an API is defined in terms of a small range of criteria, fulfilment of which are pre-requisite before the form of the API can be considered for use in a drug product. These criteria include how the physical properties of each of the available forms impacts the physical and chemical stability of the API, its solubility, bioavailability and manufacturability.[14] The chosen form must be suitably physically and chemically stable, have sufficient solubility/bioavailability to be effective, have favourable physical properties such as flowability, compactability/tabletting, morphology, hygroscopicity[27] and if there is the possibility of phase transitions under manufacturing conditions they should be well understood and easily controlled.[28] Potentially the optimal form could be a polymorph (either stable or metastable), an amorphous form, a hydrate, or a salt of the parent API. Choice of the optimal form for development ideally happens early in the development of the API prior to the first clinical trials, and a change of form later in the development process will prove costly (in terms of both time and money).

Ideally the thermodynamically most stable polymorphic form of the API is chosen for development which aids the formulation of a robust method of manufacture and imparts confidence in the manufacturing specification submitted to regulatory

bodies such as the Food and Drug Administration (FDA). The choice of the thermodynamically most stable form minimises the possibility of unplanned phase changes due to variations in the manufacturing process or during storage of the drug product. In the case of highly soluble APIs, often the variation in solubility between different polymorphic forms is not sufficient to affect the bioavailability of the API and a less stringent specification is allowed that may lead to any one of the known polymorphs[29] (though even in this scenario it is best practice to specify a single polymorph). It is estimated that the variability in solubility between polymorphs is commonly less than a factor of 10.[30] In the most usual case, where the polymorphic form of the API critically determines the efficacy of the drug product (through either solubility or stability considerations), it is required that only a single form is manufactured and included in the drug product, and that controls are developed and put in place to monitor amounts of other polymorphic forms.[4;14] Commonly a range of experiments, including slurrying and thermal cycling experiments, is carried out on what is thought to be the most stable form to ensure that there is not a more stable form possible.

When the stable form does not have the required physical properties, especially low solubility and thus poor bioavailability, alternative solid state forms can be considered. Both metastable crystalline forms and the amorphous form are unstable with respect to the thermodynamically most stable form which leads to them having greater solubility, though there is the danger that a metastable form could convert to a more stable form or that an amorphous formulation could crystallise. Use of either as the selected form of an API in the drug product requires proof that it is suitably kinetically stable, and detailed consideration of the manufacturing specification is required including monitoring that the chosen form is being produced reliably and monitoring for the possible production of other forms. The choice of either a metastable polymorph or amorphous form that subsequently underwent an unexpected phase change during either drug manufacture or drug product storage would infer that there was inadequate control of the manufacturing process and would lead to the withdrawal of approval for the drug product. Examples of the use of an amorphous form include novobiocin, where the crystalline form is not readily absorbed, but the amorphous form is and is therefore therapeutically active, and cefuroxime axetil (Ceftin®), which is

formulated as a fully amorphous form, but which is very unstable to the presence of the crystalline form.[31] In solid suspensions of novobiocin additives are required to suppress the inherent tendency for the amorphous form to crystallise,[32] and some formulations of cefuroxime axetil are wax coated to prevent crystallisation.[31]

A thorough knowledge of the polymorph and solvate profile of the API is essential, even in cases where the most stable form is chosen for development. The manufacturing process must be fully controlled at each stage with respect to the solid form of the API. Processes used in pharmaceutical manufacture such as temperature variation, relative humidity changes, milling, wet granulation and drying processes can cause phase conversions within or between any of the following categories: polymorphs, solvates, desolvated solvates and amorphous forms.[8] Environmental conditions under which either the bulk API or the drug product may be stored must be considered because of their potential impact on the solid form of the API. For example, the hygroscopicity of the chosen form of the API must be well understood – if it has the potential to absorb water and convert to a hydrate, this may not cause a problem in low humidity areas, but production at, or shipping to, more humid regions may cause a phase transition to the hydrate. A form that undergoes a hydration phase change below 60% relative humidity will require all water to be excluded from manufacture and upon storage of the final drug product.[14]

A salt version of an API is often used in circumstances where it has desirable physical properties that the parent API does not possess. For a selected salt form such properties can include greater solubility, and hence bioavailability or circumstances where the salt may be crystalline while the parent API can only be produced as an amorphous form, allowing structural characterisation by single crystal X-ray diffraction. The parent API may have a complex polymorphic profile that is difficult to control, whereas the salt may only have one form or the parent API may be hygroscopic whereas the salt may not. In a similar analogy to optimisation of the polymorphic form, a salt screen would be carried out to determine the optimal salt form based on its physical solid state properties. It should be noted that because each salt produced is chemically distinct from the parent API, each would have its own polymorph and solvate profile and if a salt was chosen for development, a polymorph and solvate screen would have to be performed upon that version.[33]

The classic example of a drug substance for which the solid state characterisation was crucially incomplete is the anti-HIV drug ritonavir (Norvir ©) developed and marketed by Abbott Laboratories. The form chosen (Form I) for development was assumed to be the thermodynamically most stable form and the drug product was formulated as both a liquid and as a semi-solid capsule. Almost two years after the drug was released to market, the semi-solid formulation started to fail the quality control dissolution test, with a large proportion of the drug substance unexpectedly precipitating out. The semi-solid capsule contained a near saturated solution of Form I, but this was super-saturated with respect to a new, more stable form, Form II. This new polymorph was the cause of the failure of the dissolution tests and resulted in withdrawal of the drug from the market. It was viewed as desirable to reformulate Form I, even though metastable with respect to Form II and that potential for seeding by Form I will always be a risk, as it had more favourable manufacturing properties and indeed the reformulated product released, after a delay of 18 months, contained Form I.[34]

1.4.2 Intellectual property protection

A fundamental reason to thoroughly research the polymorphism of pharmaceutical ingredients is for robust intellectual property protection. A patent granted on an invention bestows upon the holder of the patent a limited term of exclusivity to exploit the invention, in return for its public disclosure and the freedom for anyone to employ the invention upon expiration of the patent. The substantial expenditure on research and design, of both the products brought to market and those that fail the development process and never get to market, is recovered by the premium that can be charged for the marketed products due to the exclusivity afforded by the patent for the lifetime of the patent. To ensure recovery of the initial investment, the patent protection must cover all eventualities. Alternative newly discovered polymorphs to the form marketed or those covered by the company's original patent could be used by a rival company to circumvent the original patent, and produce the drug in the new polymorphic form. For a pharmaceutical, the rival company would only have to establish bioequivalence of their new form with the original form before submitting an Abbreviated New Drug Application (ANDA) which would draw on the clinical trial data contained in the original company's New Drug Application (NDA) for proof of safety and efficacy

against the disease in question.[35] The ANDA would be considerably less expensive for the rival company compared to the cost of the original NDA, as it does not have to take on the expense of the clinical trials, or the risk that the drug may fail in the clinical trial stage. A key example is the case of the GlaxoSmithKline drug ranitidine hydrochloride (Zantac ®), one of the first attempts to exploit polymorphism to extend the patent protection on a blockbuster pharmaceutical. During the pre-clinical development process, two forms were discovered and patented, though Form 1 was prepared and patented approximately three years prior to Form 2. Form 2 was the form used in manufacture of the drug product. At the expiration of the patent on Form 1 Novopharm attempted to manufacture a generic version of Form 1 of ranitidine hydrochloride. Manufacture according to the original Form 1 patent only yielded Form 2, and Novopharm argued that Form 1 had never been produced using the patented route and that this invalidated the Form 2 patent, which would have allowed it to produce either form without restriction three years prior to the expiry of the Form 2 patent. GlaxoSmithKline successfully defended the validity of the Form 2 patent, though their attempts to stop Novopharm and others from marketing Form 1 with small amounts of Form 2 (present as an impurity) by arguing that it violated their Form 2 patent were ultimately unsuccessful.[1]

1.5 Thermodynamics of polymorphism

At a given temperature and pressure usually only one polymorph is stable, and all other polymorphs are metastable. Metastable forms can be short lived or kinetically stabilised and longer lived. The relative stability of two polymorphs is determined by the free energy difference:

$$\Delta G = \Delta H - T\Delta S$$

The enthalpy difference, ΔH, is the lattice energy difference between the two forms, which occurs because the intermolecular interactions are different in the two polymorphs. The entropy difference, ΔS, is due to differences in the lattice vibrations and any disorder present in the polymorphs.[5] The enthalpy of a crystal form at constant pressure is a sum of the internal energy, U, due to all of the intermolecular interactions in the crystal plus a PV term, with P usually the atmospheric pressure:[3]

$$H = U + PV$$

The difference in the PV term between two polymorphs, $P\Delta V$, is so small that it is negligible in the calculation of the free energy difference. Over the full range of temperature in which a crystalline form can exist, the enthalpy of the crystal increases with rising temperature due to increasing molar volume and weakening of the intermolecular interactions. The entropy, initially zero at 0 K, increases due to the increasing population of higher energy levels of lattice vibrations. The increase in the $T.S$ term outweighs the increase in the enthalpy term, leading to a net decrease in the free energy with increasing temperature (figure 1.1).
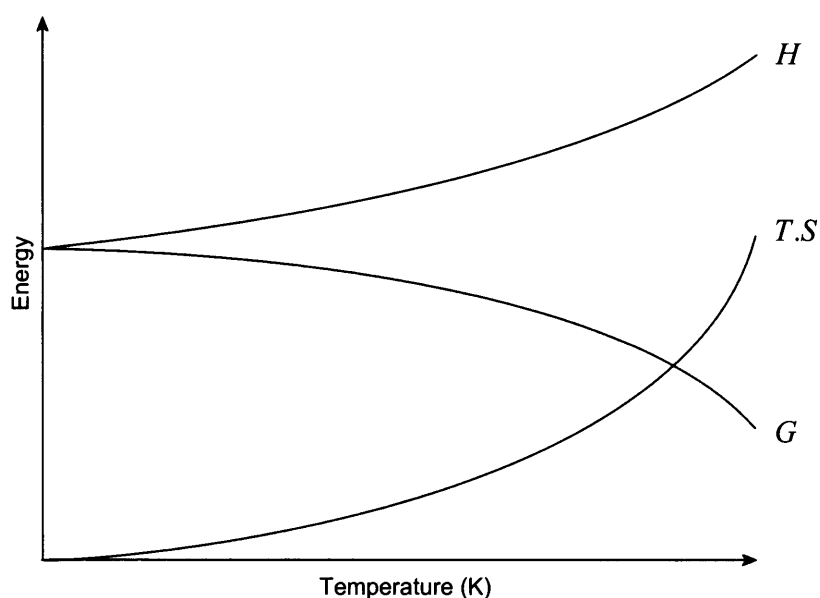
Figure 1.1: Variation in enthalpy, entropy and free energy with temperature for a crystalline phase. At 0 K the entropy $S_0 = 0$, so $G_0 = H_0$

A pair of polymorphs can be thermodynamically related in one of two ways. One polymorph can be metastable with respect to the other at all temperatures below their melting points, in which case the relationship is monotropic. If each of the polymorphs has a temperature range in which it is stable with respect to the other, and consequently there is a transition temperature at which the order of stability crosses, the two forms are said to be enantiotropically related. In figure 1.2 polymorphs A and B are monotropically related with A more stable than B. In figure 1.3 A and B are enantiotropically related with A more stable than B at lower temperatures. Included in both figures are the free energy and enthalpy curves for the liquid.
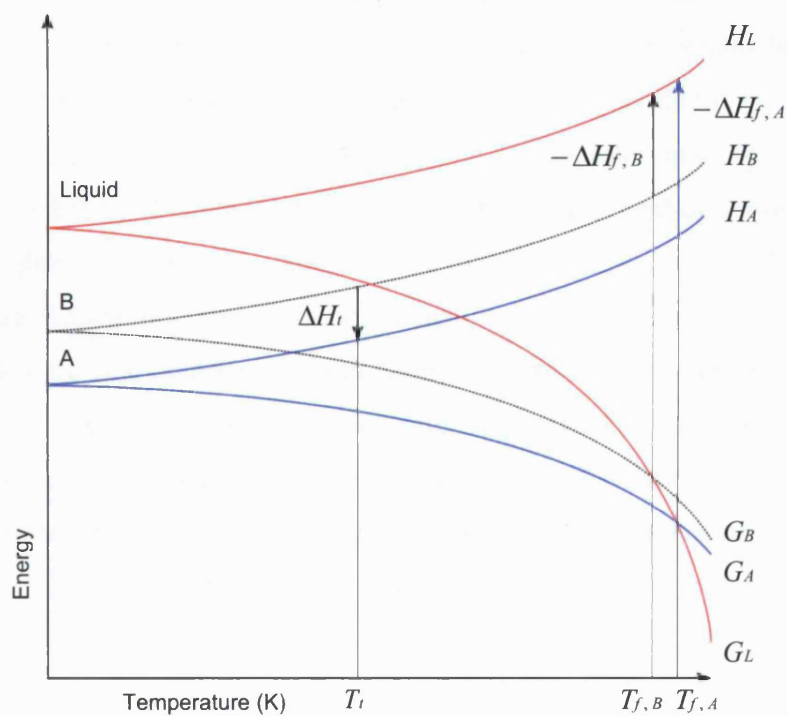
Figure 1.2: Free energy relationship between monotropically related polymorphs
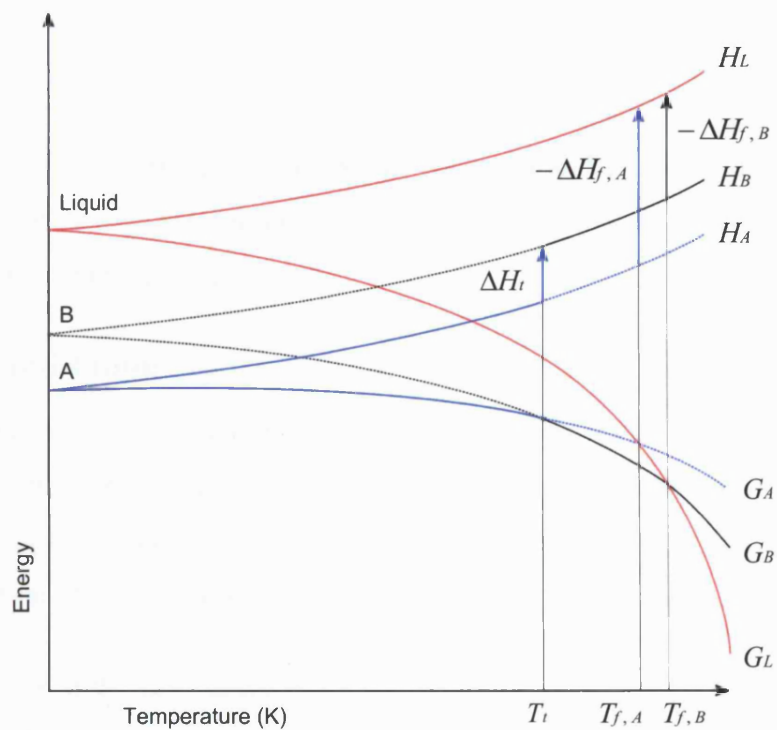
Figure 1.3: Free energy relationship between enantiotropically related polymorphs

These diagrams have provided the basis for two thermodynamic rules – the Burger and Ramberger rules[36] – used to determine the thermodynamic relationship between two polymorphs from thermal data. The heat of transition rule states that if the enthalpy of transition (ΔH_t) between the two polymorphs is endothermic then they are enantiotropes, and if exothermic the two are monotropes. The heat of fusion rule states that if the higher melting polymorph has the higher heat of fusion then the two polymorphs are monotropes, and if the higher melting polymorph has the lower heat of fusion then they are enantiotropes. In both cases it is assumed that the appropriate data can be collected: in the case of the transition rule the phase transition must be observed and for the fusion rule the phase transition must not occur. The thermodynamic stability difference between polymorphs is the fundamental basis for many of the physical property differences, such as the exhibition of different saturated vapour pressures and different solubilities in a given solvent:

$$\Delta G \approx RT \ln\left(\frac{p_2}{p_1}\right)$$

$$\Delta G \approx RT \ln\left(\frac{s_2}{s_1}\right)$$

where p_1 and p_2 are the saturated vapour pressures and s_1 and s_2 are the equilibrium solubilities of the two polymorphs.[3] The more stable polymorph has the lower solubility and saturated vapour pressure.

1.6 Crystallisation

The process of crystallisation is comprised of two steps – a nucleation step followed by crystal growth. While thermodynamics defines the stability domains for different polymorphs, kinetics controls the crystallisation pathway and consequently can lead to the formation of metastable crystal structures.[37]

1.6.1 Nucleation

Gavezzotti[38] has postulated the steps that may constitute crystal growth from solution:

- A. When super-saturation is reached the solute de-mixes in the form of minute droplets with liquid structure

- B. As super-saturation increases, the size of the droplets increases, up to a point where the transition to a semi-ordered structure starts; there is a long period during which the droplet structure is fluxional

- C. Due to as yet largely unknown reasons (which might be related with kinetic factors such as cracks in the container walls or impurities, rather than thermodynamics) there are transitions to a pre-crystalline structure, which then starts its way to growth

- D. In solution steps A-C are largely influenced by interaction with the solvent

- E. Growth of different crystalline nuclei proceeds at different rates, in a competition regime, and the appearance of a given crystal structure depends on the stability of the structure itself, but also very much on the growth speed

Each of the steps A-E will be addressed in more detail below to outline the theory of nucleation.

A. When super-saturation is reached the solute de-mixes in the form of minute droplets with liquid structure

A saturated solution is one in which the solution state is in equilibrium with the solute in a specified solid state. The specified solid state can be the thermodynamically most stable polymorph or a metastable polymorph, in which case the equilibrium is a metastable equilibrium.[2] The condition of super-saturation is required for nucleation to commence, and this is commonly achieved by either solvent evaporation or cooling of a solution saturated at elevated temperature. The super-saturation, σ, is defined as the difference in chemical potential between a molecule in the super-saturated state and the equilibrium state, but is commonly expressed in terms of activities (a). Where solubility

(x) can approximate for the activity, this substitution is also made. For small values of $\left(x_{ss} - x_{eq}\right)$ the final approximation can be made below[39]

$$\sigma = \ln\left(\frac{a_{ss}}{a_{eq}}\right) = \ln\left(\frac{x_{ss}}{x_{eq}}\right) \approx \frac{\left(x_{ss} - x_{eq}\right)}{x_{eq}}$$

A solution that is saturated can commonly be cooled into the super-saturated region before crystallisation occurs. The difference between the onset of super-saturation and crystallisation is known as the metastable zone (figure 1.4) and its presence is related to the activation energy required for nucleation. Within this zone if seeds of a polymorph are introduced, the super-saturation is relieved and the product will be the same polymorph as that of the seeds.
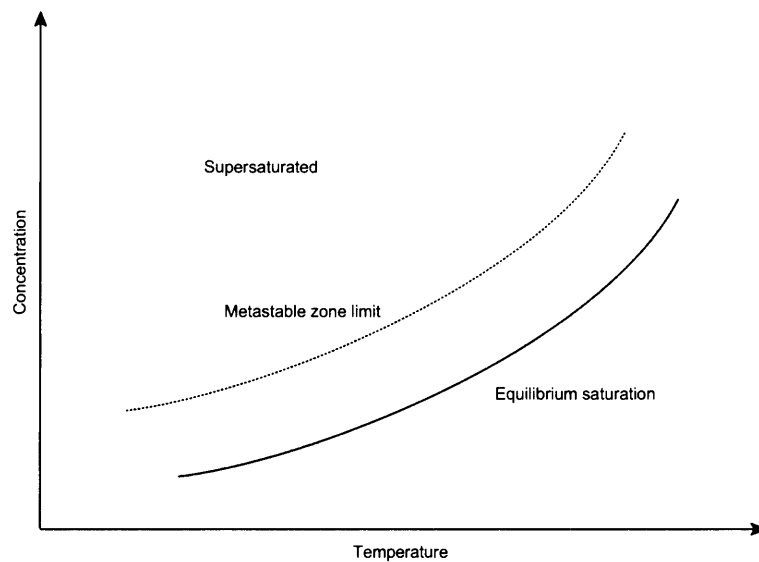
Figure 1.4: Metastable zone width. The metastable zone limit defines the super-saturation limit beyond which spontaneous nucleation can occur

B. As super-saturation increases, the size of the droplets increases, up to a point where the transition to a semi-ordered structure starts; there is a long period during which the droplet structure is fluxional. C. Due to as yet largely unknown reasons (which might be related with kinetic factors like cracks in the container walls or impurities, rather than thermodynamics) there are transitions to a pre-crystalline structure, which then starts its way to growth

The super-saturation provides the driving force for crystallisation, and the available energy, upon which it acts, must overcome the interfacial tension, that is the energy required to form the new solid-solution interface:[39;40]

$$\Delta G = -zkT\ln\left(\frac{x_{ss}}{x_{eq}}\right) + \beta\gamma z^{2/3}$$

where z is the number of molecules in the nucleus, β is a nucleus shape factor and γ is the interfacial tension. Plotting the free energy difference for different super-saturations, shows that the energy barrier is lower at higher super-saturations and that the nucleus size (z_c – the critical nucleus size) required to overcome the barrier is smaller (figure 1.5). When the available energy can overcome this energy barrier, nucleation occurs.
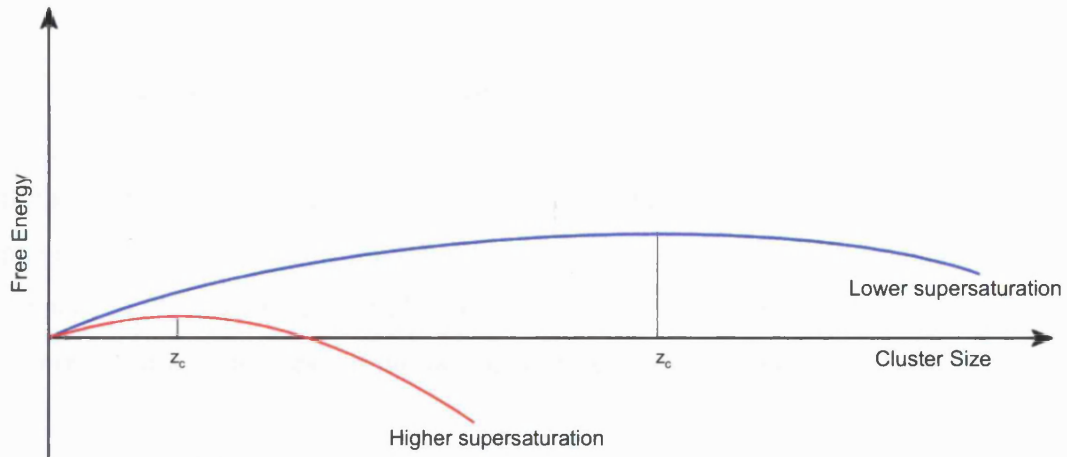
Figure 1.5: Barrier to nucleation at different super-saturations. At higher super-saturation the energy barrier to nucleation is lower and the critical nucleus size is smaller[39]

The above consideration of nucleation assumes that the nucleation is spontaneous from solution, known as primary homogeneous nucleation. Primary heterogeneous nucleation is where the nucleation event is induced by foreign particles, such as dust or insoluble impurities in the solution. Secondary nucleation occurs when the nucleation event is induced by crystals of the crystallising solute that are either unintentionally present in the crystallisation vessel, or have been intentionally added to ensure a predictable crystallisation outcome, known as seeding.[41]

D. In solution steps A-C are largely influenced by interaction with the solvent

The solvent used for the crystallisation is of critical importance in determining the outcome of the crystallisation. During the pre-nucleus stage self-assembled units are thought to exist in solution, usually associated through strong intermolecular interactions such as hydrogen bonds, and it is these units that aggregate to form the critical nuclei and the final macroscopic crystalline product.[40] The solute when dissolved in solvents with different properties such as polarity or proton-donating ability, may form different self-assembled units which can lead to different polymorph or solvate outcomes.[42] For this reason solvent is the primary variable parameter exploited in crystallisation screens. The nucleation process can hence be envisaged, from a supramolecular viewpoint, as including a step involving self-assembled units in solution prior to nucleation:[43]

Molecules in solution ↔ Self-assembled units in solution ↔ Crystallisation

It should be noted, however, that solvent is not the only variable that can change the polymorphic outcome of a crystallisation; other factors to consider include, but are not limited to, temperature, initial concentration of solution, seeding (intentional or unintentional), impurities in solution, agitation and cooling rate.[44]

E. Growth of different crystalline nuclei proceeds at different rates, in a competition regime, and the appearance of a given crystal structure depends on the stability of the structure itself, but also very much on the growth speed

At the pre-critical nucleus stage nuclei of several different polymorphic forms may be present in the crystallisation. These nuclei are all at the fluxional stage where dissolution is more energetically favoured than growth, and as the crystallisation proceeds the continued growth of one of the forms present is favoured over all others, while the others re-dissolve and their constituent molecules attach to the nuclei of the favoured form.[3] The favoured form will be the one that grows most quickly to its unique critical nucleus size and its further growth is then energetically advantageous (figure 1.5). This form will be observed at the expense of all others, even those that could result in more stable crystal structures – the favoured form may be the thermodynamically

most stable form or a metastable form. From this it can be appreciated that the process of nucleation is not thermodynamically controlled, and that kinetic factors influence the crystallisation outcome.[45] A crystallisation pathway that leads to a metastable polymorph may have a lower energy barrier to nucleation compared to the thermodynamic crystallisation pathway and the kinetic drive to relieve the imposed super-saturation of the system dominates the thermodynamic drive to attain the lowest point of free energy (figure 1.6).[46] The initial formation of a metastable form will only yield that form as the final crystallisation product if the crystallisation process is completely kinetically controlled and not allowed to enter a regime where thermodynamic control can take over: if the metastable form is left in contact with its mother liquor for a period of time after the kinetic driving force has been extinguished, there is a thermodynamic driving force for a reconstructive phase transition to the thermodynamically more stable form.
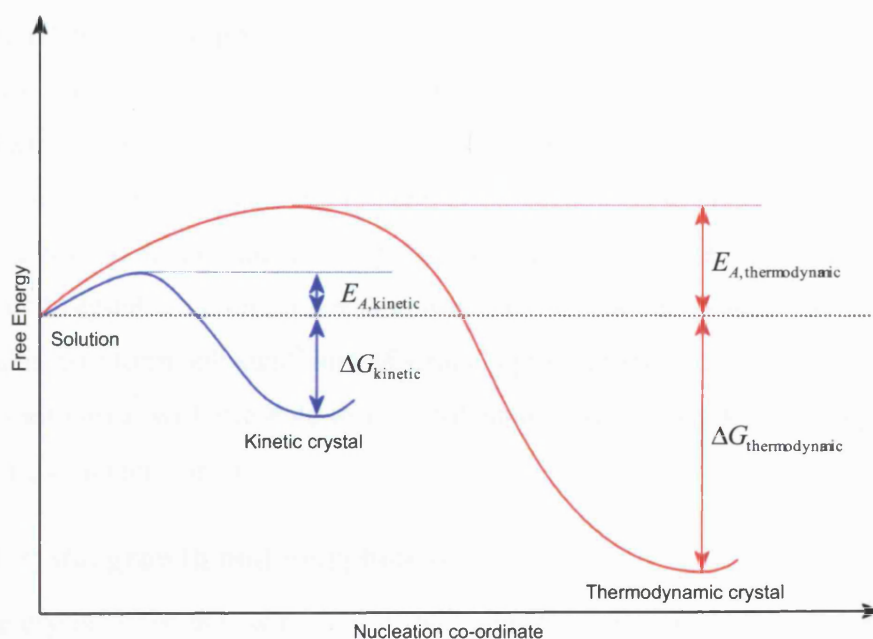
Figure 1.6: Kinetic *versus* thermodynamic control of crystallisation. The kinetic pathway has a lower nucleation activation energy ($E_{A,\text{kinetic}}$), but leads to the less stable crystal structure ($\Delta G_{\text{kinetic}}$). The thermodynamic pathway has a higher barrier to nucleation ($E_{A,\text{thermodynamic}}$) but leads to the more stable crystal structure ($\Delta G_{\text{thermodynamic}}$) [45]

1.6.2 Nucleation of solvates

The crystallisation of a solvate rather than a non-solvated form can be seen as an 'interruption' to the above model of nucleation. Self-assembled units that include solvent can form in solution, co-ordinated to solute molecules *via* multi-point hydrogen bond contacts.[47] The energy barrier for removal of the strongly co-ordinated solvent to enable the solute species to crystallise into a non-solvated crystal structure, the demixing step, can be high enough that the alternative crystallisation pathway where the solvent is incorporated into the crystal structure is followed instead. The enthalpic gain from the strong multi-point interactions between solvent and solute outweighs the entropic gain that would be achieved from the demixing process.[47] A statistical analysis[15] of the CSD in 2000 showed that 15% of the organic crystal structures therein were solvates, though again the inherent bias of the structures present in the CSD must be appreciated: the CSD is a database of single crystal structures that in the main were determined for the purpose of providing definitive proof of molecular structure. As discussed above in reference to the prevalence of polymorphism, for many of the 85% of molecules whose reported crystal structures are not solvates it cannot be concluded that they don't form solvates, but rather that for each of these systems no information is available beyond the fact that the single crystallisation experiment carried out yielded an unsolvated crystal structure rather than a solvate. It has been suggested that a third of molecules may form solvates[19] and McCrone's provocative statement on the prevalence of polymorphism[7] will also hold true for solvated crystal forms, leading to perhaps even this being an underestimate.

1.6.3 Crystal growth and morphology

For the crystal form that wins out at the competitive pre-nucleation stage and whose crystallites have passed their critical nucleus size, the crystal growth regime is entered. At this stage the nuclei have a crystalline internal structure and are bounded by faces, with the planes of these faces usually having low Miller indices.

A step growth mechanism for crystal surfaces has been proposed by Kossel.[48] The crystal can be envisaged as comprised of a three dimensional array of cubes, with each cube representing a molecule or growth unit in the case of self-assembled units in solution. Upon a face an island monolayer of several cubes can spontaneously form,

which constitutes the beginning of the next new layer on the face.[49] The edges of the island monolayer are steps in which kinks can occur, and these kinks are the primary attachment site for growth of the face. A molecule in the bulk solution will diffuse through the solution to the surface of the crystal and is adsorbed onto the face; it then diffuses along the plane until it encounters the step and diffuses along the step until a kink is found; the molecule joins the crystal by adding to the step at the point of the kink (figure 1.7). The kink moves along the step as further growth units attach to it until it reaches the edge of the face and grows out. At this point a new step is formed, in which a new kink occurs and this kink subsequently grows out. This process occurs until the steps themselves reach the edge of the face and an entire new layer has been deposited on the face. For further growth a new island monolayer must form. The higher the temperature the more surface imperfections, such as steps, kinks, surface-adsorbed growth units and surface vacancies occur in a face, enabling the step growth mechanism.[49] The surface nucleation step is the rate-limiting step in the Kossel model and the model has proved unrepresentative of true crystal growth because crystals can grow at levels of super-saturation lower than that required to induce surface nucleation.
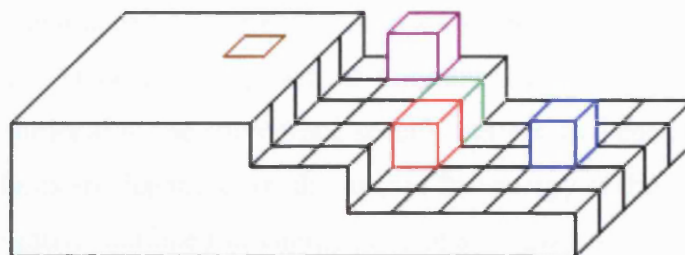
Figure 1.7: Kossel model of crystal growth on a surface. Key: brown = surface vacancy; purple = surface adsorbed growth unit; blue = surface adsorbed unit diffusing towards step; green = step unit; red = unit joining the step at kink

Burton, Cabrera and Frank[50] proposed a dislocation model, proposing that crystal faces grow *via* lattice defects such as screw dislocations (figure 1.8). Molecules can attach to the step formed by the dislocation and layers add to the face by growing in spirals around the epicentre of the dislocation. As the dislocation never grows out, the surface nucleation step of the Kossel model is avoided and the rate of growth of the face can occur close to the maximum theoretical rate for the level of super-saturation of the solution.[41]
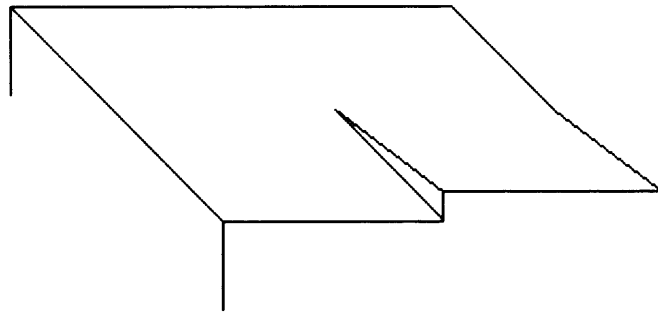
Figure 1.8: Screw dislocation on a crystal surface

Gibbs[51] showed that for a crystal in equilibrium with its environment, the faces are exhibited to minimise the surface free energy, minimising

$$\Sigma \, \sigma_i \, F_i$$

where σ_i is the specific surface free energy and F_i is the area of the ith face. This is known as the equilibrium morphology and it is distinct from the growth morphology, where the relative growth rates of faces dictates which faces have greatest morphological importance. The growth morphology was explained by Wulff[49] as deriving from "the velocities of growth of different faces in the directions of their normals are proportional to the appropriate specific surface free energies." The relative growth rates of faces are dependent on the surface free energy of the faces. Faces with a lower (greater negative) surface free energy per unit area grow more quickly than higher energy faces, and after sufficient growth of the crystal, the rapidly growing faces can 'grow out' and are no longer exhibited by the crystal as faces. The slow growing crystal faces are the morphologically important faces as it is these faces that are observed as the bounding faces in macroscopic crystals. The surface free energy is dictated by the different orientation of the growth units at the surface of each face and the functional groups presented at the face to the solution. The stronger the interactions that can be formed between the surface molecules and solute molecules joining from solution, the lower the surface free energy and the faster the growth rate. For the crystal to grow, solute molecules from solution must attach to the faces of the crystal. The process of mass deposition on to a crystal body grown from solution is comprised of several processes, some of which may occur at the same time: diffusion of solvated solute

molecules to the surface of the crystal, diffusion through the adsorption layer, surface diffusion of solvated or desolvated molecules, desolvation of molecules, integration of molecules into the lattice followed by counter diffusion of released solvent molecules away from the crystal surface.[41]

The influence of the internal crystalline structure on the growth rates of different faces and the determination of which faces have greatest morphological importance has been investigated by Hartmann and Perdok.[51] In their model, again using cubes to represent molecules, each face can be classified as belonging to one of three general types: K (kinked) faces are parallel to no strong intermolecular bond directions, periodic bond vectors, and a growth unit joining a K face can form the maximum three strong bonds upon integration onto the surface; S (stepped) faces are parallel to one periodic bond vector, and a growth unit will form two strong bonds upon joining an S surface; F (flat) faces are parallel to two periodic bond vectors, and a growth unit will only form one strong bond upon joining an F surface. K faces have a lower surface free energy and faster growth rate than S faces, and in turn S faces have lower surface free energy and faster growth rate than F faces. They note that initially, due to the driving force to relieve the super-saturation the nuclei grow with approximately isometric shape and only when the saturation condition becomes closer to 'equilibrium' does the differential energies of the faces start to determine the rate at which different faces grow. They also define 'attachment energy' as "the bond energy released when one building unit is attached to the surface of the crystal"[51] as a replacement for the surface free energy that has an entropic component, related to the solvent.

After the different attachment energies of the faces of the crystal, the second important factor that can affect the morphology of the crystal is the solvent used in the crystallisation.[52;53] When in contact with the solution the faces of the crystal will be solvated. Different faces present different orientations of the molecules that constitute the crystal, so each different face has a different functional group composition and the different faces will be solvated to different degrees in a given solvent. A face containing polar groups is going to be more strongly solvated by polar solvents than by non-polar solvents, and *vice versa* for a non-polar face. The differential solvation of the faces of the same crystal structure in two different solvents can cause the relative rates of growth of equivalent faces to be altered, giving different observed morphologies from the two

solvents. Stronger solvation of a particular face will require more energy for the diffusion of solute molecules to the surface through the adsorption layer, and for counter-diffusion of solvent molecules away from the face to allow the solute molecules to integrate into the crystal surface. This causes the face to grow more slowly than it would in a solvent that provides weaker solvation of the face.

1.6.4 Concomitant crystallisation and disappearing polymorphs

The simultaneous crystallisation of two (or more) polymorphs of a single compound from the same crystallisation experiment is known as concomitant polymorphism. Two crystallisation pathways that have very similar activation energies of crystallisation could lead to concomitant crystallisation, as both pathways will relieve the initial super-saturation at the same rate for the same available energy. It should be noted, however there will be an inherent tendency for the more stable crystal structure to form in this situation, as it releases more energy to compensate for the energy required to form the phase boundary between the solution and the new crystalline state.[46] In this case, there will be a thermodynamic tendency for the metastable form to convert to the stable form after the initial crystallisation. An example of a compound that can exhibit concomitant polymorphism is 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile (or ROY, after the observation that the different polymorphs exhibit various shades of red, orange and yellow), a pharmaceutical precursor produced at Lilly Research Laboratories.[24] This compound was initially reported to crystallise in 6 different solvent free forms from methanol, occasionally showing mixtures of polymorphs in single crystallisation experiments. This effect, if present in a drug substance, would complicate attempts to formulate a robust procedure to manufacture a single form of the drug substance.

The discovery of a new thermodynamically stable polymorph of a compound may change the crystallisation result of a previously reliable method for growing a known crystal form, giving the new polymorph in preference to the original metastable one. The mere occurrence of the new polymorph in the same laboratory may cause crystals of the original polymorph to convert to the new form by seeding of the atmosphere, and may greatly increase the difficulty of regaining the original polymorph or even completely stop it from being grown again in that locality.[54] In such a situation, the now elusive metastable polymorph is an example of a disappearing polymorph.

Intentional seeding has been used for many years as an important method for promoting crystallisation in a solution, and is useful for reliably preparing one crystal form over another.[54] Once a sample of the new more stable polymorph is opened to the laboratory space, small crystalline particles seeds can be unintentionally spread throughout the laboratory; these particles can find their way into new crystallisation experiments, by contamination of apparatus, or by airborne diffusion if small enough. This unintentional seeding with the new more stable polymorph may suppress the growth of the metastable form in crystallisation experiments. The presence of these seeds circumvents the rate-limiting homogeneous nucleation step of the crystallisation process – crystals grow out rapidly from the seed nuclei of the more stable form, giving only this form as the crystallisation product.[55] Once a more stable polymorph is discovered and displaces the original polymorph as the usual crystallisation result, it is always in principle possible to recover the 'disappeared' form – however the caveat must be included that it may require extensive time and effort to discover the experimental conditions required to yield the disappeared form.[56]

1.7 Crystallisation methods

1.7.1 Classic crystallisation methods

The super-saturation required to induce crystallisation is commonly achieved in one of three ways: by cooling a saturated solution, by allowing solvent to evaporate from a solution, or by the addition of a miscible anti-solvent to reduce the overall solubility.

The parameter most usually varied in a crystallisation screen to attempt to find polymorphs is the solvent used for the crystallisation experiments, though other factors that can be varied combinatorially with solvent include rate of cooling, rate of solvent evaporation, initial level of solution saturation and the presence or absence of agitation of the solution during crystallisation.

A recent invention in the field of crystallisation screening has been the advent of high through-put screens, that use traditional crystallisation methods, but in an automated, parallelised platform to increase the rate of crystallisation conditions that can be investigated.[57-59]

1.7.2 Novel crystallisation methods

Much research has been carried out on designing tailor made impurities, whose presence in small quantities in a crystallisation can alter the polymorphic outcome.[52;53] The mode of action of the impurity is selective adsorption onto some of the faces of the nuclei of one potential polymorph at the pre-critical nucleus stage, with this adsorption inhibiting further growth of this form, allowing nuclei of a different form to reach the critical nucleus stage and become the crystallisation product.[60] For some systems the additive does not work at the nucleation stage, but upon formation of an initial metastable form, preventing its subsequent inter-conversion to a more stable form: crystallisation from solutions with no additive yields the stable form, but solutions with additive present will yield the metastable form.[61;62]

Polymers have been employed as heteronucleation surfaces, with different polymers yielding different polymorphs. Using this method new polymorphs of the pharmaceuticals carbamazepine[63] and sulfamethoxazole[25] have been discovered, which have proved inaccessible by traditional crystallisation routes.

High pressure crystallisation has also been a rich source of new polymorphs. Most polymorphism studies are carried out at ambient pressure, and by introducing variation in the applied pressure, occurrence domains of high pressure phases can be found and the new forms characterised. High pressure crystallisations have been performed both on pure samples from the liquid phase, such as for acetone,[64] and from solutions, such as for piracetam[65] and acetamide.[66] Direct crystallisation of high pressure phases has been observed as well as solid-solid phase transitions upon application of pressure.

1.8 Intermolecular interactions

1.8.1 The origin of intermolecular interactions

Implicit so far has been the assumption that for a particular compound, the crystalline solid state is favoured below its melting point – raising the question as to why the crystalline solid state forms. Upon cooling through the melting point of the compound and crystallisation, the enthalpy gain from ordering the molecules into the crystal and the thermal randomisation of the surroundings outweighs the loss of entropy caused by the ordering of the molecules.[67] The enthalpy gain comes from attractive intermolecular

forces producing an inherent drive for the molecules to form a close packed arrangement, to minimise empty space in the crystal structure, as defined by Kitaigorodski's principle of close packing.[68] A crystalline product maximises the enthalpy gain from formation of intermolecular interactions, while minimising the repulsive interactions and produces a close packed structure.

The attractive intermolecular interactions that hold crystals together arise from the non-uniform electronic distribution in the molecules. These forces are an order of magnitude weaker than the intramolecular covalent bonds that join the atoms of the molecule together. The strongest intermolecular interactions between uncharged molecules usually have energies of 16-60 kJ mol^{-1},[69] with a typical covalent bond having a dissociation energy at least 200 kJ mol^{-1}. The intermolecular interaction between two spherical molecules can be graphically represented as a Morse curve (figure 1.9).[70] At very close intermolecular distances there is a strong repulsive interaction between the molecules due to the overlap of the electron clouds of the two molecules, which would eventually lead to electrons with the same quantum numbers occupying the same space, violating the Pauli exclusion principle.[71] At longer intermolecular separations the interactions between the molecules are attractive, and are the result of mutually advantageous electron arrangements in the two molecules, such as attractive electrostatic interactions, where partially negatively and positively charged groups on the two molecules are orientated towards one another, or polarisation, where the charge distribution in one molecule distorts the charge density in the other to lower the overall energy, or dispersion forces which are a quantum mechanical effect deriving from correlation of the instantaneous dipolar fluctuations in the charge distribution of the two molecules.[70;72] Some of the attractive forces are directional in nature, and consequently have been defined beyond their general role as an attractive intermolecular force – these include dipole-dipole interactions (the tendency for dipolar molecules to pack in a head-to-tail arrangement), π-π interactions between aromatic ring sub-units, and hydrogen bonds.
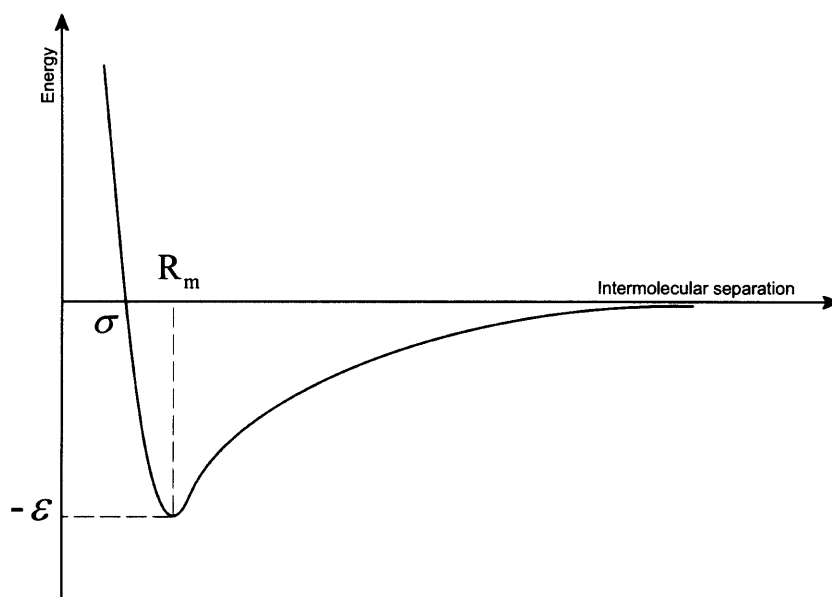
Figure 1.9: Morse potential curve for the intermolecular interaction of two spherical molecules. The energy minimum occurs at intermolecular separation R_m, of magnitude ε

1.8.2 Intermolecular potentials

The ideal potential to model intermolecular interactions and calculate the free energy of crystal structures at a defined temperature would have to use an accurate methodology that addressed all of the following terms:[73]

$$G = U^o + U_{intra} + U_{inter} - T\left(S_{inter} + S_{intra}\right) + PV$$

Currently employed methods used in crystal structure prediction usually only calculate the U_{inter} term at 0 K, neglecting the zero point energy, U^o, the vibrational energy a molecule retains even at 0 K. The intramolecular term U_{intra} is assumed to be constant between structures when the rigid body approximation is employed and the difference in the PV term is negligible between structures. The entropy based terms are neglected due to the 0 K nature of the calculation. This reduces the requirement, for rigid molecules, to that of an accurate description of the intermolecular interactions within the crystal to rank the structures at 0 K. This is modelled by partitioning the intermolecular potential into its constituent contributions and modelling each of these.

Long range forces include the electrostatic contribution, polarisation and dispersion, while the short range terms include repulsion, exchange and charge transfer.[72]

The repulsion term is a short range interaction that occurs when molecules are brought closer together than their van der Waals contact distance. The Pauli Exclusion Principle prevents the overlap because it would cause electrons with the same quantum numbers (but from two different molecules) to attempt to occupy the same space. At short range this is a steeply rising repulsive interaction usually modelled as an exponential term.[38]

The dispersion term is a long range term that is universally attractive, is always present and is a purely quantum mechanical effect. It arises from the instantaneous correlation of induced dipoles, or electronic motion, in the molecules.[70]

The electrostatic forces derive from the distribution of the valence electrons in a molecule. Coulombic interactions occur between molecules, between the undistorted charge distributions of the molecules. These forces can be either attractive or repulsive and persist over the longest range of any of the contributions to the intermolecular potential. For all species except spherical ions the electrostatic forces have an orientational dependence, and the electrostatic forces between two molecules can be either attractive or repulsive depending on this orientation. These forces are responsible for important features such as hydrogen bonding and π-π stacking, which have strong orientational dependence.[72]

Polarisation forces (or induction) are attractive and arise from the distortion of a molecule's charge distribution in the field of the undistorted charge distribution of another molecule. A similar term occurs for the reverse action – the distortion of the second molecule's charge distribution by the first molecule. These distortions only occur if they lower the overall energy, and hence are always attractive. Structures that have strong hydrogen bonding, with electron clouds in close contact in the hydrogen bonding region, may exhibit polarisation effects that could be significant.[72] The exchange interaction is an attractive force that is caused by the ability of the electron density of two individual molecules, at short range, to extend over the whole of the molecular pair. This is usually included with the repulsion term, which is the dominant contribution.[72] Charge transfer can occur when molecules are at close range with overlapping charge distributions. Electron density can be transferred from the occupied

orbitals of one molecule to the unoccupied orbitals of the other can occur giving rise to the attractive charge transfer term. This is an attractive term and it is usually responsible for part of the short-range modification of the polarisation.[72]

Two approximations are used to sum the intermolecular forces to give the lattice energy of the crystal structure. The first is the atom-atom approximation which assumes that the intermolecular potential between two molecules is the sum of all of the interactions between all atoms in the first molecule with all of the atoms in the second:[72]

$$U(R,\Omega)(R= \sum_{i\in A,\, k\in B} U_{ik}(R_{ik})$$

where atoms i belongs to molecule A and atoms k belongs to molecule B with each pair of atoms, i and k separated by distance R_{ik}. The second is the pairwise additivity approximation, where the lattice energy of the crystal is equal to the sum of the interactions between every pair of molecules in the crystal, with no treatment of many-body terms.[72] The validity of the pairwise approximation differs between the terms that make up the overall intermolecular potential. Electrostatic interactions are strictly pairwise additive, dispersion is approximately pairwise additive and repulsion is approximately pairwise additive at common molecular separations in the crystal.[70] Summation of every molecule with every other molecule in the crystal is impracticable and a cutoff radial limit is usually chosen, in conjunction with Ewald summation to accelerate the convergence of the electrostatic term.

Commonly the electrostatic, dispersion and repulsion parts of the intermolecular potential are the only ones specifically modelled for use in lattice energy minimisation. The dispersion and repulsion contributions are modelled together using empirical models, using parameterisations for each type of atom that are assumed to be transferable between molecules. The electrostatic contribution is usually calculated specifically for the molecule under investigation, from a more sound theoretical basis. Truly *ab initio* potentials, with no empirical contribution to the calculation are currently too expensive for the large number of energy minimisations required for crystal structure prediction,[74] though attempts have been made to refine the lowest energy structures of a search using an *ab initio* derived potential.[75]

1.8.3 Dispersion-repulsion models

The dispersion and repulsion terms are commonly modelled together and are modelled for each atom type using an empirical description which has been parameterised against appropriate physical data, usually for CSP crystal structures and heats of sublimation. Usually the parameters for each atom type are assumed to be transferable between molecules. The most well-known dispersion-repulsion model is the '12-6' Lennard-Jones potential:[76]

$$U_{LJ}(R) = 4\varepsilon \left[\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^{6} \right]$$

where the 12th power term models the repulsion which quickly decays with increasing interatomic separation and the 6th power term models the longer range dispersion. R is the interatomic separation; ε is the depth of the energy well; $\sigma = 2^{-1/6}R_m$ where R_m is the separation at which the energy minimum occurs and σ is the position where the repulsive arm crosses the zero of the energy scale (figure 1.9).[70] Both ε and σ are parameters that can be varied to fit the experimental data.

Buckingham modified the Lennard-Jones potential to combine an exponential-based model for the repulsive term with the same term for the dispersion to produce an 'exp-6' model:

$$U_{BUCK}(R) = Ae^{-BR} - \frac{C}{R^6}$$

where R is the interatomic separation, and A, B and C are parameters that can be fitted according to experimental data. Once the Buckingham potential is fitted for each atomic type of interest, the isotropic atom-atom approximation is employed – the molecule is assumed to be comprised of a superposition of spherical atoms. The dispersion-repulsion part of the intermolecular potential is the summation of the interaction of each atom in each molecule with all other atoms in all other molecules, within the cutoff, with a Buckingham potential centred on each atom with parameters specific to that atom's type.

Williams and co-workers developed a self-consistent potential based on the Buckingham potential and parameterised against sets of crystal structures for each atom type. Initially parameters for carbon and hydrogen were developed against hydrocarbon crystal structures[77;78] and once these parameters were optimised they were held constant for optimisation of the parameters for oxygen[79], nitrogen,[78] fluorine[80] and chlorine.[81] Coombes et al. reparameterised Williams' original hydrogen potential to give separate parameters for polar hydrogen.[82] For each atom type the homoatomic terms were derived and the heteroatomic cross terms were calculated using well-known geometric combining rules:

$$A_{\iota\kappa} = \left(A_{\iota\iota}A_{\kappa\kappa}\right)^{\frac{1}{2}} \qquad B_{\iota\kappa} = \frac{1}{2}\left(B_{\iota\iota} + B_{\kappa\kappa}\right) \qquad C_{\iota\kappa} = \left(C_{\iota\iota}C_{\kappa\kappa}\right)^{\frac{1}{2}}$$

where ι and κ are different atom types. The sets of crystal structures used by Williams were small, less than 10 structures, each containing just C, H and the atom type to be parameterised, with no hydrogen bonding in any of the structures, and the crystal structures were all determined at room temperature. The electrostatic part of the intermolecular potential was modelled using point charges.

Williams later revisited the parameterisations, developing parameters not only for atom types but for structural sub-classes of atom types, such as differently hybridised carbons, and based on larger crystal structure sets and sublimation enthalpies, while using more sophisticated electrostatic models.[83-85]

Many other force fields have been developed based on either liquids or macromolecular systems, but do not provide sufficiently accurate modelling of the crystalline solid state.[38]

1.8.4 Electrostatic models

The most commonly adopted method used to represent the electrostatic part of the intermolecular potential for a molecule is to calculate the electrostatic potential for the isolated molecule from the *ab initio* calculated wavefunction using an affordable basis set and a high level calculation, such as SCF, with or without the MP2 electron correlation correction. From the wavefunction of the molecule the electrostatic potential at a grid of points in space outside the van der Waals radius of the molecule can be

calculated and atomic point charges fitted to optimally reproduce the electrostatic potential at the grid points. The resultant charges, known as electrostatic potential (ESP) fitted atomic point charges, have no physical meaning – their values simply optimise the fit to the grid of points, giving the best description possible of the electrostatic potential at these points.

The simple nuclear site electrostatic model is limited as it assumes that the charge on a particular atom is spherical. Extra non-atomic charge sites can be included at physically realistic sites to model effects such as lone pairs, known as the ESP extended site point charge model. These methods have been compared to the original nuclear-site charge model and have been shown to be appreciably more accurate.[85;86] More recently a similar method has been developed with point charges placed not only on nuclear sites, but also at satellite positions that do not necessarily correlate with physical features such as lone pairs, but are sited to further optimise the fit of the whole set of point charges to the *ab initio* calculated charge density.[87]

Distributed multipole analysis (DMA)[88] is a more sophisticated method that still relies on the calculation of the electrostatic potential of the molecule using *ab initio* methods, but instead of fitting only point charges to the nuclear positions, it calculates a series of multipoles at each nuclear site comprising charge, dipole, quadrupole, octopole, and hexadecapole, directly from the wavefunction of the molecule. The higher terms model non-spherical features and has been shown to successfully model interactions such as π-π interactions and hydrogen bonds.[89;90] Several studies have investigated the effect of different electrostatic models on crystal structure lattice energy minimisation, predominantly concluding that multipole-based models are superior to ESP fitted atomic charges.[82;91;92]

Filippini and Gavezzotti[93] have developed a Buckingham-based potential in which they have no explicit electrostatic model, but include electrostatic contributions within the Buckingham parameters. They parameterised all of the homoatomic potentials for C, H, N, O, Cl and S and also parameterised the heteroatomic cross-terms separately rather than using geometric combining rules. Only non-hydrogen bonded crystal structures were included in the training sets used for parameter fitting and the potential was shown to reproduce many crystals properties for a range of non-hydrogen bonding molecules.

1.8.5 Hydrogen bonds

In many organic molecules the functional groups required to produce hydrogen bonds are present. Hydrogen bonding is the most important structure directing determinant, and the most useful interaction for assessing the packing modes present in crystal structures. Pauling gave an early definition of hydrogen bonding: "A hydrogen bond is an interaction that directs the association of a covalently bound hydrogen atom with one or more other atoms, groups of atoms or molecules into an aggregate structure that is sufficiently stable to make it convenient for the chemist to consider it as an independent chemical species."[94] An alternative definition is possible defining the hydrogen bond in terms of the types of atoms that are required to participate: a hydrogen bond is an attractive intermolecular interaction between a hydrogen atom that is covalently bonded to an electronegative atom, and an electron rich atom on another molecule (intermolecular hydrogen bond), or to an electron rich atom that is part of a different functional group on the same molecule (intramolecular hydrogen bond) (scheme 1.2).

$$D^{\delta-} - H^{\delta+} \cdots A^{\delta-}$$
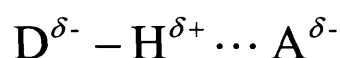
Scheme 1.2: Schematic representation of a hydrogen bond. D is the hydrogen donor electronegative atom; A is the electronegative acceptor atom

The principal components of the intermolecular potential that contribute to hydrogen bonding are the electrostatic and polarisation terms, with dispersion, repulsion and charge-transfer components being less influential on both the strength and directionality of hydrogen bonding.[71] The electronegative nature of the donor atom withdraws electron density from the D-H bond, leaving a partial positive charge on the hydrogen atom. This partial positive charge interacts with the electron rich acceptor atom, such as the lone pairs on oxygen, to form the hydrogen bond. Hydrogen bonds have been classed according to their strength, ease of attainment and reliability of formation (table 1.1) [95;96]

Very strong		
[F···H···F]⁻	[N···H···N]⁻	P-O-H···O=P
Strong		
O-H···O=C	N-H···O=C	O-H···O-H
Weak		
C-H···O	N-H···F-C	O-H···π

Table 1.1: Hydrogen bond classification

Hydrogen bonds that contain ionic groups are beyond the scope of this work. Strong hydrogen bonds occur when the hydrogen is polarisable and is covalently bonded to an electron-withdrawing donor atom, oxygen or nitrogen, and interacts with a partially negatively charged and comparatively less polarisable acceptor atom, such as oxygen.[71] In the examples of weak hydrogen bonds, carbon acts as a weak hydrogen bond donor as it is not particularly electronegative producing a less positively charged hydrogen atom, fluorine as a hydrogen bond acceptor is not sufficiently polarisable to produce a strong interaction and the π-electron cloud formed by aromatic groups do not have a focus of electron density to act as a strong hydrogen bond acceptor.

The strong hydrogen bonds defined in table 1.1 have D···A lengths that are less than the sum of the van der Waals radii of the donor and acceptor atoms, with the hydrogen atom completely subsumed into the van der Waals sphere of these two atoms.[70] The formation of a strong hydrogen bond lengthens the D-H distance, by 0.01-0.05 Å.[96] The sum of the van der Waals radii[97] gives an upper limit to O-H···O hydrogen bonds of 3.04 Å (with 2.75 Å a common average[98]) and 3.07 Å for N-H···O hydrogen bonds (with 2.85 Å a common average[98]). It should be stressed that such upper limits should not be construed as absolute, as in many cases hydrogen bonds longer than the sum of the van der Waals radii can be identified that, while weak in nature, may well have a role in the production of the structure observed. This is most particularly true with hydrogen bonds of the form C-H···O, because of their inherent weakness, which have C···O distances in the range 3.0 to 4.0 Å[98] (with 3.56 Å the mean distance calculated in a recent study[99]). Strong hydrogen bonds exhibit preferred directionality, associating with the regions of electron density on the acceptor molecule, such as lone pairs. The angle D-H···A can be assumed to adopt a near linear disposition in the absence other forces. However, in crystal structures deviation from linearity

commonly occurs because of the adoption of the optimal compromise between all intermolecular interactions in the crystal structure, with distortion of a particular hydrogen bond contact repaid by a more favourable interaction elsewhere in the structure. Weaker hydrogen bonds are more easily distortable from the ideal geometry that strong hydrogen bonds.

Etter[100] produced a set of general rules to determine the likely hydrogen bonding that a molecule will adopt in the crystalline state:

- 1. All good proton donors and acceptors are used in hydrogen bonding

- 2. Six-membered-ring intramolecular hydrogen bonds form in preference to intermolecular hydrogen bonds

- 3. The best proton donors and acceptors remaining after intramolecular hydrogen bond formation form intermolecular hydrogen bonds to one another

Exceptions to these rules are known such as alloxan[101;102] where, even though hydrogen bond donors and acceptors are present, they are unused in the crystal structure. While these rules pertain to the strong hydrogen bonds that can potentially be formed, the influence of weaker interactions, especially C-H$\cdots$O,[103-105] should not be discounted. Where strong hydrogen bonds are absent, weaker hydrogen bonds determine the crystal structure adopted and C-H$\cdots$O interactions can be structure directing in some cases even when strong hydrogen bonds are present.[98]

The reliability of formation of strong hydrogen bonds and their directional nature enables crystallographers to describe the packing of the crystal structure in terms of the hydrogen bond pattern. Graph set analysis[106;107] has been employed in a systematic method developed for this purpose, based on the general scheme shown in scheme 1.3.

$$G_d^a(n)$$

Scheme 1.3: Graph set descriptor

G is the descriptor of the hydrogen bond pattern: C designates the pattern as an infinite chain, S an intramolecular hydrogen bond pattern, R a ring pattern and D other finite patterns. The superscript a is the number of acceptors in the repeat unit and the subscript d is the number of donors in the repeat unit. n is the degree of the pattern and is the number of atoms in the repeat unit. All graph set descriptors that include only one unique hydrogen bond are termed first order (scheme 1.4A), while those that include two hydrogen bonds are second order, and this continues to higher order (scheme 1.4B). Higher order graph sets can contain the most descriptively useful graph sets. First order graph sets are also defined as hydrogen bonded motifs, though this term will be used more permissively in this work for the most descriptive hydrogen bonded feature in the structure.
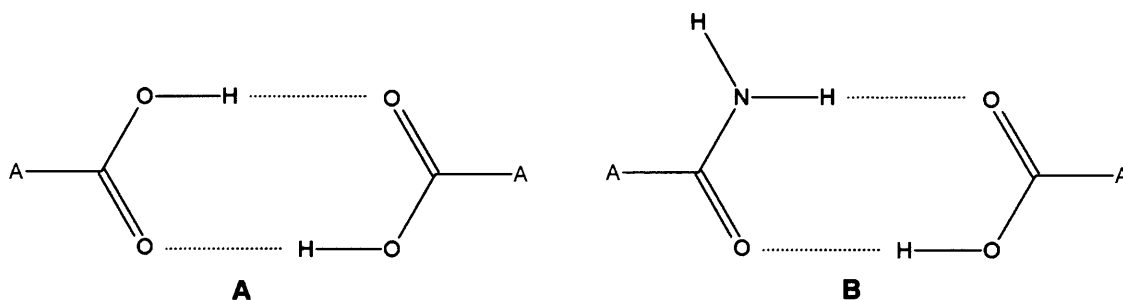
Scheme 1.4: Graph set examples. A – dimerisation of a carboxylic acid with first order graph set $R_2^2(8)$; B – dimerisation of an amide and a carboxylic acid with two first order $D_1^1(2)$ graph sets and one $R_2^2(8)$ second order graph set

The identification or rejection of weak hydrogen bonds in a crystal structure can have a critical influence on the calculated graph set.

1.9 Crystal structure prediction

The *ab initio* prediction of the crystal structure, or structures, that a molecule will adopt, prior to the possibly expensive and time consuming synthesis of that molecule, would be a powerful tool to those interested in developing organic materials with specific physical properties as the correctly predicted crystal structure(s) would contain all the necessary molecular packing information to be able to subsequently predict the physical and chemical properties of the crystalline form(s). In the cases where the crystal structure(s) exhibit undesirable properties then an alternative course of action can be considered.

Crystal structure prediction (CSP) is related to crystal engineering, though distinct from it. Crystal engineering seeks to design and control the way that molecules crystallise[55] by exploiting intermolecular interactions that have been proven to occur reliably in other crystal structures. These intermolecular interactions are usually strong hydrogen bonds such as O-H···O, N-H···O and these well defined hydrogen bonds, when present in crystal structures, are the building blocks of 'supramolecular synthons',[108] defining the structure in terms of intermolecular interactions. Crystal engineering comprises the analysis of already known crystal structures to draw conclusions about the intermolecular interactions that direct the assembly of the crystals, and the computational analysis of intermolecular interactions using quantum mechanical methods and empirical potentials, to design new molecules which exhibit predicted intermolecular interactions.[109] The ability to determine the intermolecular interactions that a designed molecule will exhibit leads to the ability to design the crystal supermolecule with specific desirable properties.[110] Moulton and Zaworotko[111] provide delineation between crystal structure prediction and crystal engineering – crystal engineering deals with less precise design, concentrating on network prediction, while crystal structure prediction is more precise, aiming to predict space group, unit cell and the precise packing details.

In a *Nature* editorial in 1988, Maddox[112] proclaimed that the inability to predict crystal structures from the molecular structure was a "continuing scandal" in physical science. The question "Are crystal structures predictable?" is a prudent question to ask[113;114] prior to attempting to develop a methodology to do so. In the first instance the problem seems relatively simple, usually dominated by strong directional interactions such as hydrogen bonding to provide the principal molecular associations and the close packing principle requiring dense structures. However, for a particular compound, there is often more than one way to pack the molecules into a crystal structure to satisfy the hydrogen bonding rules and produce a dense structure – it is the more subtle interactions that can be the key determinant of which of the possible structures is observed. Indeed the implicit assumption that it is thermodynamic stability that drives the outcome is not completely correct – nucleation kinetics (section 1.6) can determine the outcome of a crystallisation, requiring more than just lattice energy to be used as a

structure discriminator between possible crystal structures. Gavezzotti[113] has summarised the steps required for successful crystal structure prediction:

- 1. Generation of a set of polymorphs

- 2. Unequivocal detection of the most stable one at a given temperature

- 3. Modelling of the nucleation kinetics to determine which phase[s] which actually appear under given conditions

He concluded his account by summarising that this was not yet achievable (in 1994). Success in step 1, the generation of a set of energetically feasible potential polymorphs (the crystal energy landscape), depends upon the method used, its inherent limitations and the computational power available to thoroughly explore the lattice energy surface to locate all possible minima thereon. Upon generation of a set of hypothetical crystal structures by a CSP method, each structure must be energy minimised using an intermolecular potential to rank the structures by stability. Step 2, the identification of the most stable phase, is dependent upon the intermolecular potential employed to calculate the lattice energies of the structures generated in step 1, and the errors in the potential *versus* the energy gap between potential polymorphs. Step 3, the modelling of nucleation kinetics, is far from achievable in the near future, and yet is identified as the crucial aspect missing from current attempts to predict crystal structures.[45] Current methods usually rely on satisfying the first two of the three criteria above – the generation of a set of potential polymorphs, the reliable ranking of these structures using available intermolecular potentials and the identification of the subset of these structures within the usual energy window for polymorphism (usually up to 10 kJ mol^{-1} above the global energy minimum structure[1]).

The possible outcomes of crystal structure prediction have been schematically summarised in figure 1.10.[90] Figure 1.10a shows the most favourable situation where the known structure has been found by the crystal structure prediction method and is significantly more stable than all other hypothetical structures. It is desirable for the known structure to have the lowest lattice energy and to be the densest structure, to

satisfy the close packing principle. Figure 1.10b shows an outcome where crystal structure prediction indicates that there may well be a single thermodynamically more stable crystal structure. Figure 1.10c is the most common result of crystal structure prediction where the known structure is low in energy, but there are many other distinct structures within the same energy window.
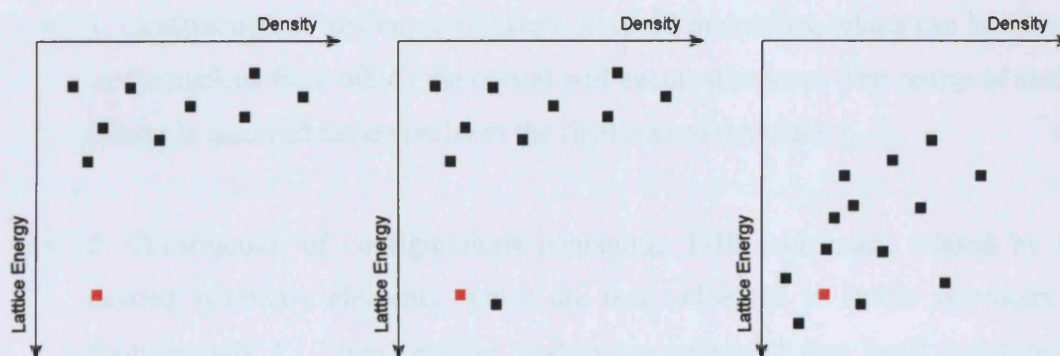
Figure 1.10: Possible crystal structure prediction outcomes; left to right a) ideal scenario, with no further low energy structures; b) the known structure may not be the thermodynamically most stable; c) the known structure is among a range of low energy structures. In all cases the experimental structure is shown in red [90]

1.9.1 Crystal structure generation methods

A reliable crystal structure generation method has to produce dense crystal structures in a range of space groups including all of the common space groups that organic molecules crystallise in. The exploration of the crystal packing energy surface has to locate as many of the low energy minima as possible, including the true global energy minimum. The degrees of freedom that can be varied, for a $Z' = 1$ system, to produce different hypothetical structures include: the unit cell constants, a, b, c, α, β, γ, ranging from one variable (cubic symmetry) to six (triclinic symmetry); the rotations of the molecule within the asymmetric unit, the Eulerian angles; the translations of the molecule within the asymmetric unit; the internal degrees of freedom within the molecule.[115] For molecules without any flexible torsion angles, the assumption that the molecular conformation is rigid is commonly employed – the isolated gas phase quantum mechanically optimised molecular structure is usually used as the search input. Internal rotational degrees of freedom introduce extra complexity into the crystal structure prediction of flexible systems, requiring not only a more expensive search procedure to address the extra intramolecular freedom, but also quantification of the

intramolecular energy differences between conformations to correct the lattice energy for the energy required to bring about the conformational differences.[75;116]

Verwer and Leusen have categorised the different methods that have been used to generate crystal structures:[117]

- 1. Construction of low energy clusters of 10-50 molecules, which can be viewed as the nucleus from which the crystal will eventually grow. The centre of such a cluster is assumed to be similar to the final crystal structure. [...]

- 2. Construction of configurations containing 1-10 molecules, related by the desired symmetry elements, which are then subjected to lattice symmetry to form crystals. [...] non-periodic clusters are generated first [and] translational symmetry is introduced to simulate a bulk environment.

- 3. Generation of a large set of crude molecular packings, subject to the desired space group symmetry, which are then energy minimised. Periodicity is assumed at all stages [...]."

Programs in current use include Polymorph Predictor which uses a simulated annealing algorithm, (category 3),[115] PROMET which generates small clusters exploiting inversion centre, glide, screw symmetry operators which are translated to produce crystal structures (category 2),[118] UPACK which uses systematic or random methods to vary molecular orientation and unit cell parameters within fixed space groups (category 3)[119] and MOLPAK which uses common coordination environments in common space groups as the basis for crystal structure generation (category 3).[120]

1.9.2 International blind tests of crystal structure prediction

On three occasions in the past seven years, 1999, 2001 and 2004, the Cambridge Crystallographic Data Centre (CCDC) has organised an international blind test of crystal structure prediction,[121-123] open to academic groups and commercial companies involved in the field. In each blind test an independent referee chose three molecules whose crystal structures had been determined, but unreleased. The molecular structures

were passed to the participating groups and three predictions for each molecule were requested. The three structures were chosen to present different degrees of difficulty to the CSP algorithms. The degrees of complexity of the three molecules used during the first blind test (CSP1999) were defined to aid participants:[121]

"the maximum size of the molecule was restricted to 30 atoms, including hydrogen, and the space group was stipulated to be one of the more commonly observed ones with one molecule per asymmetric unit, i.e. $Z'=1.0$, although no defined list of allowed space groups was provided. It was also specified that submissions should belong to three categories of perceived difficulty for prediction:

(i) a small, rigid molecule with only C,H,N,O atoms allowed and less than 20 atoms

(ii) a small, rigid molecule with some less common elements

(iii) a molecule with some small amount of conformational freedom"

For the second blind test (CSP2001) the number of atoms allowed was increased to 40 and the flexible molecule test (iii) was increased to two flexible torsion angles.[122] The third blind test (CSP2004) lifted the restriction on space groups and allowed $Z' = 2$ structures as well as $Z' = 1$ structures.[123] In none of the tests were disordered structures, co-crystals or solvates included. Comparison of the submitted predictions with the experimental structures, which were released upon the expiration of the submission deadline, was by coordination sphere overlay, allowing predictions that lay within a defined root mean square overlay tolerance of the corresponding experimental structure, but in an incorrect space group, to be classified as successful.[124]

The results of the three blind tests are summarised in table 1.2. The overall success rate is low, but with more success for the smaller rigid systems. The flexible systems, group (iii), showed only one success from a total of 111 predictions, indicating that the CSP methods currently employed cannot thoroughly search the crystal packing energy surface when extra degrees of intramolecular freedom are introduced. The smallest molecule included in any of the blind tests, the CSP2004 category (i) molecule azetidine, was the only test subject used that crystallised with two independent molecules in the asymmetric unit, and this was a major contributory factor to the lack of

success for this molecule. The $Z' = 1$ limitation is still a very real constraint for CSP methods – only 8 of the 18 participating groups explicitly addressed the possibility of $Z' = 2$ crystal structures in their structure generation method for azetidine.

The least successful blind test has been the most recent one, CSP2004, and suggests that the pace of methodological development has not proceeded as fast as the test subject constraints have been lifted – indeed the original flexible molecule constraints as defined in 1999 are still as much of a challenge today. The problems of thoroughly searching the crystal packing energy surface, ranking the generated structures using a reliable intermolecular (and potentially intramolecular) potential, and then choosing which three of the myriad low energy structures to submit are as evident in the last blind test as the first – from Gavezzotti's requirements for successful crystal structure prediction[113] the third step still proves to be the decisive hurdle yet be addressed.
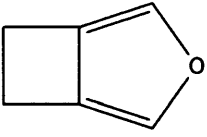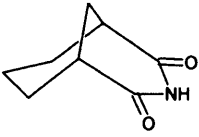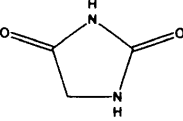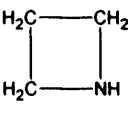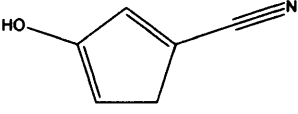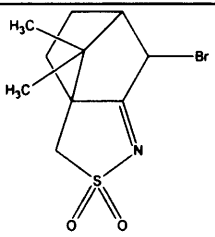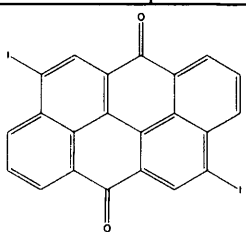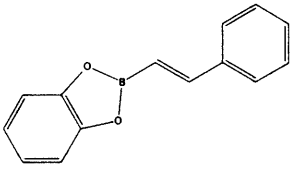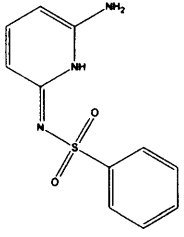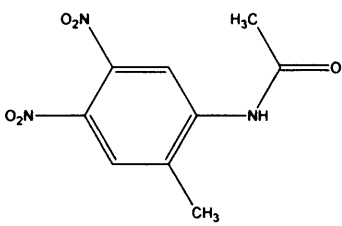
	CSP1999	CSP2001	CSP2004	
(i) Rigid				
Designator	I	IV	VIII	XI
Participants/ Success	11/[0/4] *	15/2	15/4 †	18/0
(ii) Atom types				
Designator	II	V	IX	
Participants/ Success	8/1	15/4	15/1	
(iii) Flexible				
Designator	III	VI	X	
Participants/ Success	11/1	11/0	15/0	

Table 1.2: Summary of results from the three CCDC international blind tests; * two polymorphs were characterised – one was not predicted, while the other was predicted by four groups; † molecule was only partially blind

1.10 Outline of project of study

The aim of using crystal structure prediction to identify the potential motifs present in crystal structures, and especially those that pack to give low energy predicted structures, and then use this to influence the crystallisation strategy is an important goal for the pharmaceutical industry for compounds where only a small amount of material is available in early development.[125;126] The studies presented in this work combine extensive experimental crystallisation screening to discover new polymorphs of the

subject molecules in conjunction with crystal structure prediction studies. For each molecule studied the low energy predicted structures could be classified according to a small range of principal hydrogen bonded synthons, and the experimental structures discovered in the course of the crystallisation screens were rationalised in terms of these motifs to lead to further understanding of the origin of the polymorphism of these molecules.

Two related fluorinated molecules, 5-fluorouracil and 5-fluorocytosine, both widely used in medicine, were studied using CSP and manual crystallisation techniques (chapters 3 & 4). In each case neither of these molecules was previously known to be polymorphic, and both were also chosen because they were capable of forming a range of strong hydrogen bonds. CSP studies were carried out on coumarin and its derivatives, 4-hydroxycoumarin, 7-hydroxycoumarin, 6-methoxycoumarin (chapter 5), none of which were known conclusively to be polymorphic. Coumarins are a class of compounds many of which are used as laser dyes, with some having applications as pharmaceutical precursors and as fine chemicals. A manual crystallisation screen was performed on 4-hydroxycoumarin and an abbreviated screen on 6-methoxycoumarin. The rigid test molecule from CSP2001, 3-azabicyclo[3.3.1]nonane-2,4-dione, was subjected to a state-of-the-art automated high throughput crystallisation screen to attempt to discover a hydrogen-bonded dimer based polymorph, the possibility of which is strongly suggested by the results of the blind test (chapter 6). The prevalence of hydrates of pharmaceutical molecules led to an investigation into the viability of extending the CSP method to predicting the crystal structures of monohydrates using 5-azauracil monohydrate as a test system (chapter 7).

Chapter 2 Experimental and computational methods

2.1 Introduction

This chapter describes the manual experimental polymorph screening undertaken in this thesis, along with the analytical techniques used, and the computational crystal structure prediction method.

Section 2.2 outlines the manual crystallisation screening methods used in this work and provides an introduction to high through-put crystallisation techniques. Details of the two manual crystallisation techniques, solvent evaporation and vapour diffusion, are given.

Section 2.3 details the analytical methods used during this work including single crystal X-ray diffraction, powder X-ray diffraction and simultaneous thermal analysis.

The workflow of the specific crystal structure prediction used in this thesis is presented at the beginning of section 2.4 along with an outline of each of the steps in the workflow. The principal programs and the intermolecular potential used in the method are then presented in greater detail in separate sub-sections.

2.2 Crystallisation screening method

2.2.1 Manual crystal screening

The manual crystallisation screening carried out in this work used solvent evaporation and vapour diffusion crystallisation methods, changing the crystallisation conditions by varying the solvent. Each crystallisation used between 2-10 ml of solvent in which the target compound was dissolved. Crystallisations were controlled to attempt to yield single crystals of sufficient size (approximately 0.2 mm^3) for single crystal X-ray diffraction. The production of single crystals allowed relatively quick determination of the unit cell dimensions and full structure determinations were undertaken each time a new unit cell was encountered. The weakness in using single crystal X-ray diffraction as the principal characterisation method is that only one crystal is checked, but the resulting crystal structure may not be representative of the bulk sample. Indeed the common practice of choosing the 'best' crystal from a sample for structure determination may aggravate this weakness: the chosen crystal may be a minor

crystallisation product, whereas the remaining, more microcrystalline, material may be the predominant crystallisation result. To address this weakness, crystals of 'typical' morphology were chosen whenever possible. The strength of this technique is that only one crystal need grow for a full structure determination. X-ray powder diffraction was used to characterise all commercially supplied samples and also for cases where single crystal X-ray diffraction was unsuitable, such as characterising the products of desolvation or where a crystallisation product was microcrystalline. In many circumstances, however, insufficient quantities of microcrystalline precipitate were produced for analysis by X-ray powder diffraction. Manual crystallisation screens are part of the studies presented in chapters 3, 4 and 5.

2.2.2 High throughput polymorph screening techniques

The number of experiments that can set up in a defined period of time, and the number of crystallisation products that can be analysed by a single scientist will practically limit the range of experiments in a manual crystallisation screen. The advantage of high through-put crystallisation methods is that a larger matrix of crystallisation conditions can be addressed in the same period of time. Such methods have been developed in the pharmaceutical industry on a micro scale, usually because in the development phase of a drug substance a very limited quantity is available for solid form screening. At the same time the screen must identify as many forms as possible in order to advance the optimal form to the clinical stage of development and as a consequence of this constraint on the quantity of crystallised sample, the 'gold standard' of phase identification, powder X-ray diffraction,[28] is unsuitable and Raman spectroscopy is more commonly employed. Upon identification of distinct forms, the crystallisation is scaled up from the micro scale to give a sufficient quantity of each form for full phase characterisation by X-ray powder diffraction, thermal methods and optical microscopy. Another consideration specific to automated screening is that the data generated from the increased matrix of crystallisations requires more powerful correlation and analysis techniques, usually requiring a custom designed computer database.[33;58] Several proof-of-concept studies have been published, applying high through-put screening to a small number of high profile APIs: paracetamol,[127] carbemazepine[33] and ritonavir.[57] Each of these screens found at least all of the known forms of the API targeted plus either difficult to

characterise forms (paracetamol) or new forms (carbemazepine and ritonavir). Chapter 6 details the application of a high through-put crystallisation screen, and the specific method is described therein.

2.2.3 Solvent evaporation

Crystallisation through solvent evaporation (EV) from a saturated solution is a common method used to grow crystals. A solvent containing the target compound in solution is placed in a crystallisation vessel and the solvent is allowed to evaporate steadily through the top of the open vessel. As solvent evaporates the solute is forced out of solution and crystallisation occurs (figure 2.1). The rate of evaporation of volatile solvents can controlled using a pierced lid to slow the rate of solvent evaporation and this control generally leads to fewer, larger crystals as the crystallisation product.
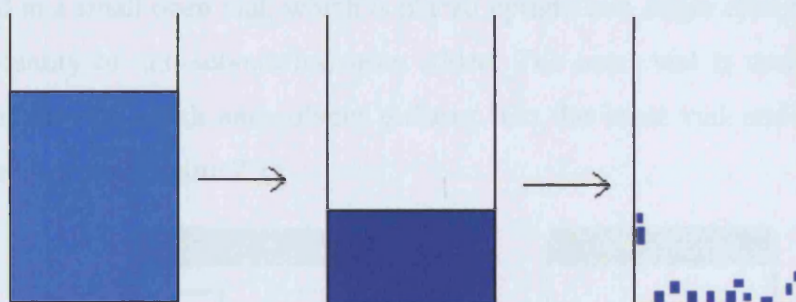
Figure 2.1: Solvent evaporation. Upon evaporation of solvent the solution becomes supersaturated, followed by crystallisation

Super-saturated solutions are more likely to lead to a kinetic crystallisation product – once a crystalline phase nucleates this phase will grow out rapidly relieving the super-saturation. Equilibration between solution and crystalline phase is not reached so there is less opportunity for phase changes to give successively more stable phases. Conversely sub-saturated solutions are more likely to lead to the thermodynamic crystalline form, as the slower crystallisation provides more opportunity to accommodate a series of phase changes *via* solution mediated phase tranformations.

Saturated solutions were made by standing solvent over an excess of the target compound for 24-48 hours followed by filtration. Sub-saturated solutions were made by diluting filtered saturated solutions with the appropriate amount of fresh solvent to give 25% and 50% saturated solutions. With all crystallisation methods the possibility of unintentional seeding, with either foreign particles or seeds of a known form of the

target molecule, must be addressed because of its potential to suppress polymorphism. In an attempt to remove all seeds, solutions were filtered through fine grained compacted Celite clay prior to use in crystallisations. In this work two different types of flat-bottomed straight-sided sample tubes were used for solvent evaporation. These sample tubes could be sealed with snap-on plastic lids. The large sample tubes (LCO) were 25 mm in diameter, 75 mm in height and could contain a maximum volume of 32.5 ml. The smaller sample tubes (SCO) were 15 mm in diameter, 50 mm in height and could contain up to 5 ml of solvent. All crystallisations were set up in virgin glassware to prevent possible cross-contamination that could occur with the use of washed vials.

2.2.4 Vapour diffusion

In crystallisation by vapour diffusion (VD) a saturated solution of the target compound is placed in a small open vial, which is placed upright in a larger closed vial in which a small quantity of anti-solvent has been added. The outer vial is sealed to produce a closed system in which anti-solvent diffuses into the inner vial and solvent diffuses from the inner vial (figure 2.2).
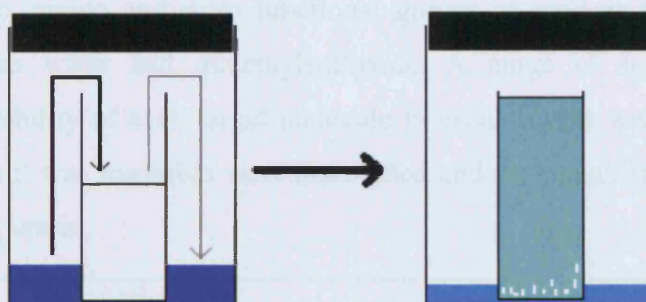
Figure 2.2: Vapour diffusion. Anti-solvent diffuses into the inner vial and solvent counter-diffuses out of the inner vial

The anti-solvent is required be more volatile than the solvent, to give a net increase in the amount of solvent in the inner vial. When the solvent/anti-solvent mixture in the inner vial reaches the point where it can no longer contain all of the solute, crystallisation occurs. The solvent must initially be saturated, otherwise the vapour diffusion may not induce crystallisation. If crystallisation occurs late in the diffusion process, it could be either solvent/solute or anti-solvent/solute interactions that are predominant, which could potentially influence the polymorph produced. An advantage

of this crystallisation method is that the crystals are grown in an increasing volume of solvent, usually yielding better quality crystals than solvent evaporation. Both the solvent and the anti-solvent can be varied, as long as they have the correct relative volatilities and miscibilities.

In this work vapour diffusions were carried out using SCO vials as the inner vial and LCO vials as the outer, closed vial. The choice of anti-solvent was dictated by initial solubility tests and solvent volatility considerations.

2.2.5 Solvent range

For each of the three manual screens a core range of 26 solvents were used (table 2.1). For each individual screen further solvents were also included, depending on their availability at the time of the screen. In total 51 different solvents were used in at least one manual screen. Of these 26 solvents 20 were included within the 67 solvents used in the high through-put screen described in chapter 6. This solvent range samples alcohols, saturated and aromatic hydrocarbons, chlorinated and fluorinated solvents, solvents containing oxygen in ketone, ether and ester functional groups, solvents containing nitrogen in cyano, amido and nitro functional groups as well as those with distinct properties such as water and dimethylsulfoxide. A range of solvent polarities is sampled. The solubility of each target molecule in each solvent was tested, and those solvents in which it was insoluble were discounted and the remaining solvents used in the crystallisation screen.

Solvents		
2,2,2-Trifluoroethanol	Chloroform	Methylbenzoate*
1,2-Dichloroethane	Cyclohexane	n-hexane
1,4-Dioxane	Dichloromethane*	Nitromethane
2-Butanol	Diethylether*	o-xylene
2-Chloroethanol*	Dimethylformamide	Tetrachloroethylene*
2-Propanol	Dimethylsulfoxide	Tetrahydrofuran
Acetone	Ethanol	Toluene
Acetonitrile	Ethylacetate	Water
Benzonitrile*	Methanol	

Table 2.1: The core range of 26 solvents used in all three manual crystallisation screens. * not included in the high through put screen solvent range

2.3 Analytical techniques

2.3.1 Single crystal X-ray diffraction

Single crystal X-ray diffraction (SXRD) data were collected on a Bruker AXS SMART APEX diffractometer with a charge coupled device (CCD) detector. The X-ray source was a molybdenum anode, Mo Kα λ = 0.71073 Å, with a graphite monochromator. The diffractometer was fitted with a Bruker AXS Kryoflex open flow cryostat. Unit cell determinations and data sets were collected at either 298 K or 150 K. The Bruker SMART[128] diffraction suite was used to control data collection. Data for unit cell determination were generated by taking 20, 30, or 50 frames with 5, 10 or 20 second exposures at three separate crystal orientations, scanning ω in 0.3° steps. A full sphere of data was collected for all crystal structure determinations. The data was integrated using the software package SAINT+[129] and a semi-empirical absorption correction applied using SADABS.[130] Structure solution by direct methods and subsequent structural refinement were carried out using the SHELXTL[131] suite. Full matrix least squares refinements were based on F^2. Friedel opposites were not merged for structures where the space group contained an inversion centre, but were merged for structures in non-centrosymmetric space groups (light-atom structures). All non-hydrogen atom positions were refined anisotropically. Hydrogen positions were usually found from the electron density map and freely refined. In some circumstances methyl hydrogens were placed on the carbon atom using a rigid rotor model and refined using a riding model. At the end of the refinement process the maximum shift/parameter ratio was less than 0.001. The final electron density map was usually featureless with a maximum electron density of less than 0.5 e Å^{-3}. Tables are provided at the end of each chapter summarising the crystal structure determinations presented in that chapter.

In this work, asymmetric unit figures were produced using the SHELXTL[131] package, with thermal ellipsoids for all non-hydrogen atoms plotted at the 50% probability level, and hydrogen atoms shown as spheres of arbitrary radius. Where hydrogen bonds (O-H···O, N-H···O, N-H···N) occur in the asymmetric unit they are shown as solid dashed lines. Figures showing crystal packing were produced using either Mercury[132] (wireframe and overlay figures) or CAMERON[133] (ball-and-stick figures). In all crystal packing figures hydrogen bonds are shown as dotted lines.

2.3.2 X-ray powder diffraction

X-ray powder diffraction (XRPD) data were collected on a Siemens D500 diffractometer, in a θ-2θ Bragg-Brentano geometry. The X-ray source was a copper anode, Cu Kα λ = 1.54056 Å, with a quartz pre-sample monochromator. Samples were mounted on a silicon monolith, cut slightly offset from the 111 plane to give zero background, in either a 0.3mm recess, or no recess. Samples were rotated in the X-ray beam to improve counting statistics. A typical scan measured from 5-40° in 2θ, with 0.05° stepping and 10 seconds exposure per step.

Two crystal structures determined from capillary XRPD data, by Dr. P. Fernandes and Dr. A. J. Florence (Strathclyde University), are presented in Chapter 6, and the general method used is described therein.

2.3.3 Simultaneous thermal analysis

Simultaneous thermal analysis (STA) comprising combined differential scanning calorimetry (DSC) and thermogravimetric analysis (TGA) was carried out using a Netzsch STA 449-C Jupiter instrument. Samples were placed in open or pierced-lid 10 μl aluminium pans and heated from room temperature until melting or decomposition, at 10 K min^{-1}. All samples were heated under a flow of nitrogen at 10 ml min^{-1}.

2.4 Crystal structure prediction method

2.4.1 The MOLPAK-DMAREL crystal structure prediction workflow

The MOLPAK[134]-DMAREL[135] crystal structure prediction method used in this thesis can be summarised in seven steps (scheme 2.1). This methodology refers to the case where at least one crystal structure is already available to use as a basis for the crystal structure prediction, but where no crystal structure is available energy minimisation of experimental crystal structures can be carried out upon their discovery.

CSP workflow	Programs
1. Energy minimisation of known crystal structure using experimental molecular conformation	DMAREL + GDMA
2. *Ab initio* optimisation of molecular conformation	Gaussian03
3. Energy minimisation of known crystal structure using *ab initio* molecular conformation	OptimalPaste DMAREL + GDMA
4. Generation and energy minimisation of possible crystal structures	MOLPAK DMAREL
5. Analysis to ensure all structures have energy minimised to a true minimum	Analysis DMAREL
6. Clustering of duplicate structures to give energy landscape of thermodynamically feasible structures	Cluster
7. Comparison of predicted structures with experimental structure(s)	Mercury PLUTO

Scheme 2.1: Crystal structure prediction workflow and programs used at each step

1. Energy minimisation of known crystal structure using experimental molecular conformation (ExptMinExpt)

Initially the experimental crystal structure, containing the experimental molecular conformation from this crystal structure, is lattice energy minimised using DMAREL (section 2.4.1).[135] Prior to energy minimisation, the bond lengths to hydrogen atoms in the crystal structure are standardised to counteract the foreshortening introduced during structure determination by single crystal X-ray diffraction. Bonds to hydrogen are

standardised to average values derived from neutron data[136]: C-H = 1.08 Å; N-H = 1.01 Å; O-H = 0.97 Å. The intermolecular potential used for this step comprises a distributed multipole electrostatic model (section 2.4.3), analysing the MP2/6-31G(d,p) charge density, unless otherwise specified, and the FIT dispersion-repulsion potential (section 2.4.4). The result of this step is the energy minimised structure using the experimental molecular conformation – the ExptMinExpt structure. This step tests the accuracy with which the intermolecular potential reproduces the experimental crystal structure.

2. Ab initio optimisation of molecular conformation

The second step is the *ab initio* optimisation of the molecular conformation from the experimental crystal structure to obtain the energy minimised gas phase molecular conformation. This molecular conformation is used as the search input because using an experimental molecular conformation from a known polymorph could unfairly bias the search results in favour of that polymorph to the detriment of other potential (or real) polymorphs. The optimisation is performed using the quantum mechanics package Gaussian03,[137] using the 6-31G(d,p) basis set and MP2 level calculations (section 2.4.5).

3. Energy minimisation of known crystal structure using ab initio molecular conformation (ExptMinOpt)

The third step is closely analogous to the first step, but with *ab initio* optimised molecular conformation pasted into the experimental crystal structure in place of the experimental molecular conformation. The computer program OptimalPaste[138] substitutes the *ab initio* optimised conformation for the experimental conformation in the crystal structure by minimising the root mean square discrepancy between all corresponding non-hydrogen atomic positions of one molecule. This crystal structure is energy minimised to give the ExptMinOpt structure. Due to the use of the optimised molecular conformation, the reproduction of the experimental crystal structure in this step is generally inferior to that of the first step as it ignores any genuine effect of crystal packing on the molecular conformation. However the difference should not be large if the rigid body approximation is appropriate, and the reproduction should be within the usual tolerance of under 5% error in the cell parameters. This step tests the

ability of the potential to reproduce the crystal structure, using the *ab initio* optimised molecular structure. It allows direct comparison of the resultant ExptMinOpt version of the experimental structure with the hypothetical crystal structures generated in step 4. The ExptMinOpt structure is the closest that any predicted structure can come to reproducing the experimental structure. It should be noted that while MOLPAK limits the CSP procedure to only generating structures with $Z' = 1$, DMAREL does not have this limitation and experimental structures can be lattice energy minimised regardless of the number of symmetry independent molecules.

4. Generation and energy minimisation of possible crystal structures

The fourth step is the computational crystal structure search procedure, using the program MOLPAK (section 2.4.6).[134] This program uses predefined packing types in a range of the most common space groups, but with a $Z' = 1$ limitation, to generate hypothetical crystal structures for organic molecules. The *ab initio* optimised molecular conformation is input and different packing arrangements are used to define the relative positions of the molecules. A systematic search places different molecular orientations in these packing arrangements to generate dense crystal structures by condensation using a pseudo-hard sphere repulsion potential. This produces the initial hypothetical crystal structures. For each packing type up to 200 of the most dense structures are passed to DMAREL for full energy minimisation, with its more sophisticated intermolecular potential.

5. Analysis to ensure all structures have energy minimised to a true minimum

The fifth step is the analysis of the hypothetical crystal structures generated to ensure that all structures have reached a true energy minimum. Any structure that energy minimises to a saddle point rather than a true minimum in its initial space group is re-minimised without the symmetry element that led to the saddle point, producing a structure in a sub-group of the original space group with two molecules in the asymmetric unit. This procedure is carried out by the program Analysis[138] and it is usually defined that up to two symmetry operators can be removed serially for each saddle point structure. This leads to a small number of $Z' = 2$ structures being present in

the search results. Reduced cell parameters[139] are then calculated for each structure and they are ranked according to lattice energy.

6. Clustering of duplicate structures to give energy landscape of thermodynamically feasible structures

The sixth step is to cluster duplicate predicted structures – each minimum on the crystal packing surface may be reached by DMAREL minimisation of more than one of the MOLPAK generated structures, and these duplicate structures must be discarded to leave only the lowest energy example of each of these clusters. The program Cluster[138] uses both peak profile matching of simulated powder patterns[140] and the co-ordination environment[124] around a central molecule in each structure to find hypothetical crystal structures that are effectively equivalent. The co-ordination environment comparison commonly used the 18 molecules with centres of mass closest to the central molecule. All of the distinct structures are ranked according to their lattice energy and volume and the dense, low energy structures are regarded as those most likely to correspond to experimental polymorphs. These structures are usually plotted on a scatter graph of lattice energy against cell volume per molecule (density) to give the crystal energy landscape.

7. Comparison of low energy structures with experimental structure

The accuracy with which the experimental crystal structure is reproduced by the ExptMinExpt and ExptMinOpt energy minimisations can be quantified by calculation of the F-value, which takes into account the differences introduced in the cell parameters, and molecular positions and orientations (section 2.4.7).

The ExptMinOpt energy minimised crystal structure is the only version of the experimental crystal structure that can be directly compared with the hypothetical structures generated by the CSP search because it has been energy minimised using the same intermolecular potential and with the same molecular conformation as that used in the search. The ExptMinOpt structure can be compared to computed structures using the visualisation program Mercury,[132] which has a structure overlay function. Graph sets for all low energy generated structures and discovered polymorphs were calculated using PLUTO.[141]

2.4.2 Energy minimisation by DMAREL[135]

DMAREL is used for energy minimisation of both MOLPAK-generated and experimental crystal structures, using the FIT dispersion-repulsion potential and electrostatic distributed multipole analysis of the charge density calculated by Gaussian. For a crystal structure the parameters that are varied in the process of arriving at an energy minimum are three translational and three rotational components for each molecule in the unit cell plus six strain vectors to describe the changes in the cell dimensions. The energy is minimised as a function of these 6Z+6 where Z is the number of molecules in the unit cell.[135] No kinetic energy factors are calculated such as zero-point energy or lattice vibrational effects, meaning that the energy surface upon which the energy of the crystal structure is calculated is a '0 K surface with neglect of zero-point energy'.[142]

An iterative procedure using a modified Newton-Raphson method is used for the energy minimisation. An unmodified Newton-Raphson calculation has the form:

$$\delta_i = -\sum_j^N (W^{-1})_{ij} G_j$$

where δ_i is the step direction, G is the initial position first derivative vector and W is the second derivative matrix (Hessian).[143] The second derivative matrix, the second partial derivative of the internal energy with respect to each pair of variables, defines whether the point is a saddle point or a true minimum:

$$W_{ij} = \frac{\partial^2 U}{\partial q_i \partial q_j}$$

where q_i and q_j are variables that the minimisation is being carried out with respect to.[70]

Each iteration of the minimisation includes a choice of the step direction and the size of that step. The step direction is provided by the Newton-Raphson equation: G is the gradient vector that includes the partial derivative of the lattice energy with respect to all coordinates of a point and gives the direction of greatest downward slope.[144] A series of line searches are carried out to find the optimal step size in the direction δ_i, though a user-defined maximum step size is imposed to prevent the step from taking the

minimisation beyond the region in which the second order Taylor expansion, upon which the Newton-Raphson method is based, is valid.[145]

Energy minimisations are carried out with space group constraints imposed. If the eigenvalues calculated from the components of the Hessian matrix are all positive then the point is a true minimum. If one of the eigenvalues is negative then the point is a transition state, a saddle point, and the eigenvector corresponding to this eigenvalue will lead the minimisation procedure to a true minimum.[70] In the case of a saddle point minimisation result a lower symmetry sub-group will have a lower energy, and the minimisation can be repeated with the symmetry operator corresponding to the most negative eigenvalue removed. This will increase the number of molecules in the asymmetric unit by a factor of two for each symmetry operator removed.

The atom-atom approximation is used to sum the interaction energy for each pair of molecules and pairwise additivity approximation is used to sum of these two molecule interactions over all of the molecules in the crystal. The lattice energy could be summed for each molecule with every other molecule in the crystal structure, however this would be computationally expensive, and therefore summation limits are employed. Charge-charge (R^{-1}), charge-dipole (R^{-2}) and dipole-dipole (R^{-3}) electrostatic terms converge slowly and a direct cutoff would not give reliable result. Consequently for these terms the Ewald accelerated summation technique is used.[146;147] The higher multipole terms of the electrostatic interactions converge more rapidly and are cut off using a direct cutoff at 15 Å. The 15 Å limit refers to the maximum distance between the centres of mass of two molecules and all of the site-site interactions in molecules within this distance are calculated, regardless of the actual site-site separation.[135] Dispersion-repulsion terms are summed with a direct cutoff of 15 Å between atoms.

2.4.3 Calculation of distributed multipoles using GDMA[88]

GDMA derives a series of point multipoles to represent the charge distribution around the molecule, from the wavefunction calculated using *ab initio* methods. Atomic multipoles are particularly useful for the modelling of electrostatic effects such as hydrogen bonding and π–π interactions,[89] producing a more accurate electrostatic model than potential derived nuclear point charges for a slight increase in computational expense.

GDMA[88] calculates the distributed multipoles directly from the wavefunction of the molecule – with only the exponents of the gaussian functions in the basis set used and the nuclear charges and positions required. The products of the primitive gaussian functions that comprise the basis set are calculated, with the product of two gaussians a further gaussian function for the overlap of the two functions. The charge density, $\Psi^*\Psi$, can be expressed in terms of these gaussian products. A multipole expansion about the point of overlap of two gaussians can be calculated exactly and then moved to a nearest chosen site and incorporated into a multipole expansion at this site.[88] The distributed multipoles are calculated at nuclear sites and are calculated up to hexadecapole. Nuclear sites are generally sufficient to ensure good convergence of the electrostatic potential and interaction energies, as calculated from the distributed multipole analysis.

Multipole expansions do not give a correct potential inside the charge distribution, because they neglect penetration effects.[88] They also exhibit basis set dependence: a particular part of a wavefunction may be described by basis functions on different atoms between different basis sets, leading to differences in how the electrostatic potential is modelled using multipoles.[88] However this is not seen as a particularly important effect because outside the molecule the electrostatic potential should only vary with the quality of the wavefunction used, provided the expansion is converged.

2.4.4 The FIT dispersion-repulsion potential

The dispersion and repulsion terms for each atomic type are modelled together, using an empirical method and, unlike the electrostatic model which is not transferable and must be generated specifically for each different molecule, they are assumed to be transferable for each atomic type. The Buckingham potential form, used by DMAREL to model the dispersion and repulsion contributions to the intermolecular potential, combines an exponential-based repulsive term with the London dispersion term.

During the 1970s and 1980s, Williams and co-workers produced a consistent set of values for the homoatomic Buckingham parameters for a number of elements, deriving the Buckingham parameters for each by optimisation against a range of experimental crystal structures. The isotropic atom-atom approximation was employed, where the molecule is assumed to be an overlay of spherical atoms,[148] and the

interactions between two molecules were assumed to be equal to the sum of the interactions between each atom in the first molecule with each of the atoms in the second molecule. This allowed the parameters for each atomic type to be transferable. Optimised parameters were generated for carbon,[77;78] hydrogen,[77;78] oxygen,[79] nitrogen,[78] fluorine[80] and chlorine.[81] Their method employed the general expression for the intermolecular potential:

$$E_{ik} = A_{\iota\kappa} \exp(-B_{\iota\kappa}r_{ik}) - C_{\iota\kappa}r_{ik}^{-6} + q_i q_k r_{ik}^{-1}$$

where i is an atom of type ι on one molecule, k is an atom of type κ on a second molecule and r is the distance between atoms i and k and where the three parameters A, B, and C were those to be derived for each atomic type. The modelling of the electrostatic term for each molecule used nuclear-site, extended-site or bond-site potential derived point charges. The extended- and bond- site models were used when the nuclear site point potential derived point charge model did not give a satisfactory representation of the electrostatic potential. It should be noted that while the functional form of the Buckingham potential principally models the dispersion and repulsion, the empirical fitting partially absorbs other effects (such as exchange and polarisation) and differences between the true charge density and the electrostatic model.

For each atomic type a range of crystal structures was chosen as a training set against which the Buckingham potential parameters were optimised for that atomic type. Initially parameters for carbon and hydrogen were developed,[77] by optimising the A and C parameters, while fixing the B parameter for both atomic types to values obtained from previous work using different methods. Once C and H parameters were derived they were held constant for optimisation of the parameters for oxygen,[79] nitrogen,[78] fluorine[80] and chlorine.[81] The carbon and hydrogen A and C parameters were slightly reparameterised at the same time as the nitrogen parameters were derived.[78] For each atomic type the homoatomic parameters were calculated and the heteroatomic cross terms were calculated using the well-known geometric combining rules (section 1.8.3). Uncertainty in the validity of the calculated heteroatomic values led, in two chapters, to further validation for specific cases – for fluorine (chapter 3) and for the interactions of water with organic molecules (chapter 7).

The structures in the training sets were chosen so that they did not include hydrogen bonding, to minimise the effect of the electrostatic term and maximise the effect of the dispersion-repulsion terms on the experimental crystal structures. Carbon and hydrogen were originally parameterised against the crystal structures of a small number of hydrocarbons, and then slightly reparameterised during the nitrogen parameter development; parameters for oxygen were found from the crystal structures of β-oxygen, carbon dioxide, trioxane, tetraoxocane and succinic anhydride; nitrogen parameters were developed against α-nitrogen and aza- and cyano- containing hydrocarbons and fluorine against a small range of perfluorocarbon crystal structures. Williams only produced terms for non-polar hydrogen atoms attached to carbon (and hereafter denoted non-polar, H_{np}). A separate model was desirable for polar hydrogen atoms, those attached to oxygen and nitrogen. Coombes *et al.*[82] further optimised Williams' hydrogen parameters against the crystal structures of a range of N-H containing molecules to produce optimised parameters for polar hydrogens (H_p). An SCF DMA electrostatic model was used. Williams' optimised parameters for carbon, non-polar hydrogen, nitrogen, oxygen, along with Coombes' polar hydrogen parameters, are used with the Buckingham potential as the dispersion-repulsion model for DMAREL, and known as the FIT potential. These parameterisations were each based on a limited number of crystal structures and sample each atomic type in a limited number of environments. The environment around a particular atom in a particular crystal structure may be very different to the average environment sampled during the empirical parameterisation for that atomic type, potentially causing this particular atom to be poorly represented by the dispersion-repulsion parameters optimised for its generic atomic type.

2.4.5 Calculation of molecular wavefunctions and *ab initio* optimisations of molecular structures

The program Gaussian (98 or 03)[137] is used for two tasks in the CSP method: it is used to calculate the wavefunction and charge density of the molecule prior to calculation of the distributed multipoles, and it is used to produce the *ab initio* gas phase optimised molecular conformation, usually starting from the molecular conformation from a known crystal structure, by minimising the total energy of the conformation.

In this thesis both tasks were always carried out using the 6-31G(d,p) basis set and using, for the most part, MP2 (Møller-Plesset 2nd order) level calculations. This is a split-valence double-zeta basis set, with two basis functions representing each valence orbital, but only a single basis function represents each core orbital. The 6-31G(d,p) basis sets represents the core 1s atomic orbitals of second row atoms with a contracted basis function made up of a fixed linear combination of 6 primitive Gaussian functions. The valence 2s and 2p atomic orbitals are represented by two basis functions – the inner part of each valence orbital is represented by a contraction of 3 primitive Gaussian functions and the outer part of the orbital by a single Gaussian function. This basis set includes d-type polarisation on non-hydrogen atoms and p-type polarisation on hydrogen atoms.[145]

MP2 calculations provide a correction to the corresponding Hartree-Fock self-consistent field (HF-SCF) calculation to account for electron correlation effects. The MP2 level calculation provides an extra term to the Hartree-Fock term to better approach the true electronic energy.[144]

2.4.6 Crystal structure generation by MOLPAK[134]

The MOLPAK method for generation of crystal structures is based on the observation by Kitaigorodsky[149] that all organic molecules show close packing in their crystal structures. He defined the packing coefficient as the ratio of the molecular volume to the unit cell volume per molecule.[149] It has been noted that virtually all crystal structures of organic molecules have a packing coefficient between 0.65 and 0.75, close to the hard sphere packing coefficient of 0.74.[150]

MOLPAK uses a rigid molecular conformation, most usually the *ab initio* optimised conformation, as the input for the generation of hypothetical crystal structures. This is known as the rigid body approximation, and while it is not significant for molecules that do not have degrees of torsional freedom, it is a limitation for flexible molecules, where a series of molecular conformations must be used as the input for separate MOLPAK searches to compensate.[116] MOLPAK is also limited to producing structures with only one molecule in the asymmetric unit. The symmetry operations used to generate hypothetical crystal structures are: inversion through a point; mirror plane; glide plane; two-fold rotation axis and two-fold screw axis. A central molecule is

taken and the application of the symmetry operations of a space group is used, along with a simple repulsion potential, to generate a crystal structure. Most of the space groups employed in MOLPAK can generate more than one arrangement of molecules in the coordination sphere around a single molecule, and for these space groups more than one packing type is defined. The process for building the crystal structure in a packing type can be summarised:

- 1. A molecule is placed in an orthogonal axis system and a second molecule is placed on axis-1 an arbitrary distance from the central molecule. The distance between the central molecule and this neighbour is reduced until the repulsion potential dictates that the molecules are as close as they can be without incurring significant repulsion. This intermolecular separation is applied to produce a row of translationally related molecules.

- 2. A second row, identical to the first, is placed in the axis-1/axis-2 plane parallel to the first row, and moved towards it until a similar repulsion criterion is met. In the case of packing types in triclinic and monoclinic space groups, the two rows are moved parallel to one another (parallel to axis-1) until the repulsion is minimised. For orthorhombic space groups the relation of the first row to the second is fixed by the space group and only the separation of the two rows is determined by the repulsion potential. A plane is produced from this condensation of rows of molecules.

- 3. Step 3 is the three-dimensional analogue of the first two steps: a second plane is generated parallel to the first plane and moved towards the first plane, in the direction of axis-3 until the repulsion criterion is again met. For triclinic space groups the relative positioning of the planes with respect to axis-1 and axis-2 is varied until the repulsion is minimised. In this case for both monoclinic and orthorhombic space groups the relative positioning of the two planes is dictated by the space group symmetry and only the separation of the planes need be determined.

The intermolecular potential used is a pseudo-hard sphere potential using the atom-atom pairwise approximation. The level of restriction of positioning of packing units (for example 2 generated planes) depends on the number of symmetry operators in the space group. This was explained by Holden and Ammon[134] "For monoclinic and orthorhombic space groups, the positions of the molecules of the 2D grid, relative to its intersection with axis-3 [adjacent 2D grids] are restricted, and procedure 3 determines only the length of a unit cell axis. However for the two triclinic space groups the grid must be shifted in both the axis-1 and axis-2 directions to determine two angles in addition to the length of an axis." Hence if the packing units are not restricted by symmetry operations, then their positioning has one (step 2) or two (step 3) extra degrees of freedom associated and this positioning is dictated solely by the repulsion potential.

Different orientations of the molecules with respect to the global co-ordinate system are systematically sampled. The rigid molecular probe is used in different orientations in different iterations in each packing type, by varying the Euler angles in 10° steps over a 180° range around each of the three axes. 5 to 10 of the most densely packed volume arrangements are subjected to a more detailed search over a ±10° Eulerian range, with a 2° step size.

Two different MOLPAK searches were used in this thesis: the standard search used 37 packing types covering the most common 18 space groups with the 125 most dense structures from each packing type passed to DMAREL for further energy minimisation; the extended search used 47 packing types covering the most common 22 space groups and passed the 200 most dense structures in each packing type to DMAREL.

2.4.7 Comparison of experimental and energy minimised structures

All polymorphs of the target compound known prior to a crystal structure prediction search, or found during subsequent crystallisation screening, had to be energy minimised using the same potential and molecular conformation as that used in the search, to give the ExptMinOpt energy minimised structure, as this allows direct comparison between the predicted structures and the ExptMinOpt energy minimised experimental structure.

The reproduction of the experimental crystal structure by its ExptMinExpt or ExptMinOpt energy minimised structures is quantified by the calculation of the F-value:[151]

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

where $\Delta\theta$ is the rotational displacement of the rigid molecule after minimisation (°), Δx is the translational displacement (Å) and the remaining terms are the changes in the cell parameters (Å and °). The multiplication factors in the first five terms bring the errors onto a comparable scale.[152] This so-called 'figure-of-shame' quantifies the accuracy of the reproduction of the crystal structure by the intermolecular potential and the approximations applied in the method. For comparisons between experimental and ExptMinOpt structures this also includes variation due to the modified molecular conformation used. As a guide, an ExptMinOpt structure with an F-value under approximately 50 and with all cell parameters reproduced to within 5% is viewed as a successful energy minimisation.

Qualitative comparison of the ExptMinOpt version of the experimental structure with the experimental structure or with structures generated by the CSP search can be made using the root mean square overlay function in Mercury,[132] which allows the user to define the atoms or molecules upon which to match the two structures. PLUTO[141] can be used to calculate graph sets for both experimental and hypothetical crystal structures, and the user can define which hydrogen bond types are included in the calculation and the distance and angle criteria for these hydrogen bonds.

2.5 Supporting information

The compact disc that accompanies this thesis contains all supporting information. This includes detailed results for all experimental crystallisation screens, summaries of all crystal structure predictions carried out, and crystallographic information files (.cif) and a summary of crystallisation conditions for all newly determined crystal structures.

Chapter 3 5-Fluorouracil

3.1 Introduction

3.1.1 Historical and medicinal background

5-Fluorouracil (5-fluoropyrimidine-2,4-dione, 5FU, figure 3.1) was originally synthesised in 1957[153] specifically because it was thought to be a potential anti-cancer agent, based on observations that cancerous tissue incorporates the nucleobase uracil faster than non-malignant tissue and it was thought that the fluorinated analogue could have an inhibitory effect.[154] Preliminary studies in rats and mice proved this supposition correct.[155] 5FU has been shown to have both DNA and RNA modes of action.[156] In DNA it inhibits the thymidylate enzyme that readies thymine for DNA incorporation, and can be incorporated itself into DNA in place of thymine. In RNA, 5FU substitution for uracil leads to both enzyme inhibition and changes in protein expression. In the intervening 50 years since its discovery 5FU has become widely used to treat a range of solid tumours,[157] usually in combination therapies with other anti-cancer pharmaceuticals or ionising radiation.[156] As a side-effect 5FU is known to be a powerful immunosuppressant.[158]
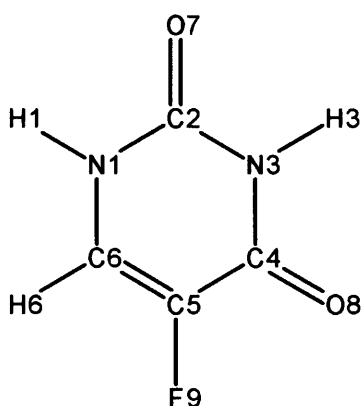
Figure 3.1: Chemical structure of 5-fluorouracil (5FU) with its crystallographic numbering scheme

3.1.2 Previously published crystal structures

The CSD contains one crystal structure of 5FU (FURACL[159]) hereafter referred to as form 1. Form 1 crystallises in the triclinic space group $P\bar{1}$, and unusually it contains

four molecules in the asymmetric unit. An analysis[160] of Z' > 1 structures in the CSD in 2001 found that only 600 structures were reported to have Z' = 4, though this may be an under-representation of its true occurrence in nature. The unit cell of form 1 was originally reported in an unconventional setting with β = 43.9(3)°. The unit cell and the conventional setting are summarised in table 3.1.

	5-Fluorouracil Form 1[159]	Conventional unit cell
Empirical formula	$C_4H_3N_2O_2F$	
Crystal system	Triclinic	Triclinic
Space group	$P\bar{1}$	$P\bar{1}$
a, b, c (Å)	9.22(3), 12.66(3), 12.67(3)	8.79(5), 9.22(3), 12.66(3)
α, β, γ (°)	89.7(3), 43.9(3), 98.6(3)	98.6(3), 99.5(5), 89.4(4)
V (Å³)	998	998
Z', Z	4, 8	4, 8

Table 3.1: Unit cell parameters for the published structure of 5FU form 1. The conventional unit cell was calculated using PLATON[139]

The form 1 crystal structure is also unusual in that it contains four fluorine atoms in close proximity forming a tetramer (figure 3.2). Each of the four close F···F intermolecular distances are less than 3.2 Å, which when compared to the van der Waals radius of fluorine 1.47 Å,[97] suggests that the fluorine atoms are almost in van der Waals contact. All four crystallographically independent molecules are present in the tetramer unit. Interactions involving fluorine, such as F···F, C-H···F and C-F···π are not commonly used as structure directing synthons in crystal engineering as they are very weak, and are easily overwhelmed by stronger interactions, though it has been noted that these interactions can act in a stabilising supporting role.[161;162] In the crystal structure each 5FU molecule forms a $R_2^2(8)$ hydrogen bonded ring comprised of two N-H···O hydrogen bonds and two single N-H···O hydrogen bonds. These hydrogen bonds produce a hydrogen bonded sheet with no strong interactions between adjacent sheets. Each sheet is comprised of three different hydrogen bonded rings: the $R_2^2(8)$ association, a $R_4^4(16)$ ring and a $R_8^8(40)$ ring with the F···F tetramer at its centre (figure 3.2).

$R_4^4(16)$

$R_8^8(40)$

F29 F19

F9 F39
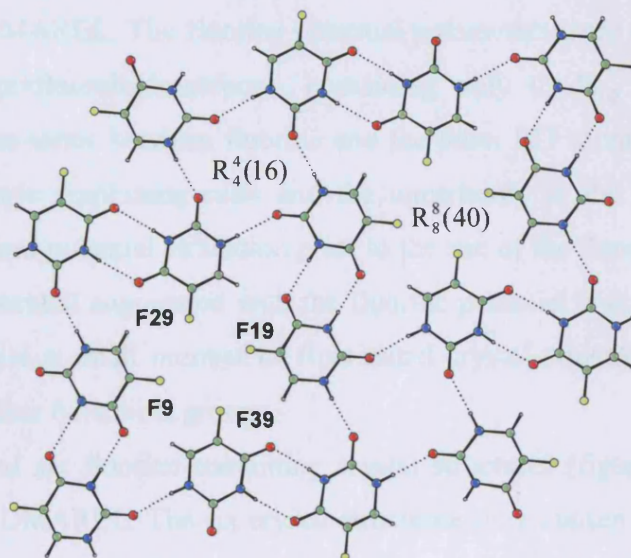
Figure 3.2: Hydrogen bonded sheet present in 5FU form 1. The F···F tetramer and the three different hydrogen bonded rings can be observed

The crystal structures of several co-crystals containing 5FU have also been published; with cytosine as a monohydrate,[163] in a 2:1 ratio with theophylline as a monohydrate[164] (both co-crystal solvates), with 1-methylcytosine,[165] and with 9-ethylhypoxanthine.[166]

It was postulated, in light of the unusual features exhibited by the 5FU form 1 crystal structure, that there may be a more stable and 'conventional' polymorph awaiting discovery, possibly with only one independent molecule in the asymmetric unit in a higher symmetry space group without close contacts between the fluorine atoms. To investigate this proposition, both computational crystal structure prediction and an experimental crystallisation screen were carried out.

3.2 Crystal structure prediction

3.2.1 Validation of the fluorine dispersion-repulsion potential

The FIT dispersion-repulsion parameters for C, H_{np}, H_p, N, O have been used extensively in the past to energy minimise crystal structures and, within the inherent limitations of their empirical derivation and the Buckingham potential form, have been found to be fit for purpose. The fluorine dispersion-repulsion potential was derived by Williams[80] and co-workers at the same time as the other atom types, but prior to this

study had not been used in the energy minimisation of any fluorine-containing crystal structures with DMAREL. The fluorine potential parameters were originally fitted to a training set of perfluorohydrocarbons, containing only C, H_{np} and F atoms. The heteroatomic cross-terms between fluorine and the other FIT atom types were derived using the geometric combining rules and the uncertainty in the performance of the cross-terms required potential validation prior to the use of the fluorine potential during CSP. The FIT potential augmented with the fluorine potential was tested for its ability to energy minimise a small number of fluorinated crystal structures containing other atom types and other functional groups.

A series of six fluorine-containing crystal structures (figure 3.3) were energy minimised using DMAREL. The six crystal structures were chosen from the CSD, with the atom types limited to C, H, N, O and F, with only one molecule in the asymmetric unit, and with all hydrogen atom positions determined in the crystal structure. Four of the molecules in the test set included other atom types beyond C, H and F, and were capable of hydrogen bonding, hence providing a more demanding test of the fluorine potential's ability to reproduce heteroatomic interactions. None of the molecules in the test set included flexible torsion angles, and it was expected that the *ab initio* optimisation of their molecular conformations would not lead to significant changes compared to the experimental conformations for any of the molecules. Each crystal structure was energy minimised using both the experimental molecular conformation (ExptMinExpt) and the *ab initio* optimised molecular conformation (ExptMinOpt), and then compared to the corresponding experimental structure (table 3.2). For all molecules the electrostatic potential was modelled using a distributed multipole analysis of the MP2/6-31G(d,p) wavefunction, and the molecular conformations were optimised using the same level of theory.
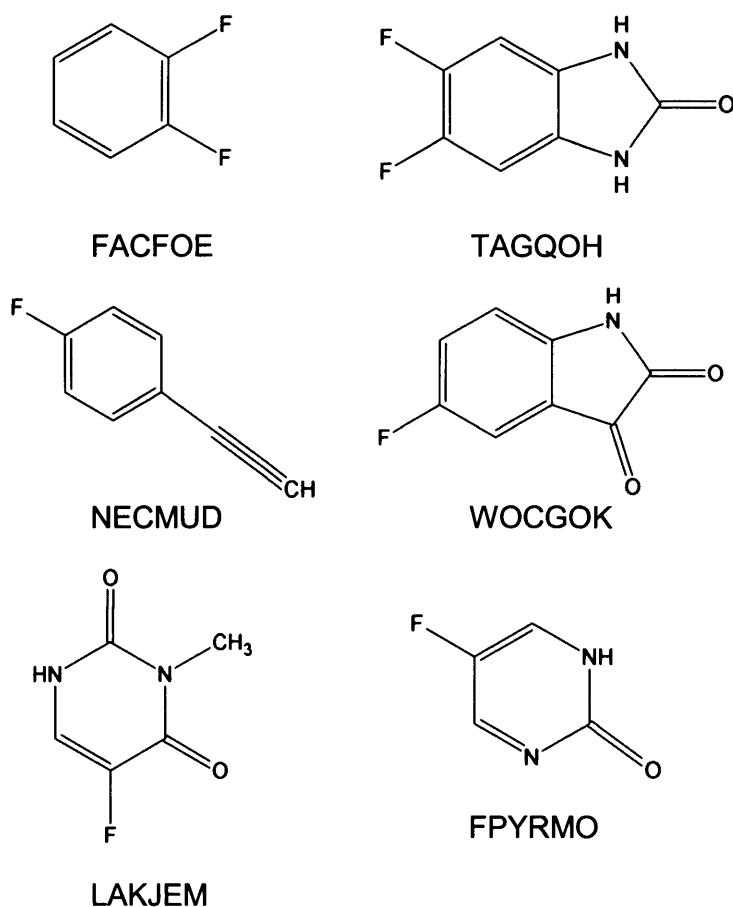
Figure 3.3: Molecules used to test the fluorine potential. FACFOE
(1,2-difluorobenzene); NECMUD (4-fluoroethynylbenzene); LAKJEM
(3-methyl-5-fluorouracil); TAGQOH (5,6-difluoro-2-benzimidazolone);
WOCGOK (5-fluoroindoline-2,3-dione); FPYRMO (5-fluoropyrimidine-2-one)

FACFOE	Experimental	ExptMinExpt	% error	ExptMinOpt	% error
Space group	$P2_1/n$				
a (Å)	7.481	7.737	3.42	7.706	3.02
b (Å)	5.961	6.063	1.72	6.072	1.86
c (Å)	11.725	11.600	-1.06	11.701	-0.20
β (°)	103.815	102.644	-1.13	102.316	-1.44
Vol. (Å³)	507.70	530.95	4.58	534.92	5.36
Density (g/cm³)	1.493	1.427	-4.38	1.417	-5.09
F		25.30		22.41	
FPYRMO					
Space group	$P2_1/c$				
a (Å)	11.118	11.198	0.72	11.305	1.68
b (Å)	6.012	6.175	2.72	6.225	3.55
c (Å)	7.147	7.232	1.19	7.243	1.34
β (°)	104.310	107.156	2.73	107.047	2.62
Vol. (Å³)	462.89	477.85	3.23	487.33	5.28
Density (g/cm³)	1.637	1.586	-3.13	1.555	-5.01
F		32.050		34.29	

LAKJEM	Experimental	ExptMinExpt	% error	ExptMinOpt	% error
Space group	$P2_1/n$				
a (Å)	4.67	4.54	-2.79	4.681	0.23
b (Å)	11.439	11.249	-1.66	11.174	-2.31
c (Å)	11.507	12.115	5.29	11.999	4.27
β (°)	101.08	101.047	-0.03	103.07	1.97
Vol. (Å³)	603.25	607.25	0.66	611.36	1.34
Density (g/cm³)	1.587	1.576	-0.66	1.566	-1.33
F		59.46		41.22	
NECMUD					
Space group	$P2_1/c$				
a (Å)	7.065	7.216	2.14	7.294	3.24
b (Å)	6.587	6.794	3.15	6.766	2.71
c (Å)	13.241	13.286	0.34	13.376	1.02
β (°)	99.640	100.931	1.30	102.061	2.43
Vol. (Å³)	607.50	639.56	5.28	645.50	6.26
Density (g/cm³)	1.313	1.248	-5.01	1.236	-5.89
F		18.204		27.78	
TAGQOH					
Space group	$P2_1/c$				
a (Å)	8.232	8.080	-1.85	8.098	-1.63
b (Å)	7.341	7.387	0.64	7.522	2.48
c (Å)	11.309	11.639	2.92	11.649	3.01
β (°)	90.530	91.684	1.27	91.681	1.27
Vol. (Å³)	683.31	694.37	1.62	709.29	3.8
Density (g/cm³)	1.654	1.627	-1.59	1.593	-3.66
F		20.974		27.15	
WOCGOK					
Space group	$P2_1/c$				
a (Å)	3.789	3.940	3.98	3.953	4.32
b (Å)	12.200	12.343	1.17	12.434	1.92
c (Å)	14.99	14.380	-4.07	14.381	-4.07
β (°)	94.41	89.338	-5.37	89.282	-5.43
Vol. (Å³)	690.87	699.21	1.21	706.71	2.29
Density (g/cm³)	1.588	1.569	-1.19	1.552	-2.24
F		89.260		96.36	

Table 3.2: Summary of energy minimisations from fluorine potential testing

All six of the structures were successfully reproduced by the FIT potential augmented with the fluorine homo- and hetero- atomic terms. The largest error in reproduction of a cell parameter in the energy minimisation of the six structures was a -5.42% error in the β angle in the ExptMinOpt minimisation of 5-fluoroindoline-2,3-dione (WOCGOK), and this structure proved to be the most poorly reproduced structure. The F-value for the ExptMinExpt minimisation was 89 and this rose to 96 when the optimised conformation was used in the minimisation. Qualitatively an overlay of the experimental and

ExptMinOpt structures showed that the structure had not significantly changed during the minimisation procedure. The structure of 3-methyl-5-fluorouracil also had an F value greater than 50, though only for the ExptMinExpt reproduction. The reproduction of this structure is of particular interest, as it is obviously closely related to 5-fluorouracil, though the methyl group on N1, rather than a hydrogen atom leads to very different hydrogen bonding. An overlay of the experimental and ExptMinOpt structures showed that the structure had not been significantly altered. The other four structures in the test set were well reproduced.

While this six molecule test set may not explore all possible fluorine hetero-interactions thoroughly, the success in their reproduction does indicate that the fluorine potential is as reliable as can be expected in light of the limitations of the empirical method, and is suitable for use in the crystal structure prediction of 5FU.

In the course of this study a new polymorph was discovered (form 2) and both polymorphs were energy minimised, further testing the potential.

3.2.2 Energy minimisation of 5-fluorouracil polymorphs

In the course of the experimental screen, the crystal structure of 5FU form 1 was redetermined at 150 K. This structure was significantly more accurate than the literature structure[159] and a superior starting point for energy minimisation. The experimental crystallisation screen also yielded a second polymorph of 5FU and the crystal structures of both forms as determined at 150 K were energy minimised using both the experimental and *ab initio* optimised molecular conformations.

The minimisation of 5FU form 1 proved successful, with the largest ExptMinOpt cell parameter error only 3.72%. The ExptMinOpt minimisation of form 2 showed an error of 6.09% in the *a* axial length, greater than the maximum 5% error normally assumed to be attributable to thermal effects, but upon visual inspection it was evident that the structure was not significantly altered by the energy minimisation process, with the hydrogen bonded ribbons still intact in the energy minimised structure. ExptMinExpt and ExptMinOpt energy minimisations of both structures are summarised in table 3.3.

5-Fluorouracil	Form 1					Form 2				
	Expt.*	ExptMinExpt	% error	ExptMinOpt	% error	Expt.	ExptMinExpt	% error	ExptMinOpt	% error
a (Å)	8.6329	8.729	1.11	8.845	2.46	5.043	5.244	3.98	5.351	6.09
b (Å)	9.156	9.329	1.89	9.274	1.29	14.935	15.245	2.08	15.262	2.18
c (Å)	12.5796	12.972	3.12	13.047	3.72	6.605	6.506	-1.49	6.507	-1.47
α (°)	99.12	97.38	-1.76	96.62	-2.53	-	-		-	
β (°)	100.02	98.10	-1.92	97.14	-2.88	108.88	109.84	0.88	110.29	1.29
γ (°)	90.02	91.30	1.42	91.69	1.86	-	-		-	
Volume (Å³)	966.4	1036.3	7.23	1053.8	9.05	470.7	489.3	3.94	498.4	5.89
Density (g cm⁻³)	1.788	1.667	-6.75	1.640	-8.3	1.835	1.766	-3.79	1.733	-5.56
Final Energy (kJ mol⁻¹)		-102.98		-96.58			-109.29		-102.49	
F		36.26		62.63			28.96		53.00	

Table 3.3: Energy minimisation of the crystal structures of 5FU. * Low temperature redetermination obtained during this study

3.2.3 Crystal structure prediction – results

The MP2/6-31G(d,p) *ab initio* optimised molecular conformation of 5FU was used as the imput for the MOLPAK search. The search generated structures in 37 MOLPAK packing types spanning 18 different space groups. The programs Analysis[138] and Cluster[138] were not used in this study, so all structures that minimised to saddle points were discarded and the resultant list was clustered by hand by comparing reduced cells, with the lowest energy member of each cluster retained. All remaining structures within 8 kJ mol⁻¹ of the global energy minimum structure were examined visually and by graph set analysis and similar structures that were not close enough to cluster by comparing reduced cells were identified and only the most stable retained. The resulting scatter plot containing 61 distinct hypothetical crystal structures (figure 3.4). A summary table of these structures is provided in the supporting information along with the atom numbering scheme used during the CSP search.
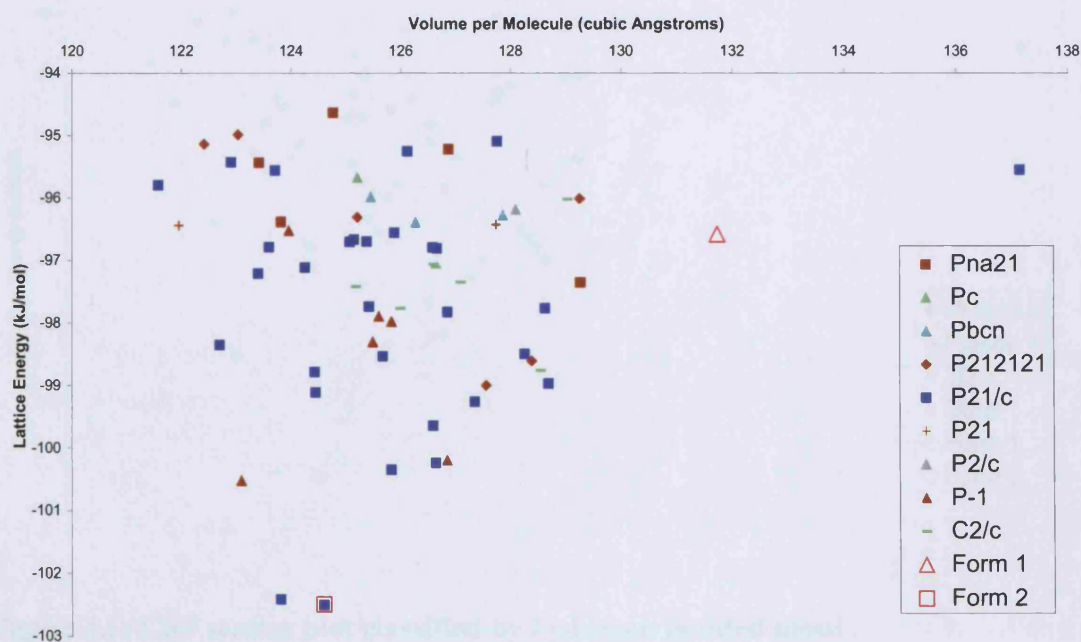
Figure 3.4: Scatter plot of all CSP generated structures for 5FU within 8 kJ mol⁻¹ of the global energy minimum

Comparison of the ExptMinOpt energy minimised structures of the two experimental polymorphs with the list of low energy hypothetical structures showed that the form 2 structure corresponded exactly to the global energy minimum structure, which was

nearly 6 kJ mol^{-1} more stable than form 1. Form 1 proved both higher in energy and less dense than many of the predicted structures.

5FU has two hydrogen bond donors and two hydrogen bond acceptors and it was expected that the computed structures would all show hydrogen bonding, and because of the equal number of donors and acceptors that all donors and acceptors would be satisfied. The set of structures within 8 kJ mol^{-1} of the global minimum showed a variety of hydrogen bond patterns so the structures were classified by their most defining hydrogen bonded feature, as identified by graph set analysis using the default graph set parameters in PLUTO.[141] Three recurring hydrogen bonded motifs were identified – two ribbon motifs and one sheet motif. Figure 3.5 shows the scatter plot of predicted structures classified by hydrogen bond motif.
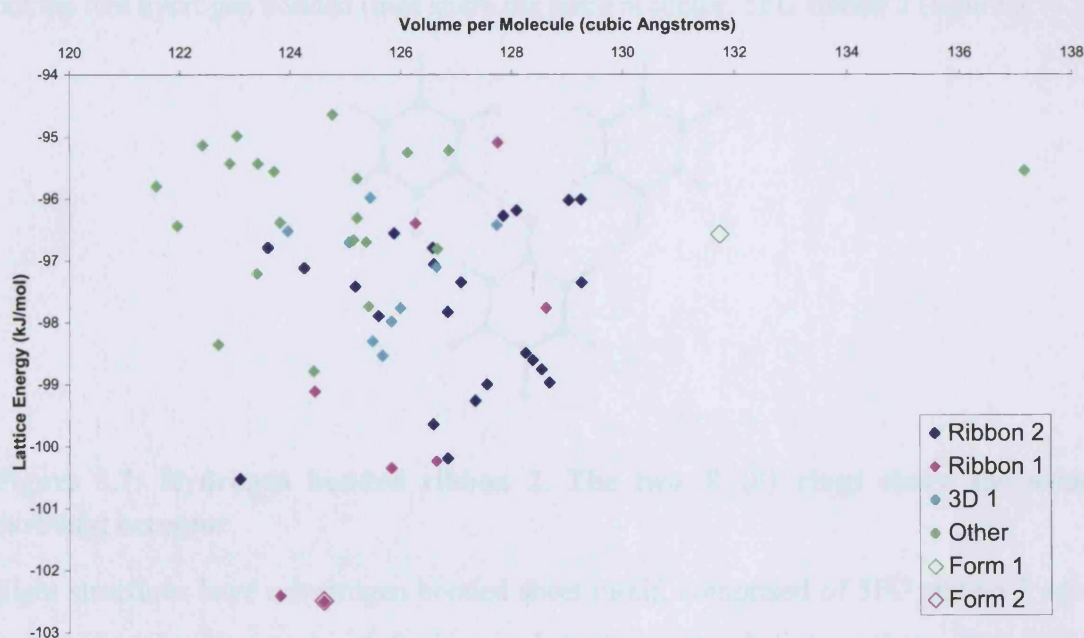
Figure 3.5: CSP scatter plot classified by hydrogen bonded motif

In eight structures, including four of five lowest energy structures, each molecule forms two $R_2^2(8)$ hydrogen bonds rings to produce a ribbon motif, 5FU ribbon 1 (figure 3.6). The two $R_2^2(8)$ interactions on each molecule do not share hydrogen bond donors.
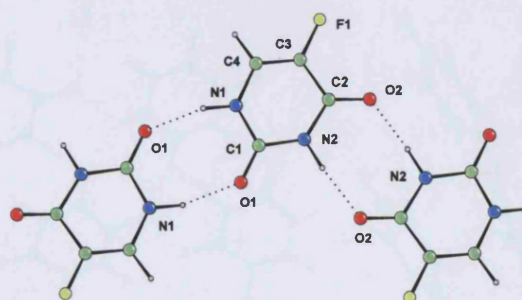
Figure 3.6: Hydrogen bonded ribbon motif 1. Two independent hydrogen bonded $R_2^2(8)$ rings propagate the ribbon. The alternate numbering scheme that was used in the CSP search and supporting information tables is shown

In 24 structures each molecule forms two $R_2^2(8)$ dimers, each to an adjacent molecule, but the two hydrogen bonded rings share the same acceptor, 5FU ribbon 2 (figure 3.7).
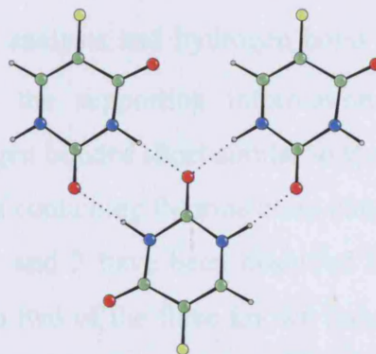
Figure 3.7: Hydrogen bonded ribbon 2. The two $R_2^2(8)$ rings share the same carbonyl acceptor

Eight structures have a hydrogen bonded sheet motif, comprised of 5FU ribbon 2 subunits within the sheet, and with further single hydrogen bonds linking these ribbons into the sheet structure, 5FU sheet 1. The sheets have a thickness of approximately 9 Å in the third dimension. Figure 3.8 shows two views of the sheet, both parallel to the plane of the sheet, and perpendicular to the 9 Å direction. The outer surfaces of the sheets are mainly made up of fluorine atoms.
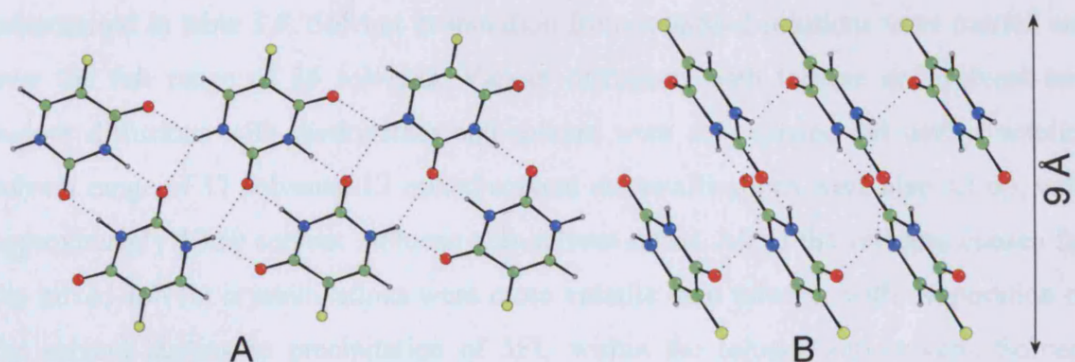
Figure 3.8: Hydrogen bonded sheet 1. Two views parallel to the plane of the sheet are shown. The first view shows the 5FU ribbon 2 sub-unit and the second view shows the single hydrogen bonds linking the ribbon 2 sub-units together to make the sheet

The other 21 structures show a variety of other hydrogen bonded motifs based on chains and $R_2^2(8)$ rings. Graph set analysis and hydrogen bond details for all 61 low energy structures are provided in the supporting information. None of the hypothetical structures produced a hydrogen bonded sheet similar to the sheet observed in the form 1 crystal structure with regions containing fluorine close contacts.

Both 5FU ribbons 1 and 2 have been observed in other 5-substituted uracils. 5FU ribbon 1 is observed in two of the three known polymorphs of 5-nitrouracil[167;168] and 5FU ribbon 2 is found in 5-chlorouracil,[169] 5-bromouracil,[170] 5-methyluracil (thymine)[171] and 5-eniluracil.[172]

3.3. Experimental crystallisation screen

3.3.1 Crystallisation screen

5-Fluorouracil was purchased from Aldrich Chemical Company (Poole, UK, 99% purity) and was used as purchased in the crystallisation screen, and was confirmed to be form 1 by XRPD. 5FU proved to be soluble in 26 solvents from the available library of 35. It was found to be insoluble in diethylether, cyclohexane, hexane, toluene, o-xylene, chloroform, dichloromethane, 1,2-dichloroethane and tetrachloroethylene. 5FU was slightly soluble in 11 solvents where the solubility was approximately 1-2 mg/ml at room temperature. It proved to be soluble in 15 solvents, where the solubility was greater than 1-2 mg/ml at room temperature. The 5FU crystallisation screen is

summarised in table 3.9. Solvent evaporation from saturated solutions were carried out over the full range of 26 solvents. Vapour diffusions with toluene anti-solvent and vapour diffusions with diethylether anti-solvent were also carried out over a smaller solvent range of 17 solvents. 12 mixed solvent recrystallisations were also set up, with approximately 50:50 solvent : toluene anti-solvent ratios. All of the solvents chosen for the mixed solvent crystallisations were more volatile than toluene, with evaporation of the solvent leading to precipitation of 5FU within the toluene anti-solvent. Solvent evaporation from sub-saturated solutions were set up over a range of eight solvents and both 25% and 50% saturation levels were used.

Of the 149 experiments summarised in table 3.9, 58 produced crystals that were large enough and of sufficient quality for characterisation by SXRD. In the other crystallisations either the crystallisation failed to give a precipitate, which was a common result for the vapour diffusion experiments, or the crystallisation gave a small amount of microcrystalline precipitate. For all crystallisations which produced single crystals the unit cell of a crystal of typical morphology was determined, and in seven cases new unit cells were observed and full data sets collected, yielding new crystal structures. Of the seven new forms six were solvates, and one was a new polymorph, form 2. The crystal structure of form 1 was also redetermined at 150 K, giving a more accurate determination of the crystal structure than the literature structure.

A crystallisation database, summarising all crystallisations and results, is provided on the supporting information CD (in Chapter_3_5Fluorouracil) and a summary spreadsheet detailing crystallisation conditions for all of the forms of 5-fluororacil is also given (Crystallisation_Summary.xls).

Solvent	Solubility (RT)	EV 100	EV 50	EV 25	VD ether	VD toluene	MX toluene
Acetonitrile	Slightly Soluble	4	2	1	2	1	1
Acetone	Slightly Soluble	4	2	1	2	1	1
1-butanol	Soluble	4	2	1	2	1	1
Ethanol	Soluble	4	2	1	2	1	1
Methanol	Soluble	3	2	1	2	1	1
1-propanol	Soluble	4	2	1	2	1	1
Nitromethane	Slightly Soluble	4	2	1	2	1	1
Water	Soluble	4	2	1	*	*	†
2-butanol	Soluble	3			2	1	1
2-propanol	Soluble	3			2	1	1
Tetrahydrofuran	Soluble	2			2	1	1
1,4-dioxane	Soluble	2			2	1	1
Ethylacetate	Slightly Soluble	3			2	1	1
Aniline	Slightly Soluble	3			2	1	†
Formamide	Slightly Soluble	3			2	1	†
2,2,2,-trifluoroethanol	Soluble	3			2	1	
Dimethylformamide	Soluble	1			2	1	†
Dimethylsulfoxide	Soluble	1			2	1	†
Acetaldehyde	Slightly Soluble	1					
Benzonitrile	Soluble	1					
2-chloroethanol	Slightly Soluble	1					
Ethylene glycol	Soluble	1					
Methylbenzoate	Slightly Soluble	1					
Methylethylketone	Slightly Soluble	1					
1-methyl-2-pyrrolidinone	Soluble	1					
tert-butylmethylether	Slightly Soluble	1					

Table 3.9: Crystallisation screen summary

Key: EV = solvent evaporation at saturation of 100%, 50% or 25%; VD = vapour diffusion with diethylether or toluene anti-solvent; all crystallisations were carried out at room temperature (RT). MX = mixed solvent; * = immiscible; † = solvent too involatile. The values in the table denote the number of repeat experiments

3.3.2 Redetermination of the crystal structure of form 1

5-Fluorouracil form 1

Crystals of 5-fluorouracil form 1 could be grown from many different solvents, with good quality crystals easily grown from ethanol and methanol. While different solvents were capable of giving a variety of different crystal morphologies, a recurring crystal habit was identified, that of a square based pyramid truncated before the apex (figure 3.9). The crystals proved prone to delamination on manipulation, with the delamination occurring parallel to the base of the pyramid and usually splitting the crystal into a multitude of very thin plate crystallites that themselves were usually twinned. This is consistent with the non-hydrogen bonded sheet composition of the crystal structure, with the sheets parallel to the 1 1 0 Miller planes. Single crystals of suitable quality for a high quality SXRD determination were grown by slow cooling a saturated solution of 5FU in nitromethane using a domestic refrigerator. The asymmetric unit is shown in figure 3.10 and the crystal structure is summarised in table 3.7. The F⋯F distances in this redetermined structure are F9⋯F29 = 3.046(2) Å, F29⋯F19 = 3.045(2) Å, F19⋯F39 = 3.091(2) Å, F9⋯F29 = 3.046(2) Å, confirming the close contacts in the fluorine tetramer.
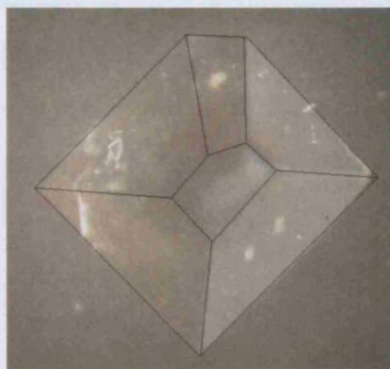
Figure 3.9: Common habit of 5FU form 1. In this example along with the four dominant side faces of the square based pyramid, a fifth smaller face is observed. The apex of the pyramid has not grown out. Lines have been added to the picture to highlight the different faces
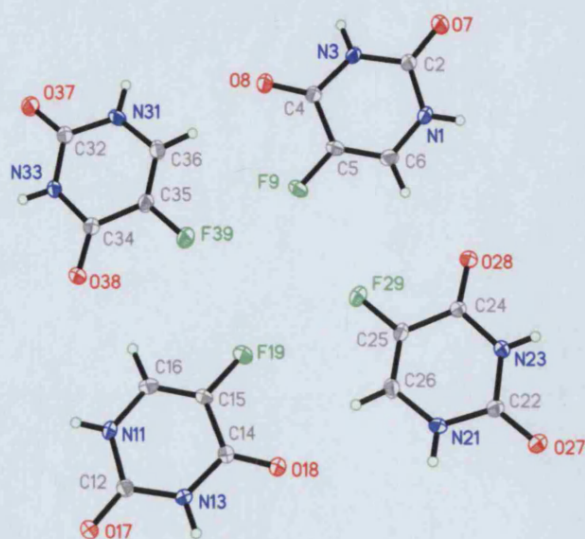
Figure 3.10: Asymmetric unit of 5FU form 1

3.3.3 5-Fluorouracil form 2

5-Fluorouracil form 2

A second polymorph of 5-fluorouracil was discovered by slow evaporation of a saturated solution of nitromethane at room temperature over a period of four months. This crystallisation method proved extremely unreliable for growing further samples of form 2 – only three of approximately 40 further repeated recrystallisations produced form 2, with all others giving form 1. Form 2 crystallises in the monoclinic space group $P2_1/c$ with one molecule in the asymmetric unit (figure 3.11 and table 3.7).
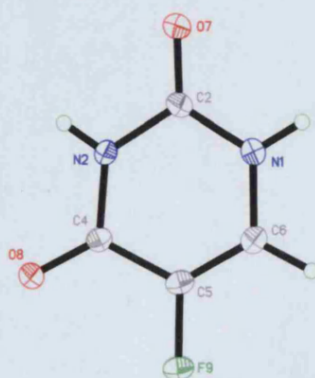
Figure 3.11: Asymmetric unit of 5FU form 2

The hydrogen bonded motif present in the crystal structure of 5FU form 2 is the ribbon 1 motif. The ribbons lie side-by-side to form rippled sheets with no strong hydrogen bond interactions between the ribbons in a sheet, or between adjacent sheets (figure 3.12). The sheets stack as the -1 0 2 Miller planes and stack directly upon one another when viewed parallel to the c crystallographic axis. There are no short F···F contacts in the crystal structure. The two hydrogen bonds, each of which is present once in each of the $R_2^2(8)$ rings, are N3-H3···O8 and N1-H1···O7.
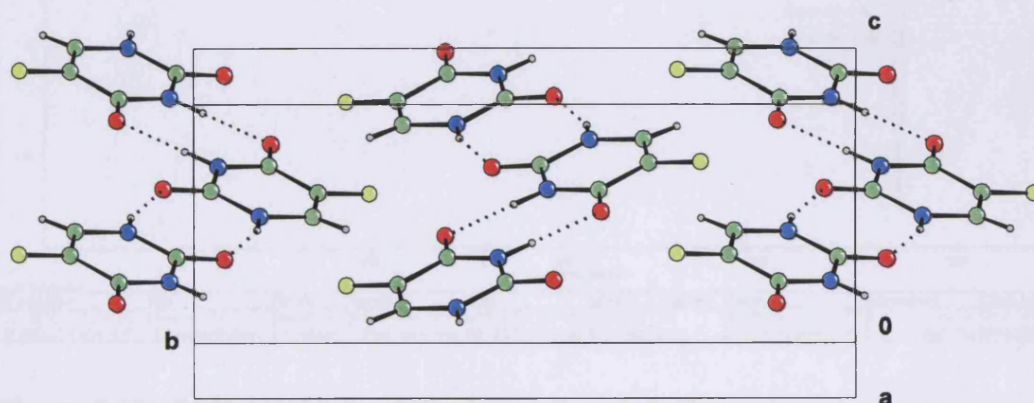
Figure 3.12: Three ribbons comprising part of one sheet in the 5FU form 2 crystal structure

Despite the unreliability of the crystallisation method, a sufficient quantity of form 2 was grown for thermal analysis. DSC experiments on both polymorphs were performed by Dr. Andrea Johnston (Strathclyde University) (figure 3.13). From these experiments it can be seen that form 2 melts at a lower temperature than form 1 and also has a smaller heat of melting (the area of the melt event).
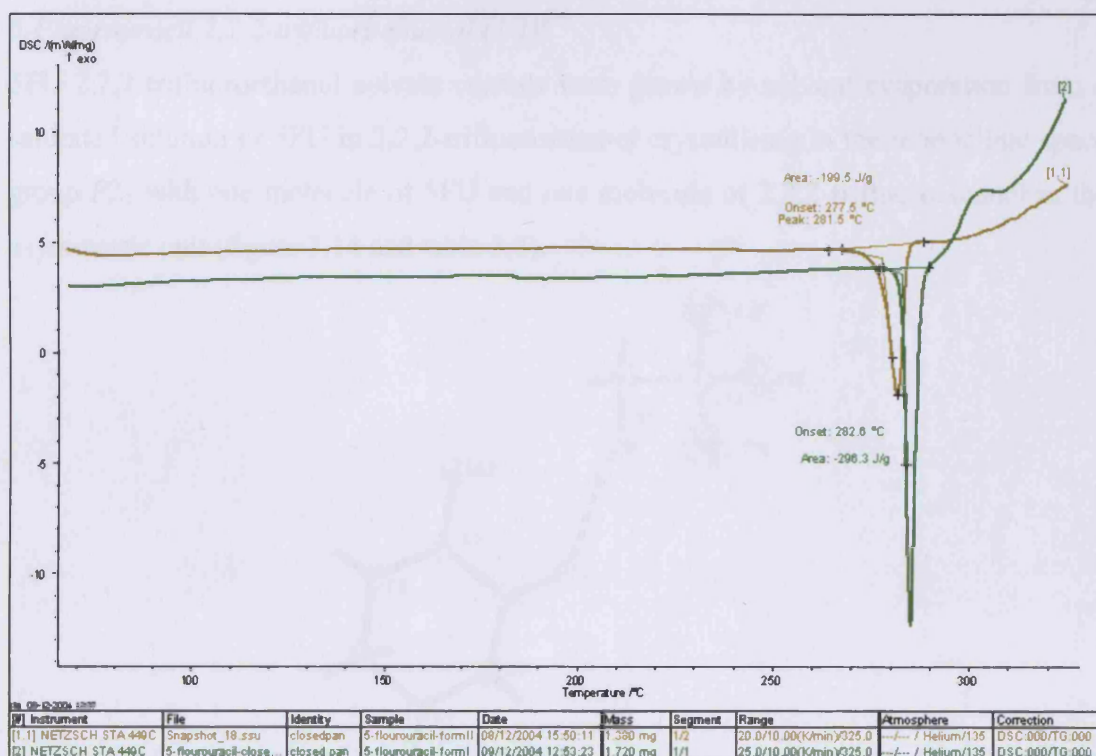
Figure 3.13: DSC results for 5FU forms 1 and 2. Form 1 is shown in green and form 2 in olive. Form 2 melts at lower temperature and has a lower heat of melting than form 1

3.3.4 5-Fluorouracil solvates

The crystallisation screen produced six new solvates, some of which show hydrogen bond features that can be related to the computed structures or the experimental polymorphs and the solvates are discussed emphasising these features. Also reported is the crystal structure of a solid solution of 5-fluorouracil and thymine.

3.3.5 2,2,2-Trifluoroethanol and benzonitrile solvates

These two solvates are reported together because both contain 5FU ribbon 2 sub-units in their hydrogen bonding patterns. In both of these crystal structures the carbonyl hydrogen bond acceptor that is left unused in the 5FU ribbon 2 structure is satisfied by interactions with solvent molecules – in the case of the 2,2,2-trifluoroethanol solvate it forms a strong hydrogen bond to the hydroxyl group of the solvent molecule and in the case of the benzonitrile solvate it forms a weak C-H···O interaction with the benzonitrile solvent, which does not contain any strong hydrogen bond donating groups.

5-Fluorouracil 2,2,2-trifluoroethanol (1/1)[173]

5FU 2,2,2-trifluoroethanol solvate crystals were grown by solvent evaporation from a saturated solution of 5FU in 2,2,2-trifluoroethanol crystallising in the monoclinic space group $P2_1$ with one molecule of 5FU and one molecule of 2,2,2-trifluoroethanol in the asymmetric unit (figure 3.14 and table 3.7).
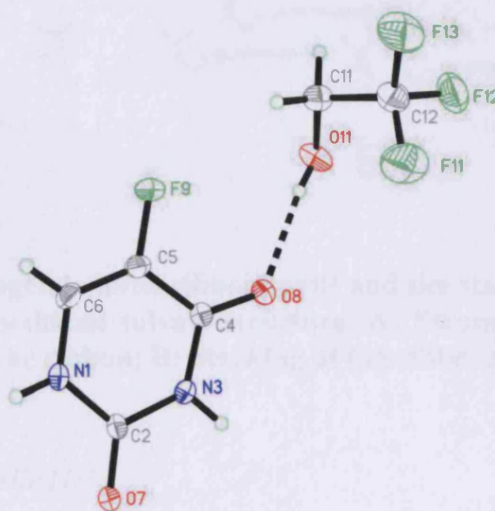
Figure 3.14: Asymmetric unit of 5FU 2,2,2-trifluoroethanol solvate

The principal hydrogen bonded motif in the structure is the 5FU ribbon 2. Each 5FU molecule participates in two $R_2^2(8)$ rings, each being comprised of one N1-H1···O7 hydrogen bond and one N3-H3···O7 hydrogen bond. The carbonyl oxygen atom in the 5FU molecule not used in the ribbon hydrogen bonding is hydrogen bonded to a 2,2,2-trifluoroethanol molecule. The 2,2,2-trifluoroethanol molecules form the outer edge of the ribbon, with the 5FU molecules forming the inside of the ribbon (figure 3.15). There is also a F···F (2.891(2) Å) close contact between one of the trifluoromethyl fluorine atoms from the solvent and the 5FU fluorine atom, thus further stabilising the ribbon. The ribbons stack parallel to the *ab* plane with the 2,2,2-trifluoroethanol molecules from adjacent stacks of ribbons in van der Waals contact.
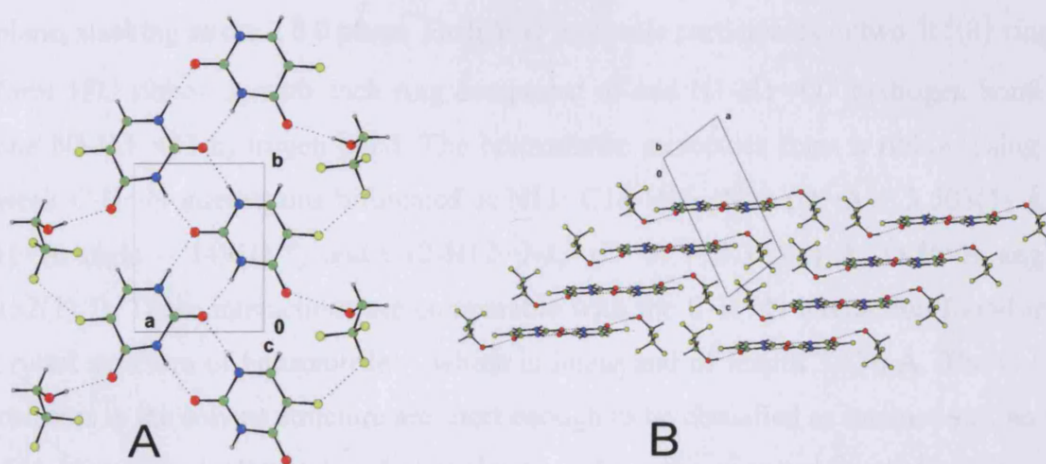
Figure 3.15: The hydrogen bonded ribbon motif and the stacking of the ribbons in the 5FU 2,2,2-trifluoroethanol solvate structure. A: Strong hydrogen bonds and short F···F contacts in the ribbon; B: Stacking of the ribbons

5-Fluorouracil benzonitrile (1/1)

Small block crystals of a 5FU benzonitrile solvate were grown by solvent evaporation from a saturated solution over a period of nearly a year. This solvate crystallises in the space group $P2_1/c$ with one molecule of 5FU and one molecule of benzonitrile in the asymmetric unit (figure 3.16 and table 3.7).
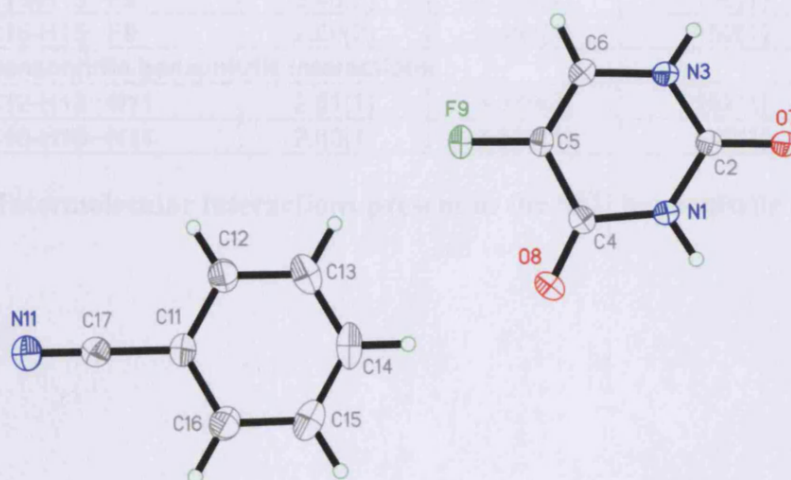
Figure 3.16: Asymmetric unit of the 5FU benzonitrile solvate

The 5FU ribbon 2 motif is present in the crystal structure, alternating with ribbons of benzonitrile molecules to form sheets (figure 3.17). The sheets lie parallel to the *bc*

plane, stacking as the 2 0 0 plane. Each 5FU molecule participates in two $R_2^2(8)$ rings to form 5FU ribbon 2, with each ring comprised of one N1-H1···O7 hydrogen bond and one N3-H3···O7 hydrogen bond. The benzonitrile molecules form a ribbon using two weak C-H···N interactions bifurcated at N11: C16-H16···N11 (D···A = 3.503(1) Å, D-H···A angle = 149(1) °) and C12-H12···N11 (D···A = 3.379(2) Å, D-H···A angle = 152(1) °). These interactions are comparable with the C-H···N interaction found in the crystal structure of benzonitrile[174] which is linear and of length 3.676 Å. The C-H···N contacts in the solvate structure are short enough to be classified as interactions, and the distortion of the hydrogen bonding angles away from the expected linear disposition can be explained by the bifurcation at N11. Two C-H···F and one C-H···O weak hydrogen bonds are the only interactions between the 5FU ribbons and the benzonitrile ribbons. The intermolecular interactions present in the crystal structure are summarised in table 3.4.

	H···A (Å)	D···A (Å)	<(D-H···A) (°)
5FU hydrogen bonds			
N1-H1···O7	1.91(2)	2.799(1)	172(1)
N3-H3···O7	1.91(2)	2.769(1)	175(1)
5FU benzonitrile interactions			
C14-H14···O8	2.52(2)	3.441(1)	159(1)
C13-H13···F9	2.45(2)	3.368(1)	150(1)
C15-H15···F9	2.59(2)	3.451(1)	150(1)
Benzonitrile benzonitrile interactions			
C12-H12···N11	2.51(1)	3.379(2)	152(1)
C16-H16···N11	2.63(1)	3.503(1)	149(1)

Table 3.4: Intermolecular interactions present in the 5FU benzonitrile solvate
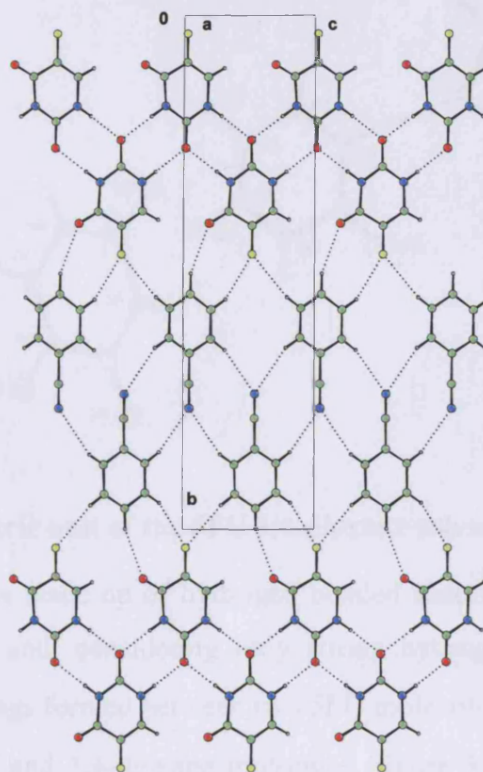
Figure 3.17: The hydrogen bonding present in a single sheet of the 5FU benzonitrile solvate. 5FU ribbon 2 sub-units can be observed forming weak hydrogen bonded interactions with benzonitrile molecules. Adjacent benzonitrile molecules interact by weak C-H···O interactions

3.3.6 1,4-Dioxane solvate

5-Fluorouracil 1,4-dioxane (4/1)[175]

The crystal structure of this solvate is comprised of hydrogen bonded sheets which include hydrogen bonded rings that resemble the largest ring found in the 5FU form 1 crystal structure. 5FU crystallises with 1,4-dioxane in a 4:1 ratio in the triclinic space group $P\bar{1}$. Block crystals were grown by solvent evaporation from a saturated solution of 5FU in 1,4-dioxane. The 1,4-dioxane molecule is located on the inversion centre at origin of the unit cell and consequently there is half of a 1,4-dioxane molecule and two 5FU molecules in the asymmetric unit (figure 3.18 and table 3.7).
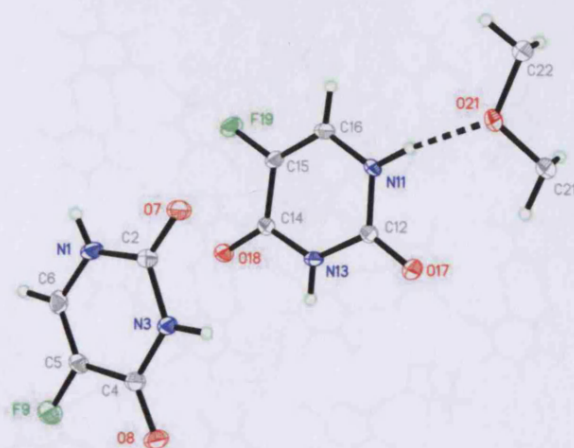
Figure 3.18: Asymmetric unit of the 5FU 1,4-dioxane solvate

This crystal structure is made up of hydrogen bonded sheets which lie parallel to the 2 -1 1 Miller planes and, considering only strong hydrogen bonds, each sheet is comprised of $R_2^2(8)$ rings formed between two 5FU molecules and large $R_{12}^{12}(54)$ rings that contain both 5FU and 1,4-dioxane molecules (figure 3.19). At the centre of the large ring a tetramer of F···F close contacts is present, and this ring is related to the $R_8^8(40)$ rings present in 5FU form 1, with the F···F tetramer and six central 5FU molecules (those overlaid in figure 3.20) common to both rings. Instead of the two further 5FU molecules that complete the form 1 ring, in the solvate two 1,4-dioxane molecules and four further 5FU molecules are incorporated to form the larger ring.

Hydrogen bond	H···A (Å)	D···A (Å)	<(D-H···A) (°)
N1-H1···O17	1.98(3)	2.798(2)	167(2)
N3-H3···O7	1.95(2)	2.857(2)	176(2)
N11-H11···O21	1.84(2)	2.746(2)	171(2)
N13-H13···O18	1.98(2)	2.824(2)	175(2)
F···F short contacts			
F9···F19	n/a	3.183(2)	n/a
F9···F19	n/a	3.034(2)	n/a

Table 3.5: Hydrogen bonds and short F···F contacts in the 5FU 1,4-dioxane solvate
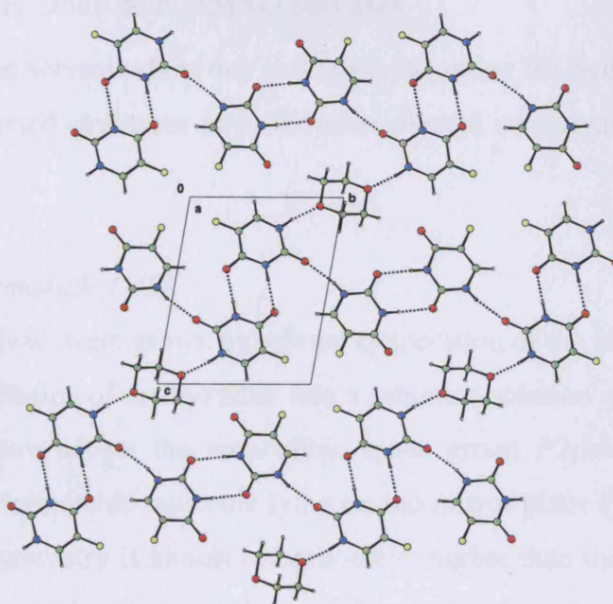
Figure 3.19: The hydrogen bonded sheet present in the 5FU 1,4-dioxane solvate
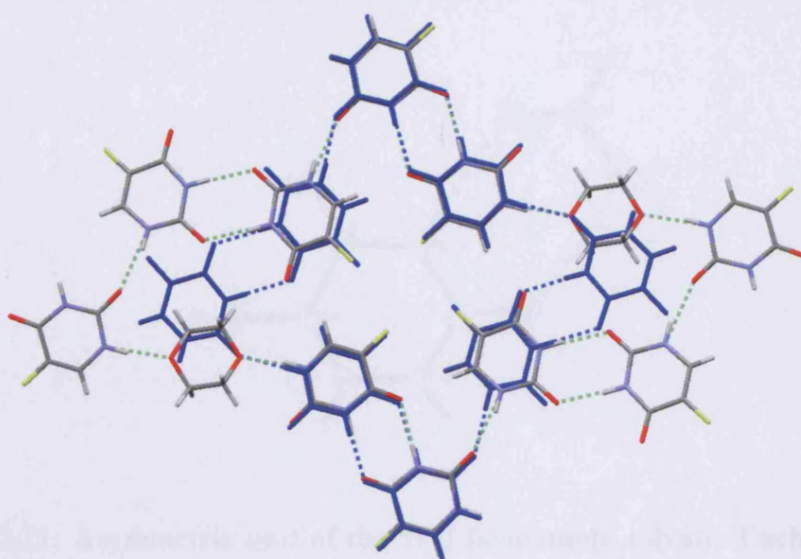
Figure 3.20: Overlay of 5FU 1,4-dioxane solvate $R_{12}^{12}(54)$ ring (coloured by element) and 5FU form 1 $R_8^8(40)$ ring (coloured blue). The overlay was matched on the six central 5FU molecules that are seen to match closely, while at the left and right of the figure the differences in the two rings can be observed

3.3.7 Formamide, DMF and DMSO solvates

None of these three solvates show any similarities to either the hydrogen bonded motifs found in the predicted structures or to the non-solvated crystal structures or indeed to one another.

5-Fluorouracil formamide (1/1)

Crystals of this solvate were grown by solvent evaporation of the resultant solution from a failed vapour diffusion of diethyl ether into a saturated solution of 5FU in formamide. The crystal structure adopts the monoclinic space group $P2_1/m$ with both the 5FU molecule and the formamide molecule lying on the mirror plane (figure 3.21 and table 3.7). The metric symmetry is almost orthorhombic, higher than the actual symmetry of the crystal structure, and in this case the crystal proved to be a pseudo-merohedral twin, which was treated during the refinement by the application of the twin law (1 0 0 0 -1 0 0 0 -1) and refined with a scale factor of 0.34.
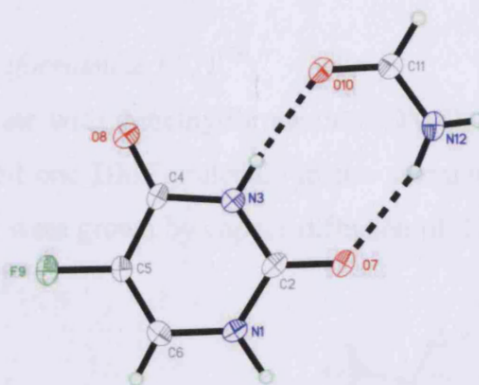
Figure 3.21: Asymmetric unit of the 5FU formamide solvate. Each atom has half occupancy as both molecules lie on the mirror plane

In this crystal structure each 5FU molecule does not hydrogen bond to any other 5FU molecules, but hydrogen bonds to three formamide molecules, figure 3.22.
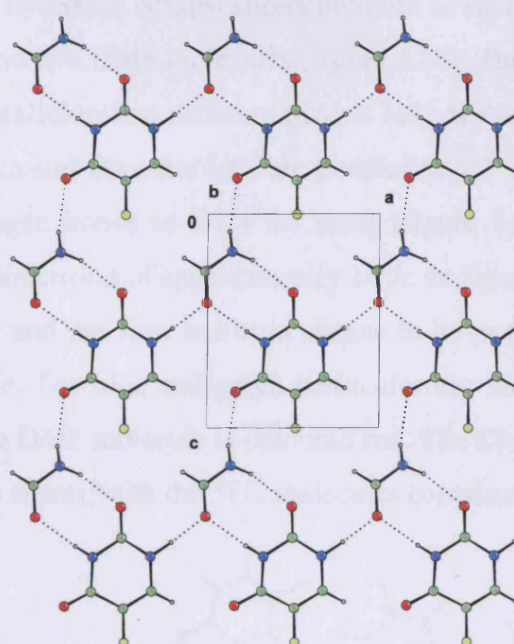
Figure 3.22: Hydrogen bonded sheet present in the 5FU formamide solvate

5-Fluorouracil dimethylformamide (2/1)[176]

5FU forms a hemisolvate with dimethylformamide (DMF), crystallising in $P2_1/c$ with two 5FU molecules and one DMF molecule in the asymmetric unit (figure 3.23 and table 3.7). The crystals were grown by vapour diffusion of diethyl ether into a saturated solution of 5FU in DMF.
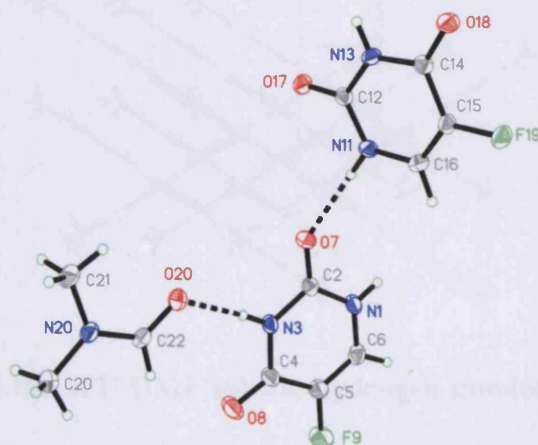
Figure 3.23: Asymmetric unit of the 5FU DMF solvate

The structure contains hydrogen bonded sheets built out of smaller sub-units comprising four 5FU molecules and two DMF molecules (figure 3.24). The sub-units do not lie flat in the sheet, but lie parallel to two different planes: half of the sub-units are parallel to the 2 -1 -4 Miller plane and the other half are parallel to 2 1 -4. These two groups are cross linked by hydrogen bonds to form the sheet (figure 3.25) and the sheet has a thickness in the third dimension of approximately 14 Å. In figure 3.25 the molecules are coloured by symmetry and the view has been chosen to be parallel to the two planes in which the sub-units lie. The blue and green molecules are the symmetry independent 5FU molecules and the DMF molecule is coloured red. The DMF molecules are present on both surfaces of the sheets, with the 5FU molecules contained within the sheet.
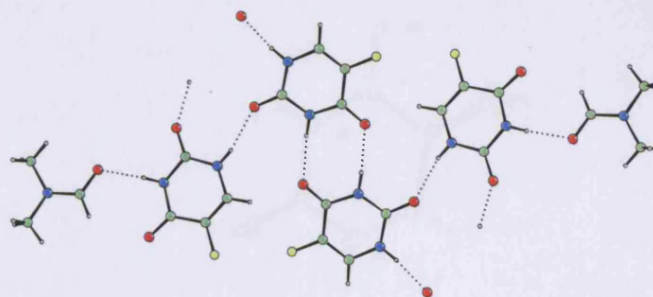
Figure 3.24: Hydrogen bonded sub-unit present in the 5FU DMF solvate. The four points of contact with adjacent sub-units used to build the hydrogen bonded sheet are shown
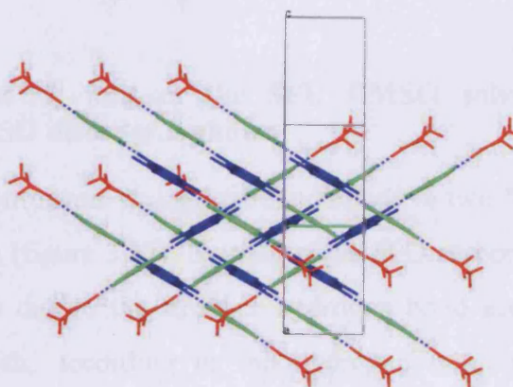
Figure 3.25: View of the 5FU DMF solvate hydrogen bonded sheet parallel to the plane of the sheet

5-Fluorouracil dimethylsulfoxide (1/1)[177]

5FU forms a 1:1 solvate with dimethylsulfoxide (DMSO), crystallising in $P2_1/c$ (table 3.7). Crystals were grown by vapour diffusion of diethyl ether into a saturated solution of 5FU in DMSO. The DMSO molecule is disordered with the sulfur atom split over two sites with the two components showing opposite pyrimidisation. Only the sulfur atom and the methyl hydrogen atoms were modelled as disordered. The ratio of the major to minor component refined to 0.94, and only the major component is shown in the asymmetric unit in figure 3.26. The following bond lengths were restrained to be equivalent in the two components, to within 0.01 Å: S20-O20 and S20'-O20; S20-C20 and S20'-C20; S20-C21 and S20'-C21.
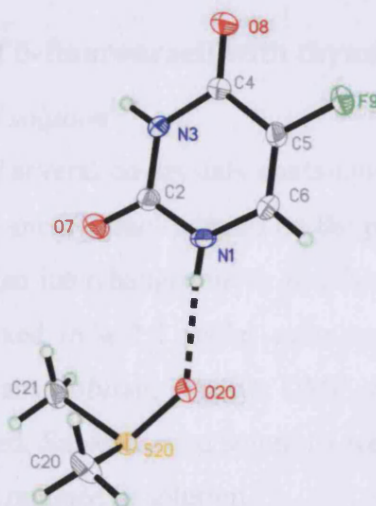
Figure 3.26: Asymmetric unit of the 5FU DMSO solvate. Only the major component of the DMSO disorder is shown

The DMSO oxygen atom forms strong hydrogen bonds to two 5FU molecules to give a hydrogen bonded chain (figure 3.27). Neither of the 5FU carbonyl oxygen atoms forms strong hydrogen bonds due to the stronger hydrogen bond acceptor capability of the sulfoxide oxygen which, according to the hydrogen bond rules,[100] will form the dominant association with the N-H hydrogen bond donors from the 5FU molecules. The ribbons propagate parallel to the *b* axis and they stack to form layers parallel to the *bc* plane. A similar hydrogen bonding pattern is found in the DMSO solvate of 5-nitrouracil.[167]
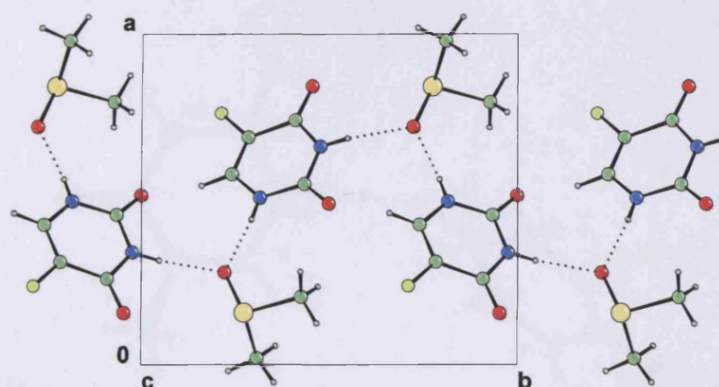
Figure 3.27: Hydrogen bonded ribbon present in the 5FU DMSO solvate. DMSO molecules and 5FU molecules alternate in the ribbon

3.3.8 Co-crystallisation of 5-fluorouracil with thymine

5-Fluorouracil thymine solid solution[178]

The presence in the CSD of several co-crystals containing 5FU led to attempts to co-crystallise it with thymine (5-methyluracil), based on the premise that the fluorine atom and the methyl group are often interchangeable to give isomorphic crystal structures.[179] 5FU and thymine were mixed in a 1:1 molar ratio and sub-saturated solutions in methanol, nitromethane, tetrahydrofuran, DMSO, DMF, water, benzonitrile and 2,2,2-trifluoroethanol were prepared. Sub-saturated solutions were used so that the 1:1 molar ratio prior to dissolution was retained in solution.

Crystallisation from 2,2,2-trifluoroethanol produced a new unit cell, a full data set was collected and a new crystal structure[178] determined. This proved to be a solid solution of 5FU and thymine (table 3.7). Solid solutions are conceptually distinct from co-crystals. Solid solutions have been defined as "a homogeneous crystalline phase in which some of the constituent molecules are substituted by foreign molecules that possess sufficient similarity that the lattice dimensions are changed only slightly".[6] The asymmetric unit (figure 3.28) contains two independent sites and both of these sites have non-integer occupancies of both 5FU and thymine. At each site the 5FU and thymine occupancies sum to one, so each site is fully occupied. Because the only molecular difference between 5FU and thymine is the functional group bonded to the 5-position, it is only this that differentiates the presence of 5FU from thymine in the crystal structure.
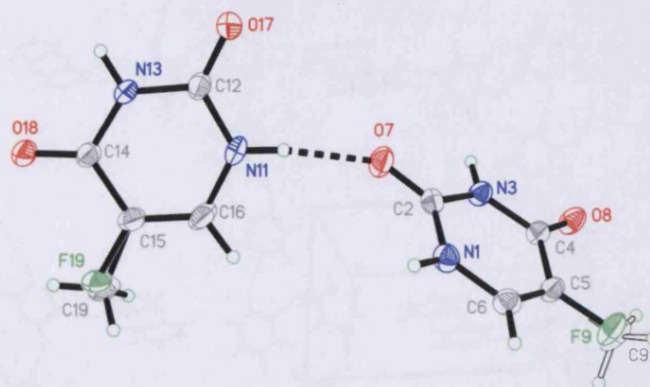
Figure 3.28: Asymmetric unit of the 5FU/thymine solid solution crystal structure. Positions 9 and 19 have non-integer occupancies of fluorine atoms and methyl groups and both are shown

Three separate crystal structure determinations using three different crystals were recorded. Two of the crystals were grown from a 1:1 ratio of 5FU : thymine and one crystal was grown from a 2:1 ratio. The ratio in the crystallisation solution influenced the relative ratios of 5FU : thymine at positions 9 and 19 in the crystal structure with the higher ratio crystallisation solution giving a higher proportion of 5FU at both sites (table 3.6, position 19 in all three structures preferentially incorporates 5FU rather than thymine). Solid solution crystals with this structure could not be grown from 3:1 or 1:2 5FU : thymine ratios – suggesting that 5FU and thymine have limited solubility in this solid solution.

Ratio 5FU:Thymine in solution	% 5FU in crystal structure	
	position 9	position 19
1:1	0.52(1)	0.70(1)
1:1	0.55(1)	0.69(2)
2:1	0.66(1)	0.76(1)

Table 3.6: 5FU : thymine occupancies in the three determinations of the 5FU:thymine solid solution crystal structure. The ratios at crystallographic positions 9 and 19 are given separately

In the crystal structure each 5FU/thymine forms four hydrogen bonds, giving a $R_2^2(8)$ ring and two single hydrogen bonds, and these interactions then form hydrogen bonded nets (figure 3.29). The nets in the crystal structure interpenetrate, and hydrogen bond together at the points of interpenetration. None of the hypothetical 5FU crystal structures corresponds to this crystal structure.
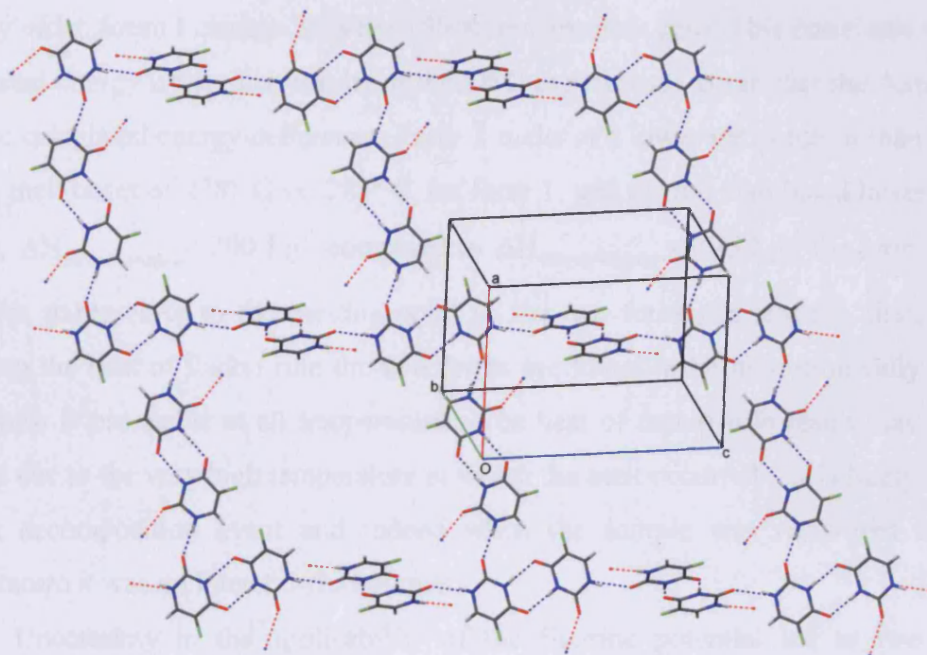
Figure 3.29: A single hydrogen bonded net present in the 5FU/thymine solid solution crystal structure. Four ring sub-units are shown

3.4 Discussion

3.4.1 The prediction of polymorphs of 5-fluorouracil

From the CSP search it was found that 5FU ribbon 1 was present in four of the five lowest energy hypothetical structures, showing that the close packing ability and intermolecular interactions available with this motif leads to lower energy crystal structures than those available based on the other observed motifs. Consequently on a calculated thermodynamic basis, the 5FU ribbon 1 motif was most likely to be found in a new polymorph. The experimental discovery of the new polymorph, form 2, that did indeed exhibit this motif confirmed the hypothesis.

The calculated energy difference from the energy minimisation of both polymorphs is almost 6 kJ mol^{-1} with form 2 most stable. It should be remembered that the energy minimisation process is carried out in effect at 0 K. The density of forms 1 and 2 from the single crystal structures determined in this study were compared. The density rule[36] states that, for a pair of polymorphs, the more dense structure is the more stable structure at 0 K. Form 2 is denser than form 1 at 150 K and so the density rule

indicates that form 2 is more stable than form 1 at absolute zero, assuming that the density order doesn't change between 150 K and absolute zero. This correlates with the calculated energy order. The results of the DSC experiments contradict the density rule and the calculated energy difference. Form 2 melts at a lower temperature than form 1, with a melt onset of 278° C *vs.* 283° C for form 1, and form 2 also has a lower heat of fusion, $\Delta H_{Form2 \rightarrow liquid} = 200$ J g^{-1} compared to $\Delta H_{Form1 \rightarrow liquid} = 296$ J g^{-1} for form 1. Form 2 is less stable close to the melting point of the two forms, as it melts first, and by applying the heat of fusion rule the two forms are found to be monotropically related, with form 2 less stable at all temperatures. The heat of fusion rule result may well be invalid due to the very high temperature at which the melt occurred – it is likely that this was a decomposition event and indeed when the sample was recovered to room temperature it was no longer 5-fluorouracil.

Uncertainty in the applicability of the fluorine potential led to two further minimisations of the form 1 ExptMinOpt structure with modified fluorine potential parameters. The fluorine dispersion term was increased by 20% in one minimisation and the repulsion pre-exponential factor was decreased by 25% in the other. In neither case was the reproduction of the form 1 structure significantly altered and the lattice energy in both cases decreased by less than 1.5 kJ mol^{-1}. It can be concluded that the reproduction of the form 1 structure by the unmodified FIT potential is satisfactory.

The method of crystallising form 2 from nitromethane proved unreliable, with form 1 returned more often than form 2. To gain an understanding of the crystallisation processes that could lead to either form 1 or form 2 from nitromethane solutions, Said Hamad (the Royal Institution) undertook molecular dynamics studies[180] of supersaturated 5FU solutions in pure nitromethane, pure water and nitromethane doped with 0.1% water. The simulation of 5FU molecules in water showed that the N-H and C=O groups were strongly hydrogen bonded by water molecules leading to strong solvation of this part of the molecule, while the fluorine atom was more hydrophobic. This relative lack of fluorine solvation led by default to F···F interactions in the simulation, because the conventional hydrogen bonds were interrupted by the strongly bound water. In nitromethane the solvation of the 5FU molecules was generally weaker and lead to a greater number of strong N-H···O=C hydrogen bonds compared to the water simulation and even the formation of $R_2^2(8)$ 5FU dimer and trimer self-

associations. In the water-doped nitromethane simulation, with four water molecules per 5FU molecule, the water molecules generally formed hydrogen bonds with the 5FU molecules again blocking the formation of strong hydrogen bonds, or in the case where one single hydrogen bond did occur, prevented the formation of the $R_2^2(8)$ dimer. Nitromethane is a hygroscopic solvent and the simulations show that even 0.1% contamination with water could be sufficient to disrupt the form 2 crystallisation pathway, yielding form 1 by default. This concurs with the experimental evidence, and suggests that even initially dry nitromethane could collect sufficient atmospheric water to alter the crystallisation pathway in favour of form 1.

3.4.2 Relationship between CSP structures and solvate crystal structures

The 5FU ribbon 2 motif was the most popular motif present in the set of low energy hypothetical structures and this, along with its presence in the crystal structures of several other 5-substituted uracils, indicated that it was a motif that could be expected to be observed in a new polymorph. However 5FU ribbon 2 has a significant disadvantage compared to 5FU ribbon 1 that could account for its lowest energy structures not being as low in energy as the lowest energy structures based on 5FU ribbon 1. In 5FU ribbon 2 the two $R_2^2(8)$ rings share the same carbonyl acceptor and the second carbonyl oxygen present in the 5FU molecule is not hydrogen bonded. There are no strong hydrogen bond donors that are not involved in the two $R_2^2(8)$ rings in 5FU ribbon 2 and consequently the second 5FU carbonyl oxygen does not form a strong hydrogen bond in crystal structures that exhibit 5FU ribbon 2. Two of the six solvates discovered in the course of the crystallisation screen show the 5FU ribbon 2 motif, with the solvent providing the extra hydrogen bonding capability to enable the second 5FU carbonyl oxygen to be hydrogen bonded. In the case of the 2,2,2-trifluoroethanol solvate the hydroxyl group from the solvent hydrogen bonds to the second 5FU carbonyl oxygen and in the benzonitrile solvate it forms a weak C-H$\cdots$O hydrogen bond to the benzonitrile molecule, which is not capable of forming strong hydrogen bonds.

The 5FU form 1 structure is not observed in the low energy hypothetical structures because of a limitation of the CSP procedure which can only generate structures with one molecule in the asymmetric unit and the most unique aspect of the form 1 structure, the F$\cdots$F tetramer, is not observed in any of the low energy structures.

The fluorine tetramer was observed, however, in the crystal structure of the 1,4-dioxane solvate. This solvate structure could be viewed as the result of an interruption to the form 1 crystallisation pathway, with the F···F tetramer feature formed, but the formation of the ring around the tetramer interrupted by the strongly hydrogen bonded 1,4-dioxane molecules, leading to the solvate rather than form 1.

The three other solvates did not show any similarities to the predicted motifs.

3.5 Conclusion

The crystal structure prediction of 5-fluorouracil generated a large range of low energy hypothetical structures, but only a small number of recurrent hydrogen bonded motifs. The discovery during this study of one new polymorph, six solvates and one solid-solution co-crystal containing 5-fluorouracil highlights the versatility of crystallisation of 5-fluorouracil, both with itself and other species.

The hydrogen bonded motif present in the lowest energy predicted structures proved to be the motif observed in the new polymorph discovered during the experimental crystallisation screen and upon energy minimisation this new polymorph corresponded exactly to the lowest energy predicted structure. This crystal structure prediction would be judged a success by the criteria of the CCDC international blind tests,[121-123] as the global energy minimum structure would be submitted as one of the three predicted structures most likely to correspond to an observed crystal structure. Attempts to elucidate which polymorph was thermodynamically most stable proved inconclusive.

The most prevalent hydrogen bonded ribbon motif found in the search structure was experimentally observed in two solvates, and its discovery in these crystal structures proved that this ribbon was indeed a realistic and experimentally realisable hydrogen bonded motif, that required the further hydrogen bonding functionality of the solvent in these crystal structures before it could crystallise. The hydrogen bonded motif found in 5-fluorouracil form 1 was not observed in any of the hypothetical structures, and even its most salient feature, the tetramer of F···F close contacts, was not found. This feature was observed however in one of the solvates, which may be an example of an interruption to the form 1 crystallisation pathway.

Crystal Data

Compound name	5-Fluorouracil form 1	5-Fluorouracil form 2	5-Fluorouracil 2,2,2-trifluoroethanol	5-Fluorouracil benzonitrile
Empirical formula	$C_4H_3N_2O_2F$	$C_4H_3N_2O_2F$	$C_4H_3N_2O_2F$, $C_2H_3OF_3$	$C_4H_3N_2O_2F$, C_7H_5N
Formula weight	130.1	130.1	230.1	233.2
Crystal system, space group	Triclinic, $P\bar{1}$	Monoclinic, $P2_1/c$	Monoclinic, $P2_1$	Monoclinic, $P2_1/c$
a (Å)	8.6329(12)	5.0433(12)	5.3976(6)	7.0460(7)
b (Å)	9.1560(13)	14.935(3)	6.7062(8)	24.035(2)
c (Å)	12.5796(18)	6.6049(15)	12.1098(14)	6.8640(7)
α (°)	99.119(2)	90	90	90
β (°)	100.021(3)	108.884(4)	102.807(2)	116.554(2)
γ (°)	90.017(2)	90	90	90
V (Å^3)	966.4(2)	470.7(2)	427.44(9)	1039.8(2)
Z', Z	4, 8	1, 4	1, 2	1, 4
D(calc (g cm^{-3}))	1.788	1.836	1.788	1.490
Data Collection				
Crystal size (mm)	0.22 x 0.20 x 0.19	0.30 x 0.17 x 0.13	1.49 x 0.34 x 0.17	0.61 x 0.37 x 0.20
Temperature (K)	150(2)	150(2)	150(2)	150(2)
hkl range (h, k, l)	-11→11, -12→11, -16→16	-6→6, -19→19, -8→8	-7→4, -8→8, -15→15	-8→9, -31→27, -8→9
Reflections measured, R$_{int}$	8688, 0.0252	3956, 0.0198	2621, 0.0159	6296, 0.0158
Independent reflections	4489	1127	1090	2458
Reflections I>2σ(I)	3470	1020	1060	2120
Refinement				
Parameters refined	373	94	160	186
R(F) (I>2σ(I))	0.049	0.054	0.027	0.037
wR(F^2) (all reflections)	0.122	0.153	0.073	0.112
Residual electron density (min, max (e Å^{-3}))	0.34, -0.29	0.42, -0.26	0.30, -0.25	0.27, -0.24

Crystal Data

Compound name	5-Fluorouracil 1,4-dioxane	5-Fluorouracil formamide	5-Fluorouracil dimethylformamide	5-Fluorouracil dimethylsulfoxide
Empirical formula	$C_4H_3N_2O_2F$, $\frac{1}{4}(C_4H_8O_2)$	$C_4H_3N_2O_2F$, CH_3NO	$C_4H_3N_2O_2F$, $\frac{1}{2}(C_3H_7NO)$	$C_4H_3N_2O_2F$, C_2H_6OS
Formula weight	152.1	175.1	166.6	208.2
Crystal system, space group	Triclinic, $P\bar{1}$	Monoclinic, $P2_1/m$	Monoclinic, $P2_1/n$	Monoclinic, $P2_1/c$
a (Å)	7.0847(11)	6.827(4)	14.7361(18)	9.8831(10)
b (Å)	8.4733(13)	6.111(3)	5.8693(7)	10.8128(11)
c (Å)	10.2291(15)	8.424(4)	16.397(2)	8.6842(9)
α (°)	98.128(3)	90	90	90
β (°)	96.913(3)	90.313(8)	100.524(2)	107.397(2)
γ (°)	99.785(3)	90	90	90
V (Å^3)	592.5(2)	351.4(3)	1394.3(3)	885.58(16)
Z', Z	2, 4	0.5, 2	2, 8	1, 4
D(calc (g cm^{-3}))	1.705	1.655	1.588	1.562
Data Collection				
Crystal size (mm)	0.35 x 0.24 x 0.03	0.33 x 0.26 x 0.15	0.42 x 0.21 x 0.11	0.29 x 0.21 x 0.11
Temperature (K)	150(2)	150(2)	150(2)	298(2)
hkl range (h, k, l)	-9→9, -11→11, -13→13	-8→8, -7→7, -10→10	-19→19, -7→7, -21→20	-13→12, -14→14, -11→11
Reflections measured, R_{int}	5320, 0.0291	2892, 0.0243	11701, 0.0315	7666, 0.0220
Independent reflections	2741	866	3331	2127
Reflections I>2σ(I)	2131	765	2768	1921
Refinement				
Parameters refined	230	92	238	140
R(F) (I>2σ(I))	0.052	0.041	0.049	0.036
wR(F^2) (all reflections)	0.114	0.111	0.108	0.090
Residual electron density (min, max (e Å^{-3}))	0.33, -0.33	0.43, -0.28	0.29, -0.22	0.40, -0.54

Crystal Data

Compound name	5-Fluorouracil thymine
Empirical formula	$0.61(C_4H_3N_2O_2F)$, $0.39(C_5H_6N_2O_2)$
Formula weight	257.1
Crystal system, space group	Monoclinic, $C2/c$
a (Å)	19.3785(15)
b (Å)	5.9918(5)
c (Å)	20.0293(15)
α (°)	90
β (°)	117.813(1)
γ (°)	90
V (Å^3)	2057.0(3)
Z', Z	2, 8
D(calc (g cm^{-3}))	1.660

Data Collection

Crystal size (mm)	0.79 x 0.22 x 0.20
Temperature (K)	150(2)
hkl range (h, k, l)	-25→25, -7→7, -26→26
Reflections measured, R_{int}	8568, 0.0155
Independent reflections	2459
Reflections I>2σ(I)	2232

Refinement

Parameters refined	207
R(F) (I>2σ(I))	0.037
wR(F^2) (all reflections)	0.096
Residual electron density (min, max (e Å^{-3}))	0.35, -0.19

Table 3.7: Crystal structure summary for all 5FU crystal structures included in this chapter

Chapter 4 – 5-Fluorocytosine

4.1 Introduction

5-Fluorocytosine (4-amino-5-fluoropyrimidin-2-one, 5FC, figure 4.1) is the 5-fluorinated analogue of the DNA nucleobase cytosine, and has important medical applications. The first reported medical use of 5-fluorocytosine,[181] in 1968, was in the treatment of systemic fungal infections. Since then it has commonly been used in combination with amphotericin B as an anti-fungal treatment.[182] A more recent application is its use as a pro-drug for the delivery of 5-fluorouracil to tumour cells in patients.[183] Delivered in combination with cytosine deaminase, this enzyme converts the 5-fluorocytosine to 5-fluorouracil *in vivo* which is then subsequently converted to pyrimidine anti-metabolites by cellular enzymes. In the intervening 37 years between its first medical use and this study[184] no anhydrous crystal structure has been published; the only structure containing 5-fluorocytosine present in the CSD[23] is that of a monohydrate.[185]
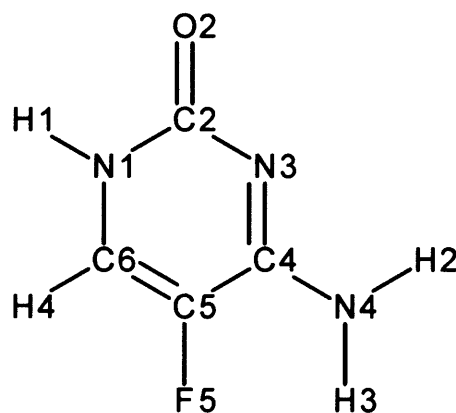
Figure 4.1: Molecular structure and crystallographic numbering of 5-fluorocytosine

5FC was investigated using computational crystal structure prediction to predict possible anhydrous crystal structures, and a simultaneous manual crystallisation screen to discover whether any of these predicted crystal structures could be found experimentally.

4.2 Crystal structure prediction

4.2.1 Energy minimisation of the polymorphs of 5-fluorocytosine

Two polymorphs of 5FC were discovered in the course of the crystallisation screen. The molecular conformation of the 5FC molecule was *ab initio* optimised and the resultant conformation was found to exhibit significant pyrimidisation of the NH_2 group, more pronounced than in any of the molecular conformations found in 5FC experimental crystal structures. This observation, combined with the known uncertainty concerning the degree of pyrimidisation induced into amino groups by *ab initio* optimisation methods,[186] led to a further, constrained, molecular optimisation in which the amino group was constrained to be planar. The crystal structures of both polymorphs were energy minimised with their experimental molecular conformations, the *ab initio* optimised molecular conformer, and the planar constrained conformer.

In all DMAREL[135] energy minimisations a distributed multipole analysis[88] of the MP2/6-31G(d,p) charge distribution was used to model the electrostatic contribution to the intermolecular potential and the dispersion and repulsion contributions were modelled using the FIT[77-79] parameters for C H N O atoms and additional parameters for the fluorine atoms.[80]

For both polymorphs the energy minimisations with the experimental molecular conformations led to excellent reproductions, with F-values under 5 and very small errors in all the lattice parameters. The energy minimisations of form 2 with both *ab initio* and planar molecular conformations produced satisfactory reproductions of the experimental crystal structure, with the largest variation in a cell parameter of 2.9% (table 4.1). The energy minimisation of form 1 with the planar conformation was also successful, returning an F-value of 11 and with a largest cell parameter error of 1.6% (table 4.1).

The energy minimisation of form 1 with the *ab initio* molecular conformation led to significant rearrangement of the molecules, resulting in a poor reproduction of the experimental crystal structure and breaking of the tetragonal symmetry. The differences between the experimental and optimised conformations were small and it was concluded that the reproduction of this crystal structure must be critically dependent on very minor conformational change. A series of constrained *ab initio* optimisations was

performed in an attempt to find the critical difference that caused the lattice energy minimisation with the fully *ab initio* optimised molecular conformation to give such a poor reproduction of the experimental crystal structure. A successful minimisation was achieved by fixing the relative positions of the six atoms in the ring of the molecule, while allowing the peripheral groups (NH_2, F5, O2, H1, H4) to optimise. The lattice energy minimisation using this partially optimised molecular structure (form 1 ring-fixed ExptMinConOpt) led to a satisfactory reproduction with an F-value of 7 and largest cell parameter error of 1.1% (table 4.1). This ring-fixed constrained conformation was as close to the fully *ab initio* optimised conformation as could be found without producing the drastic structural changes detailed above upon energy minimisation.

5-Fluorocytosine	Form 1					Form 2				
	Experimental	ExptMinConOpt ring-fixed	% error	ExptMinConOpt Planar	% error	Experimental	ExptMinOpt *ab initio*	% error	ExptMinConOpt Planar	% error
a (Å)	6.639	6.702	0.96	6.743	1.57	4.063	4.120	1.40	4.083	0.50
b (Å)	6.639	6.702	0.96	6.743	1.57	9.521	9.600	0.84	9.555	0.36
c (Å)	23.471	23.72	1.06	23.709	1.02	12.739	12.844	0.83	13.112	2.93
β (°)	-	-		-		92.99	92.17	-0.87	92.75	-0.25
Volume (Å^3)	1034.3	1065.5	3.00	1078.09	4.22	492.1	507.66	3.16	510.93	3.83
Density (g cm^3)	1.658	1.610	-2.88	1.591	-4.05	1.742	1.689	-3.06	1.678	-3.69
Final Energy (kJ mol^{-1})		-123.87		-115.85			-117.12		-116.99	
F		7.00		10.90			10.64		17.28	

Table 4.1: Summary of the energy minimisation results for the polymorphs of 5FC, using different molecular models. Planar molecular conformations were used for both polymorphs, the fully *ab initio* optimised conformation was also used for form 2 and a conformation with the ring atoms fixed was also used for form 1

4.2.2 Crystal structure prediction – results

The MP2/6-31G(d,p) *ab initio* optimised conformation was used as the input for the MOLPAK[120] search. However, because the NH_2 conformation exhibited pronounced pyrimidisation in the *ab initio* conformer, a supplementary search was carried out using the fully planar molecular conformation. The MOLPAK searches both used 37 packing types that generated structures in 18 common space groups. DMAREL[135] energy minimisation of the resultant structures was performed using the same model potential as used earlier: distributed multipole analysis of the MP2/6-31G(d,p) wavefunction to model the electrostatic term combined with the FIT plus fluorine dispersion-repulsion potential.

The crystal structure search with the *ab initio* optimised molecular conformation generated 33 hypothetical structures within 10 kJ mol^{-1} of the global energy minimum (figure 4.2) and the search with the planar conformation generated 46 hypothetical structures (figure 4.3). Structures that initially minimised to a saddle point had up to two symmetry operators removed sequentially, those that were associated with the most negative eigenvalue from the second derivative matrix. If this did not lead to a true energy minimum the structure was discarded. This resulted in approximately 1/6 of the structures from the *ab initio* search and 1/4 of the structures from the planar search having $Z' > 1$. Included in both scatter plots are the positions of the energy minimised experimental structures, with the corresponding molecular conformation. Summary tables of the low energy structures from both searches including reduced cell parameters, graph set analysis and hydrogen bond information along with the CSP molecular numbering scheme are included in the supporting information.
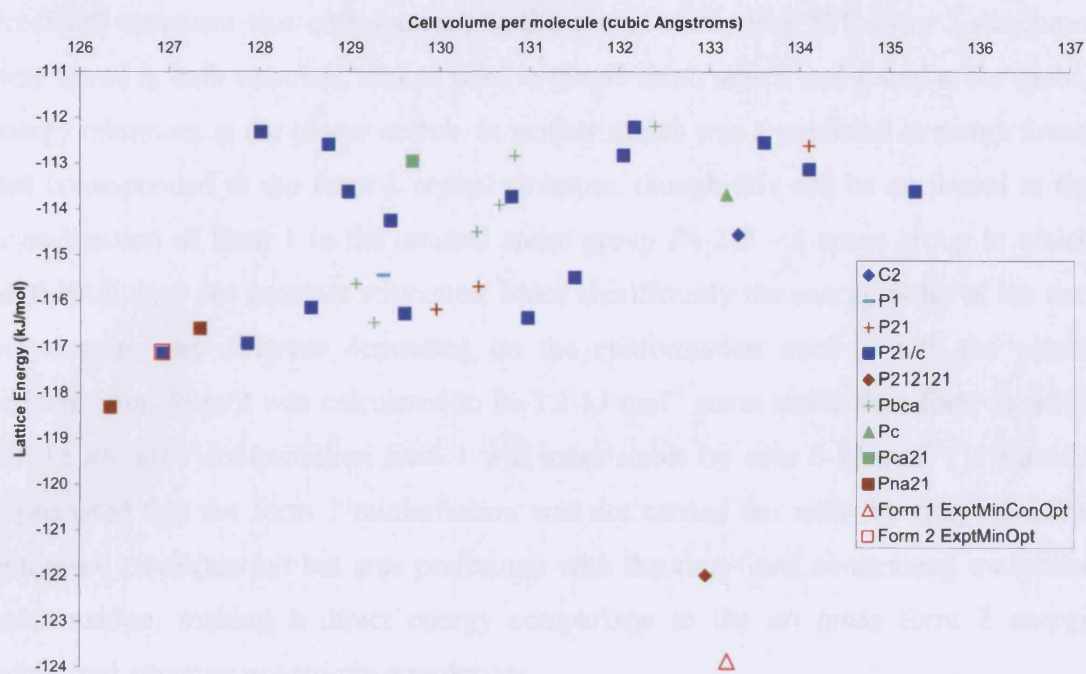
Figure 4.2: Scatter plot of all CSP generated structures for the 5FC *ab initio* conformer within 10 kJ mol⁻¹ of the global energy minimum
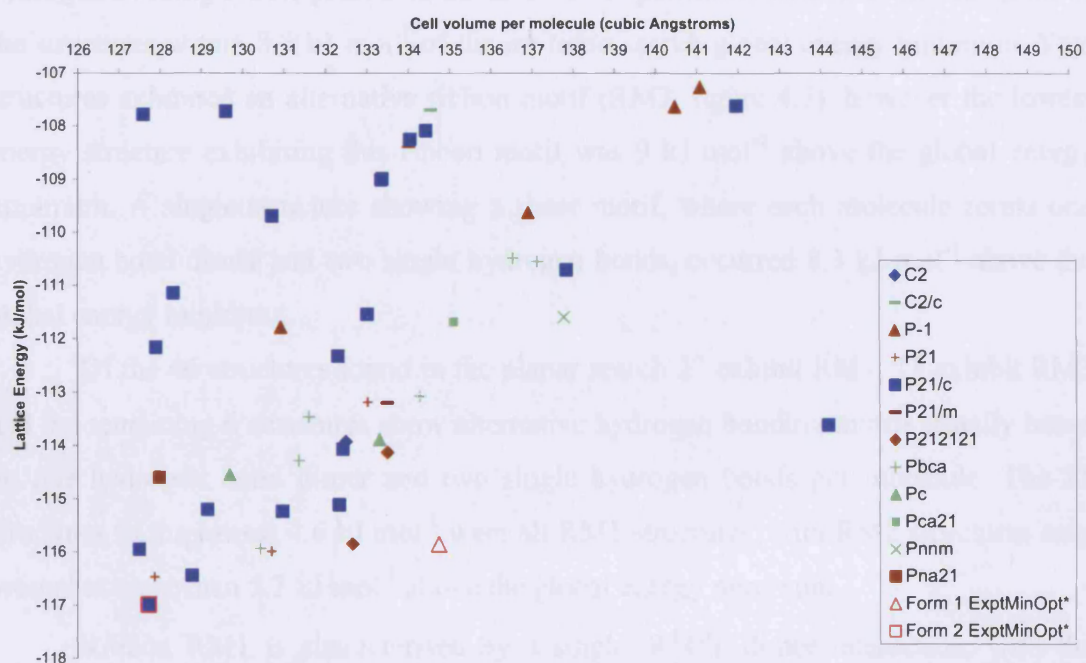
Figure 4.3: Scatter plot of all CSP generated structures for the 5FC planar conformer within 10 kJ mol⁻¹ of the global energy minimum. * Planar conformation

Predicted structures that corresponded to the energy minimised 5FC form 2 structures were found in both searches, ranked third in the *ab initio* search and found at the global energy minimum in the planar search. In neither search was a predicted structure found that corresponded to the form 1 crystal structure, though this can be attributed to the crystallisation of form 1 in the unusual space group $P4_12_12$ – a space group in which MOLPAK does not generate structures. More significantly the energy order of the two polymorphs was different depending on the conformation used – with the planar conformation form 2 was calculated to be 1.1 kJ mol^{-1} more stable than form 1, while for the *ab initio* conformation form 1 was more stable by over 6 kJ mol^{-1}. It must be appreciated that the form 1 minimisation was not carried out with the fully *ab initio* optimised conformation but was performed with the ring-fixed constrained molecular conformation, making a direct energy comparison to the *ab initio* form 2 energy minimised structure not strictly appropriate.

Two different ribbon structures occurred repeatedly within the hypothetical structures from the *ab initio* search. Ribbon motif 1 (RM1, figure 4.4) is the dominant hydrogen bonding motif, present in 30 of the 33 hypothetical structures including all of the structures within 8.3 kJ mol^{-1} of the *ab initio* search global energy minimum. Two structures exhibited an alternative ribbon motif (RM2, figure 4.5), however the lowest energy structure exhibiting this ribbon motif was 9 kJ mol^{-1} above the global energy minimum. A single structure showing a sheet motif, where each molecule forms one hydrogen bond dimer and two single hydrogen bonds, occurred 8.3 kJ mol^{-1} above the global energy minimum.

Of the 46 structures found in the planar search 27 exhibit RM1, 13 exhibit RM2 and the remaining 6 structures show alternative hydrogen bonding motifs usually based on one hydrogen bond dimer and two single hydrogen bonds per molecule. The 21 structures in the lowest 4.6 kJ mol^{-1} were all RM1 structures, with RM2 structures only present at more than 5.2 kJ mol^{-1} above the global energy minimum.

Ribbon RM1 is characterised by a single $R_2^2(8)$ dimer interaction, with the hydrogen bond components N1-H1$\cdots$N3 and N4-H2$\cdots$O2 (figure 4.4).
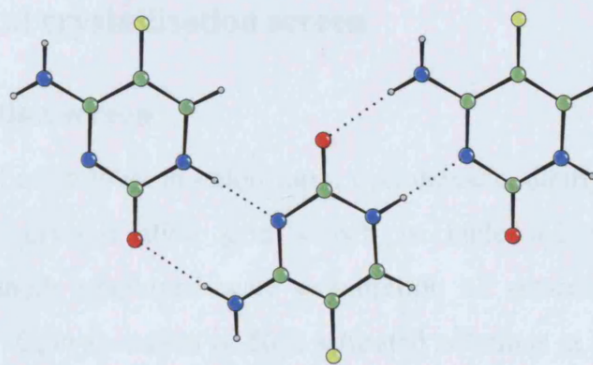
Figure 4.4: Ribbon motif 1, RM1 [first order graph set: $C_1^1(4), C_1^1(6)$; second order: $R_2^2(8)$]

Ribbon RM2 propagates via two different alternating $R_2^2(8)$ dimers: one dimer is comprised of two N4-H2···N3 hydrogen bonds and the other is comprised of two N1-H1···O2 hydrogen bonds (figure 4.5).
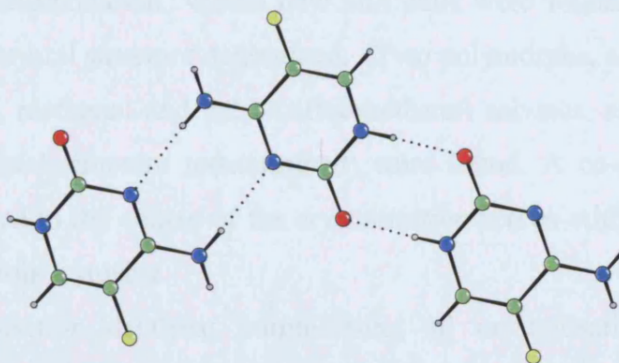
Figure 4.5: Ribbon motif 2, RM2 [first order graph set: $R_2^2(8)a, R_2^2(8)b$]

5FC has an unequal number of hydrogen bond donors and acceptors, with three N-H potential donors and only two acceptors: the carbonyl and aza groups. The three lowest energy structures from the *ab initio* search and the two lowest energy structures from the planar search all form a further 5FC···5FC hydrogen bond, N4-H3···O2 or N4-H3···N3, in addition to the hydrogen bonds in the RM1 ribbon. This contrasts with the majority of predicted structures from both searches in which the final amino hydrogen on each molecule is left unused. The search results clearly show that there is an energetic preference for forming structures that contain RM1, and also for forming inter-ribbon hydrogen bonds to allow all hydrogen bonding groups present in 5FC to be used.

4.3 Experimental crystallisation screen

4.3.1 Crystallisation screen

5FC was found to be insoluble in chloroform, cyclohexane, diethylether, DMF, hexane, and toluene. The crystallisation grid shown in table 4.2 was completed. The crystallisation methods employed were evaporation of saturated solutions at room temperature and 5° C, evaporation of 50% saturated solutions at room temperature and 5° C, vapour diffusion with diethyl ether anti-solvent and vapour diffusion with toluene anti-solvent. A smaller subset of saturated solvent evaporations from aqueous two solvent systems was also carried out. The solvents used were acetone, benzonitrile, ethanol, methanol, 2-butanol and 2-propanol. 1:9, 1:3 and 1:1 water : solvent ratios were used.

Of the 227 crystallisations set up 53 produced crystals that were characterised by SXRD unit cell determination. Where new unit cells were found, full data sets were collected and the crystal structure determined. Two polymorphs, a new monohydrate, a hemipentahydrate, methanol and 2,2,2-trifluoroethanol solvates, as well as the known monohydrate (crystal structure redetermined) were found. A co-crystal monohydrate was also discovered in the course of the crystallisation screen with the minor synthetic by-product 5-fluoroisocytosine.

A crystallisation database, summarising all crystallisations and results, is provided on the supporting information CD (in Chapter_4_5Fluorocytosine) and a summary spreadsheet detailing crystallisation conditions for all of the forms of 5-fluorocytosine is also given (Crystallisation_Summary.xls).

Solvent	EV 100 RT	EV 50 RT	EV 100 5°C	EV 50 5°C	VD toluene	VD ether
Acetone	X	X	X	X		
Benzonitrile	X	X	X	X	X	X
Ethanol	X	X	X	X	X	X
Methanol	X	X	X	X	X	X
2-butanol	X	X	X	X	X	X
2-propanol	X	X	X	X	X	X
1,4-dioxane	X	X	X	X	X	X
1-me-2-pyrrolidinone	X	X	X	X	X	X
2,2,2,-trifluoroethanol	X	X	X	X	X	X
2-chloroethanol	X	X	X	X	X	X
Acetaldehyde	X	X	X	X	X	X
Acetic Acid	X	X	X	X	X	X
Acetonitrile	X	X	X	X	X	X
Diisopropyl ether	X	X	X	X	X	X
Ethylacetate	X	X	X	X	X	X
Ethylene glycol	X	X	X	X	X	X
Methylbenzoate	X	X	X	X	X	X
Nitromethane	X	X	X	X	X	X
o-xylene	X	X	X	X	X	X
Tetrachloroethylene	X	X	X	X	X	X
Tetrahydrofuran	X	X	X	X	X	X
1,2-dichloroethane	X	X	X	X		
Dichloromethane	X	X	X	X		
Dimethylsulfoxide	X					
Formamide	X					
Water	X					

Table 4.2: Experimental crystallisation screen carried out on 5FC. x denotes that this crystallisation was set up; EV = solvent evaproration of 100% and 50% solutions at room temperature and 5°C; VD = vapour diffusion with toluene or diethyl ether

4.3.2 5-Fluorocytosine polymorphs

Two polymorphs of 5FC were discovered in the course of the crystal structure screen.

5-Fluorocytosine form 1

Crystals of form 1 were originally grown by vapour diffusion of diethyl ether into a saturated solution of 2-propanol and the structure was determined by SXRD. Form 1 exhibits a block morphology, usually observed with well-defined faces. It crystallises in the tetragonal space group $P4_12_12$, with a single molecule in the asymmetric unit (figure 4.6 and table 4.4).
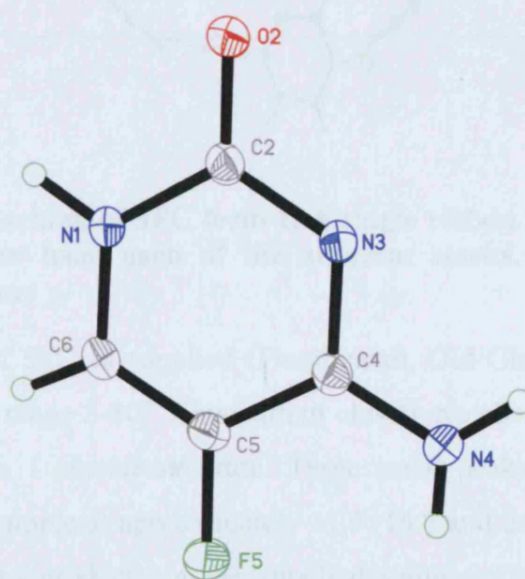
Figure 4.6: Asymmetric unit of 5FC form 1

The structure exhibits RM1 with the ribbons stacking in an ABAB repeat pattern. The planes of the ribbons in adjacent stacks are parallel to different planes – in one the ribbon planes are parallel to the 2 -2 8 Miller planes and in the adjacent stack the ribbon planes are parallel to the -2 -2 8 Miller planes. An inter-ribbon hydrogen bond, N4-H3⋯O2, is also present which satisfies the N-H hydrogen bond donor that was not used in forming the RM1 ribbon. This hydrogen bond incorporates the two dimensional RM1 motif into a three dimensional hydrogen bond network (figure 4.7).
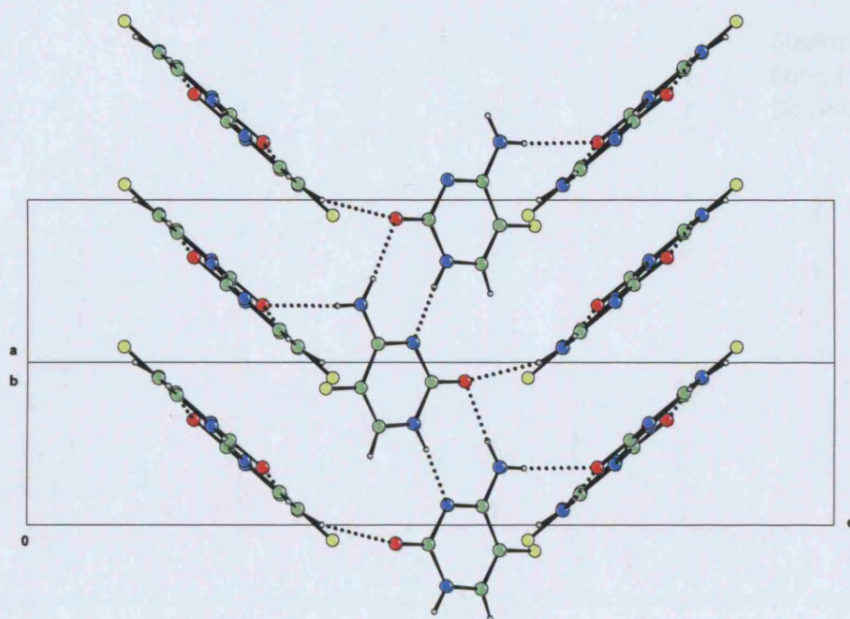
Figure 4.7: Crystal packing of 5FC form 1. A single ribbon from the central stack is included and three from each of the adjacent stacks. View parallel to the diagonal of the *ab* plane

The powder pattern of 5FC as supplied (Fluorochem, Old Glossop, 98% purity) was measured over the 2θ range 5-30°. This pattern closely matched the simulated powder pattern from the form 1 crystal structure. Three small peaks were observed in the commercial material sample at approximately 11.5, 16.7 and 24.8° and are assumed to be due to synthesis by-products in the supplied sample. Crystals of form 1 were reproducibly obtained by sublimation over a period of 30 hours at 200-215 °C under dynamic vacuum (~1 mm Hg), as confirmed by XRPD (figure 4.8).
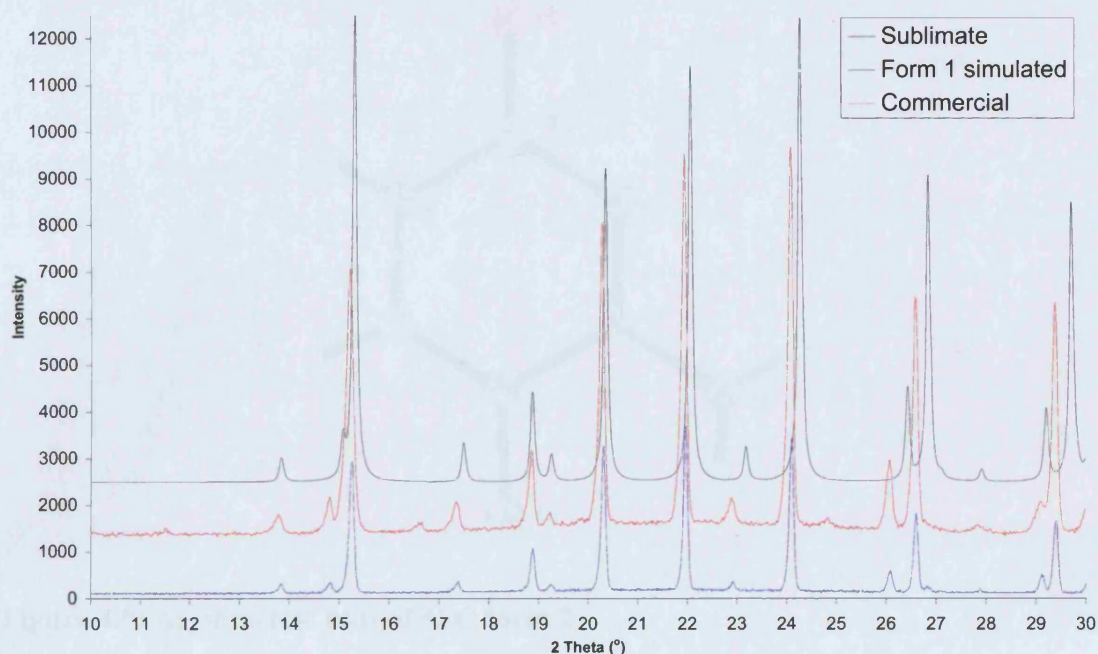
Figure 4.8: XRPD of commercial 5FC (red), the product of sublimation (blue) and the simulated pattern of the form 1 crystal structure (black)

5-Fluorocytosine form 2

A second polymorph of 5FC, form 2, was discovered by solvent evaporation from 5% aqueous 2-propanol. This crystallisation experiment exhibited concomitant polymorphism,[46] with the block crystals in the crystalline product proving to be form 1 and the plate crystals form 2. Form 2 crystallises in the space group $P2_1/n$ with one molecule in the asymmetric unit (figure 4.9 and table 4.4).
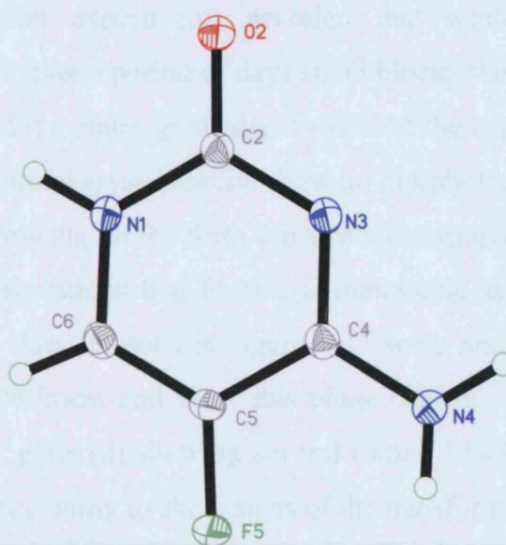
Figure 4.9: Asymmetric unit of 5FC form 2

The crystal structure contains RM1 hydrogen bonded ribbons and the additional hydrogen bond, N4-H3···O2, links the ribbons *via* $R_4^2(8)$ rings into sheets (figure 4.10). These sheets lie parallel to the 1 0 1 Miller planes, stack directly in the structure and have a stepped cross-section. The ribbons are rippled in form 2, rather than the flat topology exhibited by the RM1 ribbons in form 1.
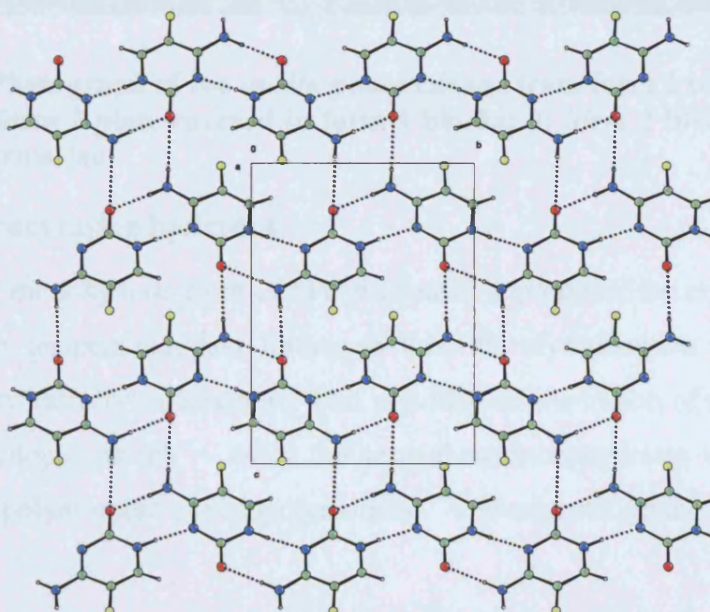
Figure 4.10: The hydrogen bonded sheet structure in 5FC form 2

Subsequent crystallisation experiments revealed that while both forms initially crystallise from solution, over a period of days small blocks start to grow on the surface of the form 2 plates, and the plates gradually dissolve at the expense of the blocks. The form 1 blocks from the initial crystallisation show no morphology change over the same period and the blocks growing on the form 2 plates are assumed to be form 1. It can be deduced from these observations that form 2 is metastable at room temperature with respect to form 1. The three crystals in figure 4.11 were photographed *in situ* in the same crystallisation experiment and show this phase change – the block crystal (1) is form 1, with the form 2 plate (2) showing several form 1 block crystals on its surface. The form 2 plate (3) is beginning to show signs of the transformation.

Figure 4.11. Photograph of the *in situ* phase change from form 2 to form 1. 1: form 1 crystal; 2: form 2 plate covered in form 1 blocks; 3: form 2 block early stage of phase transformation

4.3.3 5-Fluorocytosine hydrates

The published monohydrate form of 5FC was easily reproduced by crystallisation from water at room temperature, and further to this, the crystallisation screen yielded a second monohydrate crystal structure. This provides an illustration of the inadequacy of the term 'pseudopolymorph' – using this term these monohydrates would have to be described as 'polymorphic pseudopolymorphs'. A hemipentahydrate of 5FC was also found.

5-Fluorocytosine monohydrate form 1(h)

The published monohydrate structure, hereafter form 1(h), was redetermined by SXRD at 150 K and this produced a more accurate determination of the structure. The crystals were grown by solvent evaporation from a saturated aqueous solution at room temperature. 5FC form 1(h) crystallises in the monoclinic space group $P2_1/c$ with two 5FC molecules and two water molecules in the asymmetric unit (figure 4.12 and table 4.4).
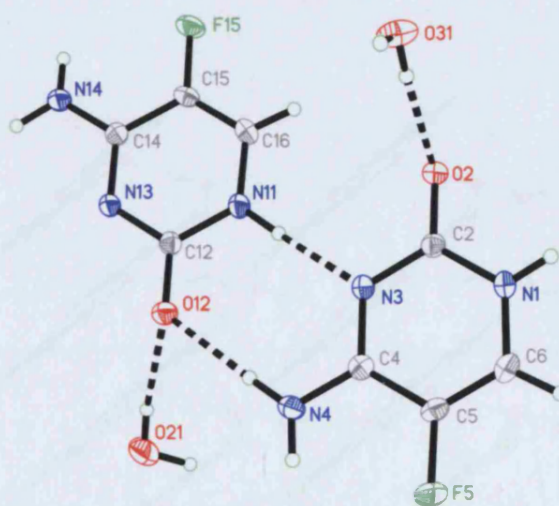
Figure 4.12: Asymmetric unit of 5FC monohydrate form 1(h)

This structure exhibits RM1, with each of the independent 5-fluorocytosine molecules from the asymmetric unit alternating in the ribbon. Each molecule participates in two $R_2^2(8)$ hydrogen bond dimers, with each dimer having the same constituent hydrogen bonds: N1-H1···N13 (or N11-H11···N3) and N14-H12···O2 (or N4-H2···O12). These interactions propagate to form the RM1 ribbons. The ribbons stack directly in columns parallel to the *ab* plane, with a ribbon spacing of approximately 3.2 Å. One water molecule, H21-O21-H22, forms four hydrogen bonds: as a donor to a 5-fluorocytosine molecule (O21-H21···O12), as an acceptor to a second 5-fluorocytosine molecule (N4-H3···O21), as a donor to another water molecule (O21-H22···O31) and as an acceptor to a different water molecule (O31-H31···O21). The second independent water molecule (H31-O31-H32) has a similar hydrogen bonding pattern, acting as a hydrogen bond donor to one 5-fluorocytosine molecule (O31-H32···O2) and as a hydrogen bond acceptor to a second 5-fluorocytosine (N14-H13···O31) along with participating in the

two water⋯water hydrogen bonds described above. The water molecules form discrete cyclic tetramers, denoted R4 in the terminology introduced by Infantes and Motherwell to classify water patterns in hydrate crystal structures.[187] These water tetramers occupy channels in the structure parallel to the *b* axis and hence the structure is a channel hydrate.[11] The water tetramers act as bridges between six 5-fluorocytosine ribbons to form a three dimensional hydrogen bonded network. Adjacent columns of 5-fluorocytosine ribbons have no interactions with each other except *via* the cyclic water tetramers (figure 4.13).
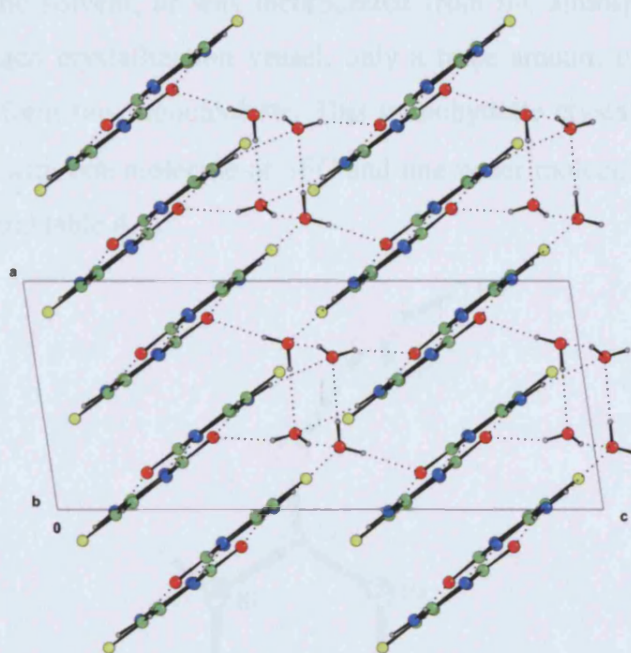
Figure 4.13: Crystal packing of 5FC monohydrate form 1(h). Water tetramers mediate all 5FC ribbon⋯ribbon contacts

DSC and TGA were performed on the sample of 5FC form 1 obtained from sublimation, the commercial material, and the form 1(h) monohydrate. The Form 1 sublimation product and the commercially supplied material decomposed with an onset temperature of 301-302° C with no other events prior to decomposition. The form 1(h) monohydrate showed a single mass loss event of 11.7% of the initial mass at an onset temperature of 99° C. The water in the crystal structure was calculated to comprise 12.3% of the mass and so this event is the loss of the water from the structure. The resulting phase exhibited a sharp decomposition event at an onset of 299° C. XRPD before and after dehydration (vacuum dessication, 72 hrs, over sodium pentoxide)

proved that the form 1(h) monohydrate underwent a phase transition upon dehydration to the form 2 anhydrous structure.

5-Fluorocytosine monohydrate form 2(h)

A new monohydrate structure, hereafter form 2(h), was obtained from a 50% saturated solution of tetrahydrofuran by solvent evaporation at room temperature, and with form 1(h), is a rare example of a pair of polymorphic monohydrates. The crystallisation experiment has no water explicitly contained within it, but residual water may have been present in the solvent, or was incorporated from the atmosphere – as only one crystal grew in each crystallization vessel, only a trace amount of water would have been required to form this monohydrate. This monohydrate crystallises in the triclinic space group $P\bar{1}$, with one molecule of 5FC and one water molecule in the asymmetric unit (figure 4.14 and table 4.4).
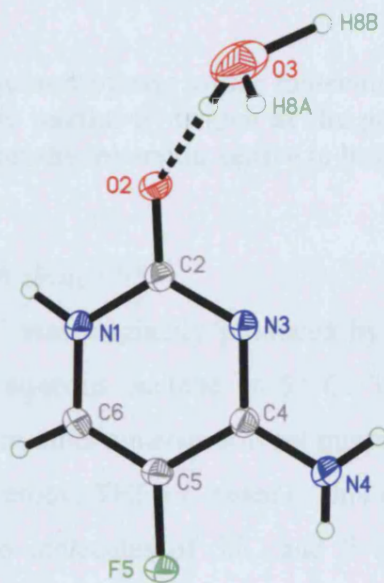
Figure 4.14: Asymmetric unit of 5FC monohydrate form 2(h)

The 5FC molecules adopt the RM2 ribbon, in contrast to all of the other crystal structures reported in this chapter, other than the co-crystal. The ribbons stack as the 0 1 –1 Miller planes. One of the hydrogen atoms in the water molecule is disordered over a general site and the inversion centre at (0, 0.5, 0.5). The other hydrogen atom (H7) in the water molecule is not disordered. The water molecule forms three hydrogen bonds, two to 5-fluorocytosine molecules (O3-H7···O2 and N4-H3···O2) and one to

either of two neighbouring water molecules, O3-H8B···O3 or O3-H8A···O3, depending on the disordered hydrogen occupation (figure 4.15). The structure is a channel hydrate,[11] with the channels parallel to the *a* axis.
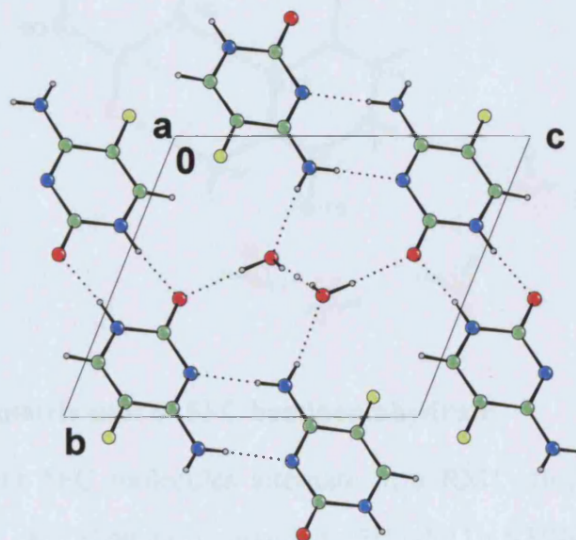
Figure 4.15: Two ribbons joined by two water molecules. In both water molecules full bonds are shown to the partial hydrogen at the general site and dotted lines joins the partial hydrogen on the inversion centre to both water oxygen atoms

5-Fluorocytosine hemipentahydrate (2/5)

A hemipentahydrate of 5FC was originally produced by solvent evaporation from a saturated solution of 25% aqueous acetone at 5° C. The structure has also been subsequently crystallised from other aqueous solvent mixtures, most commonly 25% or 50% water mixtures with acetone, THF or ethanol. This compound crystallises in the space group $P2_1/c$ with two molecules of 5FC and five molecules of water in the asymmetric unit (figure 4.16 and table 4.4).
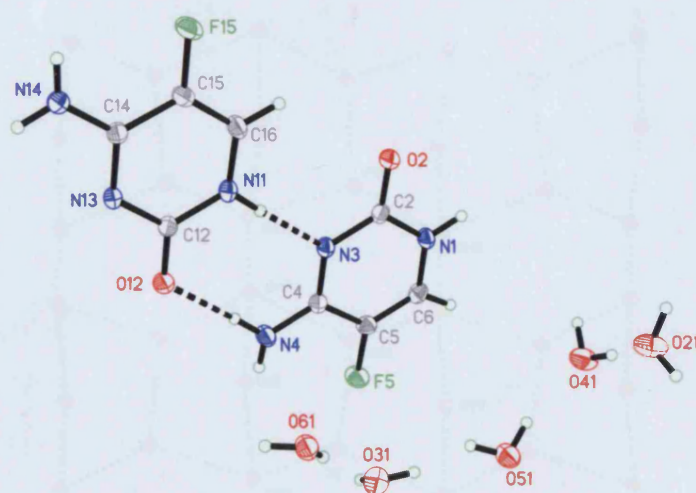
Figure 4.16: Asymmetric unit of 5FC hemipentahydrate

The two independent 5FC molecules alternate in a RM1 structure, with the ribbon propagating by the $R_2^2(8)$ dimer interactions H11-N11···N3/N4-H2···O12 and N14-H12···O2/N1-H1···N13. The structure is a planar hydrate[11] with the water molecules forming sheets parallel to the *bc* plane. Within the sheet, alternating ribbons of hexamer and pentamer units occur, with all units sharing edges. In the terminology of Infantes and Motherwell,[187] the sheet can be seen as comprised of alternating T5(2)5(3) tapes (pentamers) and T6(2) tapes (hexamers), giving an overall layer nomenclature of L5(6)6(10) (figure 4.17). In the crystal structure, the hydrogen atoms have been placed on the water molecules according to the most significant peaks in the electron density map. However there is a degree of disorder in their positions, and consequently figure 4.17 does not include any hydrogen atoms, but shows the overall water hydrogen bonding motif.
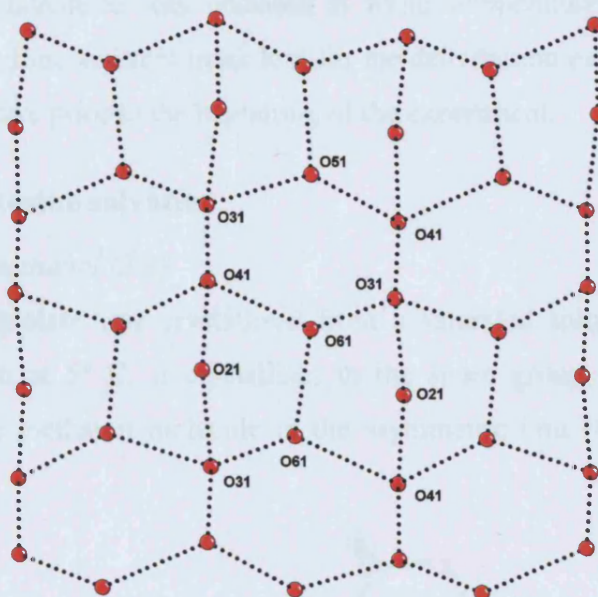
Figure 4.17: Water sheets observed in 5FC hemipentahydrate

The 5FC ribbons stack into columns and the water sheets lie between adjacent columns. Adjacent columns do not come into contact with each other, but both hydrogen bond to the water sheet separating them (figure 4.18). Four crystallographically independent 5FC···water hydrogen bonds are present in the structure.
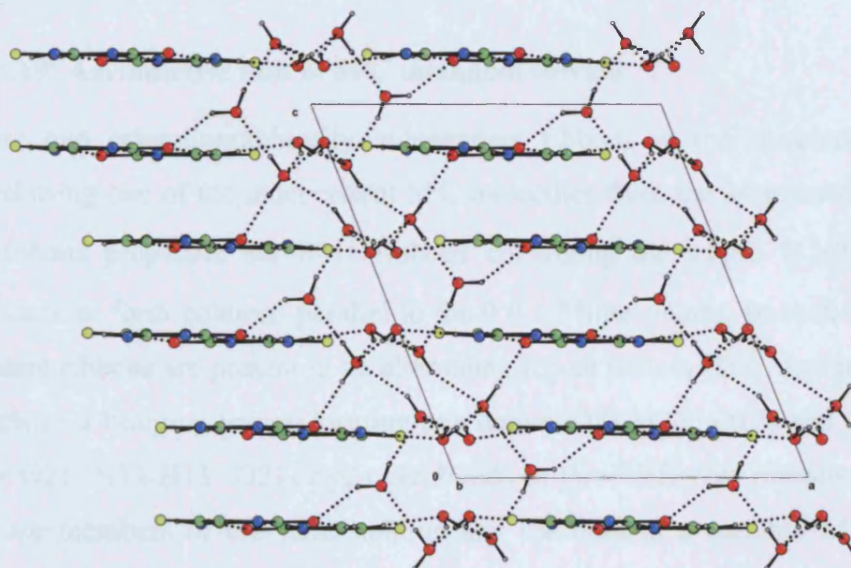
Figure 4.18: Packing diagram for 5FC hemipentahydrate showing a side view of the water sheets packing between 5FC columns

The 5FC hemipentahydrate was unstable at room temperature and repeated TGA experiments showed inconsistent mass loss for the dehydration event, because of water loss from the structure prior to the beginning of the experiment.

4.3.4 5-Fluorocytosine solvates

5-Fluorocytosine methanol (2/1)

A 5FC hemimethanolate was crystallised from a saturated solution of methanol by solvent evaporation at 5° C. It crystallises in the space group $P2_1/n$ with two 5FC molecules and one methanol molecule in the asymmetric unit (figure 4.19 and table 4.4).
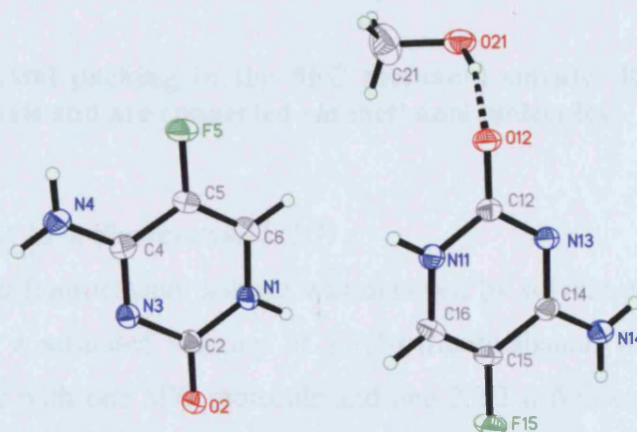
Figure 4.19: Asymmetric unit of 5FC methanol solvate

There are two crystallographically independent ribbons in the structure with each generated using one of the independent 5FC molecules from the asymmetric unit. Both of the ribbons propagate are RM1 ribbons containing the single $R_2^2(8)$ dimer. The ribbons stack to form columns parallel to the 0 0 1 Miller planes. In each column both independent ribbons are present in an alternating repeat pattern. The methanol hydroxyl group acts as a bridging group, forming one donor (O21-H24···O12) and two acceptor (N4-H3···O21; N14-H13···O21) hydrogen bonds to three different ribbons. Two of the ribbons are members of the same column and the third is a member of an adjacent column (figure 4.20).
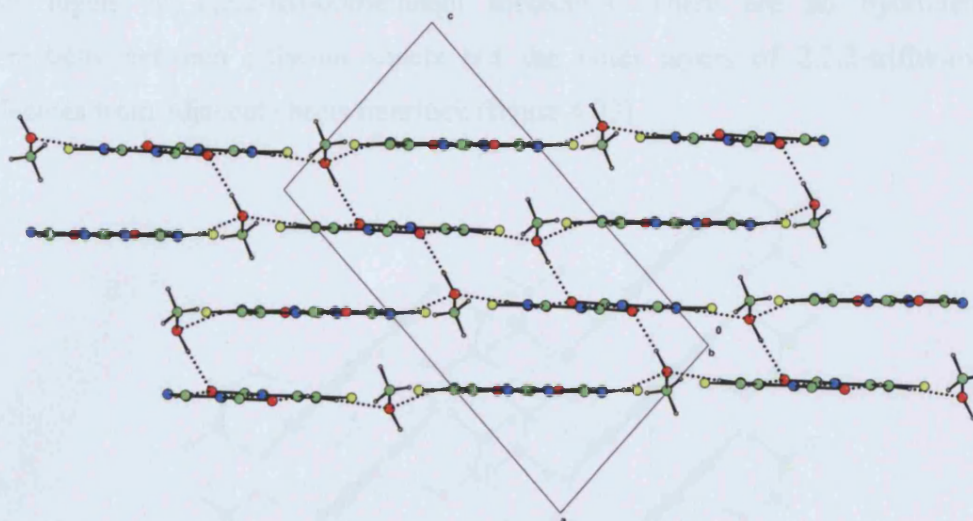
Figure 4.20: Crystal packing in the 5FC methanol solvate. Ribbons propagate parallel to the *b* axis and are connected *via* methanol molecules

5-Fluorocytosine 2,2,2-trifluoroethanol (1/1)

A 1:1 5FC 2,2,2-trifluoroethanol solvate was obtained by solvent evaporation at room temperature from a saturated solution of 2,2,2-trifluoroethanol. It crystallises in the space group $P2_1/c$ with one 5FC molecule and one 2,2,2-trifluoroethanol molecule in the asymmetric unit (figure 4.21 and table 4.4).
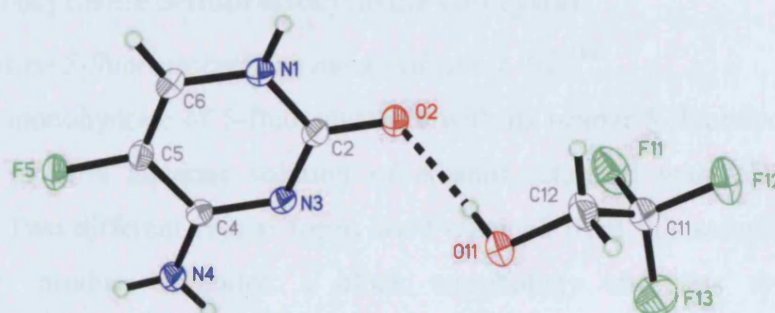
Figure 4.21: Asymmetric unit of 5FC 2,2,2-trifluoroethanol solvate

The 5FC molecules again form the RM1 ribbon, and the ribbons stack in columns with an ABAB repeat pattern parallel to the 1 0 0 Miller planes. The 2,2,2-trifluoroethanol molecules form two hydrogen bonds – acting as an acceptor (N4-H3···O11) and a donor (O11-H13···O2) to 5FC molecules from adjacent ribbons in the same column. These hydrogen bonds generate a sheet, comprised of an inner layer of 5FC molecules and

outer layers of 2,2,2-trifluoroethanol molecules. There are no hydrogen bond interactions between adjacent sheets but the outer layers of 2,2,2-trifluoroethanol molecules from adjacent sheets interlock (figure 4.22).
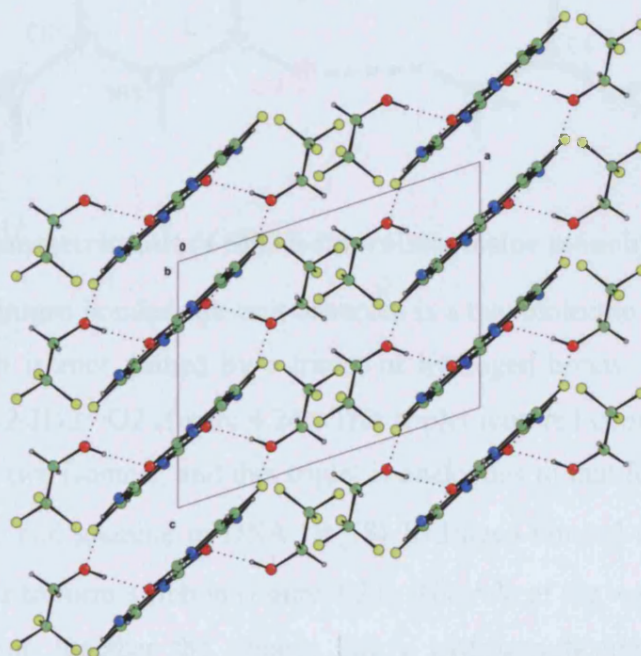
Figure 4.22: 5FC 2,2,2-trifluoroethanol solvate crystal packing. Two columns of 5FC ribbons parallel to the *bc* plane, separated by 2,2,2-trifluoroethanol molecules

4.3.5 5-fluorocytosine 5-fluoroisocytosine co-crystal

5-Fluorocytosine 5-fluoroisocytosine monohydrate (1/1/1)[188]

A co-crystal monohydrate of 5-fluorocytosine with its isomer 5-fluoroisocytosine was grown from a 50% aqueous solution of ethanol saturated with 5FC by solvent evaporation. Two different crystal forms were obtained from this solution. The major crystallisation product exhibited a block morphology and was the form 1(h) monohydrate (identified by SXRD unit cell determination). A small number of lath shaped crystals were observed as the minor crystallisation product. 5-Fluoroisocytosine is assumed to present in the commercial sample as a by-product originating from the synthesis. This co-crystal monohydrate crystallises in the space group $P\overline{1}$ with one molecule of 5FC, one molecule of 5-fluoroisocytosine and a molecule of water in the asymmetric unit (figure 4.23 and table 4.4).
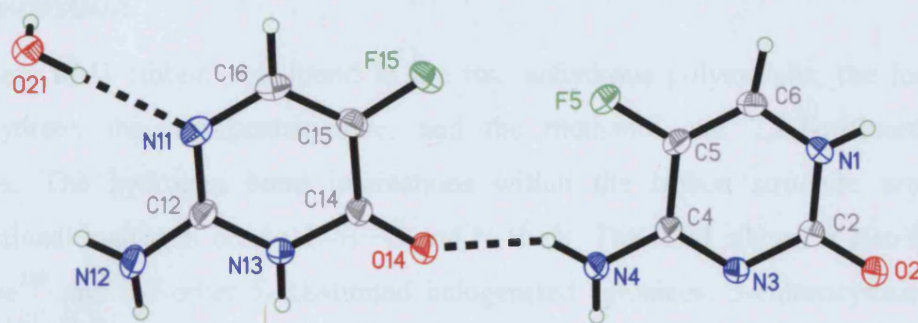
Figure 4.23: Asymmetric unit of 5FC 5-fluoroisocytosine monohydrate

The simplest hydrogen bonded sub-unit observed is a two molecule unit, containing one molecule of each isomer, joined by a triplet of hydrogen bonds: N4-H2···O14, N13-H13···N3 and N12-H12···O2 (figure 4.24). This triplet required complimentary edges to be present in the two isomers, and this triplet is analogous to that found in base pairing between cytosine and guanine in DNA. $R_4^2(8)$ hydrogen bonded rings join the triplet sub-units together to form a ribbon (figure 4.24). The role of the water molecules in the structure is to join together the ribbons into a hydrogen bonded sheet. The water hydrogen bonds to two molecules from one ribbon, acting both as donor and acceptor (O21-H21···N11 and N1-H1···O21) and as a donor to a third molecule from a different ribbon (O21-H22···O2). The ribbons form stepped sheets, parallel to the 0 1 -1 Miller planes. Within the ribbon structure, there is also a close F···F contact, between F5 and F15 of 2.9 Å, however this is likely to have arisen as a consequence of the adjacent $R_4^2(8)$ hydrogen bonded ring.
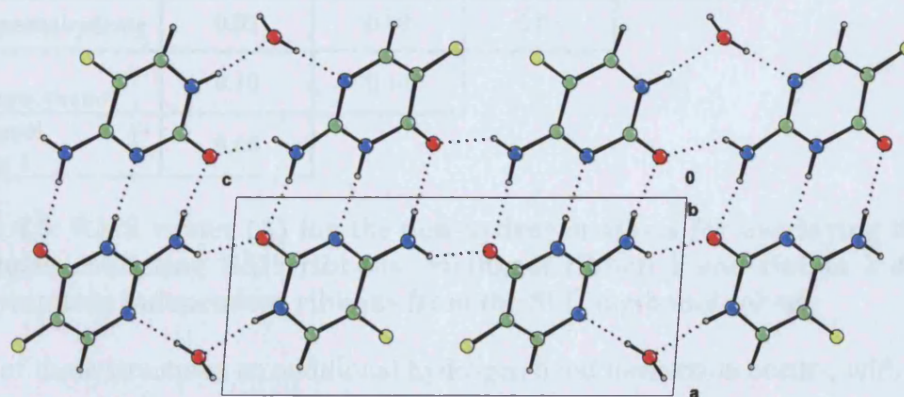
Figure 4.24: Ribbon structure present comprised of 5FC···5-fluoroisocytosine pairs joined by the base pairing triplet interaction

4.4 Discussion

The same RM1 ribbon was found in the two anhydrous polymorphs, the form 1(h) monohydrate, the hemipentahydrate, and the methanol and 2,2,2-trifluoroethanol solvates. The hydrogen bond interactions within the ribbon structure are strong conventional hydrogen bonds, N-H···O and N-H···N. The RM1 ribbon is also found in cytosine[189] and the other 5-substituted halogenated cytosines: 5-chlorocytosine,[190] 5-bromocytosine[190] and 5-iodocytosine.[191] The occurrence of the RM1 ribbon in six of the experimental crystal structures presented here, many related cytosines and the majority of hypothetical structures from the CSP search leads to the conclusion that this is a robust hydrogen bond motif for 5-fluorocytosine.

Root mean square deviation overlays,[132] minimising atom atom distances, of dimers from the RM1 ribbon for all possible pairings of the six structures show (table 4.3) that the difference in the conformation of the ribbons between structures is minimal, with the RMS deviation values ranging from 0.05 to 0.23 Å. The form 2 dimer gives the largest RMS values when overlaid with the other structures because of the pronounced undulation in the ribbon in this structure.

	Methanol ribbon 2	Methanol ribbon 1	2,2,2-trifluoro-ethanol	Hemipenta-hydrate	Form 1(h)	Form 1
Form 2	0.20	0.15	0.23	0.19	0.10	0.17
Form 1	0.06	0.05	0.13	0.10	0.08	
Form 1(h)	0.11	0.07	0.15	0.12		
Hemipentahydrate	0.07	0.09	0.06			
2,2,2-trifluoroethanol	0.10	0.14				
Methanol ribbon 1	0.06					

Table 4.3: RMS values (Å) for the non-hydrogen atoms for overlaying dimers for structures exhibiting RM1 ribbons. Methanol ribbon 1 and ribbon 2 denote the two symmetry independent ribbons from the 5FC methanol solvate

In all of these structures an additional hydrogen bond interaction occurs, with solvent in the lattice or with adjacent 5FC ribbons (in the case of the anhydrous forms) to satisfy the amine hydrogen bond donor on each molecule not used forming the RM1 ribbon. These bonds usually transform the two-dimensional ribbon motif into a three-

dimensional hydrogen bond network. Four solvate structures (monohydrate form 1(h), hemipentahydrate, methanolate, 2,2,2-trifluoroethanolate) also showed similarities in the three-dimensional hydrogen bond network formed. All of these structures pack RM1 ribbons into columns with sheets, layers or channels of solvent between the columns such that the solvent molecules act as bridging units connecting the adjacent columns.

The crystal structure of the form 2(h) monohydrate is unique in that the 5FC molecules propagate using the RM2 motif, which was only present in less stable hypothetical structures. This ribbon is related to RM1 by an approximate two-fold rotation of every second molecule in the ribbon, with the rotation axis running through the centres of the N3-C2 and C5-C6 bonds.

Form 2 crystallises in the space group $P2_1/n$, a space group included in the subset of space groups in which MOLPAK generates structures. Comparison of the form 2 ExptMinOpt structure with the structures from the *ab initio* search revealed that it corresponded to the third lowest energy structure from the search, and in a similar comparison for the planar conformation, the search found the form 2 structure as the global energy minimum structure. By the criteria of the CCDC blind tests of crystal structure prediction,[121-123] the prediction of form 2 was a success as in both searches the structure would be one of three allowed submissions, if selection was based solely on lattice energy. Form 1 could not have been predicted due to the limitations of the search algorithm, and would not have been included in such a blind test, as the space group $P4_12_12$ is sufficiently unusual for organic molecules (305 non-ionic, non-polymeric, organic crystal structures present in CSD[23]).

4.5 Conclusion

After 37 years without a published anhydrous structure, the pharmaceutical 5-fluorocytosine has been found to exhibit polymorphism. In addition to the two anhydrous crystal structures that have been discovered, two new hydrates, two solvates and a co-crystal were obtained. The results of computational crystal structure prediction showed that there was a single dominant hydrogen bonded chain, found in the vast majority of predicted structures. From this observation it was concluded that any anhydrous crystal structures found would exhibit this ribbon motif and would find a method of packing to satisfy the extra N-H hydrogen bond donor functionality not used

in forming the ribbon. This was the case for both of the anhydrous structures and the search was also successful in predicting the exact structure of form 2. Four new solvates and a redetermination of the known monohydrate structure provided additional evidence for the robust nature of the ribbon motif 1, with four of the five exhibiting this motif. It is also noteworthy that the form 2(h) monohydrate, the only solvate found not to contain the ribbon motif 1, exhibited the alternative ribbon motif 2, which was present in a minority of the predicted structures. A co-crystal monohydrate of 5-fluorocytosine with its isomer 5-fluoroisocytosine was also found and its structure determined, which exhibited the same hydrogen bonding as found in DNA base pairing.

Thus the crystal structure behaviour of 5-fluorocytosine can be rationalised as having a strongly preferred two dimensional ribbon structure, which exhibits versatile methods of packing, leading to polymorphism and a number of closely related solvate structures.

Crystal Data

Compound name	5-Fluorocytosine form 1	5-Fluorocytosine form 2	5-Fluorocytosine monohydrate form 1(h)	5-Fluorocytosine monohydrate form 2(h)
Empirical formula	129.1	129.1	147.1	147.1
Formula weight	$C_4H_4N_3OF$	$C_4H_4N_3OF$	$C_4H_4N_3OF, H_2O$	$C_4H_4N_3OF, H_2O$
Crystal system, space group	Tetragonal, $P4_12_12$	Monoclinic, $P2_1/n$	Monoclinic, $P2_1/c$	Triclinic, $P\bar{1}$
a (Å)	6.6387(4)	4.0629(4)	7.3871(6)	4.1026(5)
b (Å)	6.6387(4)	9.5211(9)	9.3940(8)	8.2731(10)
c (Å)	23.471(3)	12.7386(12)	17.5787(15)	9.9191(12)
α (°)	90	90	90	110.036(2)
β (°)	90	92.986(2)	98.608(2)	100.460(2)
γ (°)	90	90	90	96.710(2)
V (Å^3)	1034.4(2)	492.10(8)	1206.1(2)	305.14(6)
Z', Z	1, 8	1, 4	2, 8	1, 2
D(calc (g cm^{-3}))	1.658	1.743	1.620	1.601

Data Collection

Crystal size (mm)	0.25 x 0.23 x 0.19	0.70 x 0.25 x 0.09	0.21 x 0.18 x 0.12	0.41 x 0.17 x 0.13
Temperature (K)	150(2)	150(2)	150(2)	150(2)
hkl range (h, k, l)	-8→8, -8→8, -30→31	-5→5, -12→12, -16→16	-9→9, -12→12, -23→22	-5→5, -10→10, -13→12
Reflections measured, R_{int}	9022, 0.0208	4186, 0.0132	10415, 0.0268	2672, 0.0117
Independent reflections	820	1177	2876	1381
Reflections I>2σ(I)	805	1104	2317	1278

Refinement

Parameters refined	98	98	229	116
R(F) (I>2σ(I))	0.033	0.034	0.046	0.037
wR(F^2) (all reflections)	0.084	0.111	0.120	0.110
Residual electron density (min, max (e Å^{-3}))	0.34, -0.18	0.41, -0.24	0.38, -0.25	0.36, -0.23

Crystal Data

	5-Fluorocytosine hemipentahydrate	5-Fluorocytosine methanol	5-Fluorocytosine 2,2,2-trifluoroethanol	5-Fluorocytosine 5-fluoroisocytosine monohydrate
Compound name				
Empirical formula	$C_4H_4N_3OF$, $2\frac{1}{2}(H_2O)$	$C_4H_4N_3OF$, $\frac{1}{2}(CH_4O)$	$C_4H_4N_3OF$, $C_2H_3OF_3$	$C_4H_4N_3OF$, $C_4H_4N_3OF$, H_2O
Formula weight	174.1	290.3	229.2	276.2
Crystal system, space group	Monoclinic, $P2_1/c$	Monoclinic, $P2_1/n$	Monoclinic, $P2_1/c$	Triclinic, $P\bar{1}$
a (Å)	12.2384(8)	8.4486(9)	11.1490(9)	5.412(2)
b (Å)	9.4254(6)	9.2898(10)	9.5914(8)	8.447(2)
c (Å)	13.8727(9)	16.104(2)	8.5221(7)	12.083(4)
α (°)	90			89.454(5)
β (°)	111.391(1)	97.371(2)	108.139(1)	85.718(5)
γ (°)	90			77.096(4)
V (Å³)	1490.0(2)	1253.5(2)	866.02(12)	536.9(3)
Z', Z	2, 8	2, 8	1, 4	1, 2
D(calc (g cm⁻³))	1.553	1.538	1.757	1.708
Data Collection				
Crystal size (mm)	0.67 x 0.53 x 0.28	0.64 x 0.11 x 0.07	0.47 x 0.29 x 0.19	0.44 x 0.14 x 0.11
Temperature (K)	150(2)	150(2)	150(2)	150(2)
hkl range (h, k, l)	-15→16, -12→12, -17→18	-11→11, -12→12, -21→20	-14→14, -12→12, -10→10	-6→6, -11→10, -15→15
Reflections measured, R_{int}	12658, 0.0168	10796, 0.0306	7461, 0.0170	4532, 0.0178
Independent reflections	3539	3000	2072	2405
Reflections I>2σ(I)	3311	2138	1932	1884
Refinement				
Parameters refined	277	183	164	212
R(F) (I>2σ(I))	0.044	0.066	0.038	0.044
$wR(F^2)$ (all reflections)	0.132	0.162	0.097	0.123
Residual electron density (min, max (e Å⁻³))	0.37, -0.30	0.49, -0.26	0.34, -0.24	0.36, -0.24

Table 4.4: Crystal structure summary for all 5FC crystal structures included in this chapter

Chapter 5 – Coumarin and coumarin derivatives

5.1 Introduction

Coumarin (chromen-2-one) and its substituted derivatives are a naturally occurring class of compounds found in a wide variety of plants,[192] including cinnamon, strawberries and deadly nightshade. 3400 naturally occurring coumarins are known.[193] Substituted derivatives of coumarin have found uses as active ingredients in sunscreens because of their UV-absorbance properties and as precursors to pharmaceuticals. Coumarin itself is responsible for the sweet smell of freshly mown hay. The photoactivity of coumarins has led to their use in laser dyes, in which the lasing medium is a solution of the coumarin in a solvent. The principal drawback of using coumarins in laser dyes is their potential for photodimerisation. Many substituted coumarins, including coumarin itself are photoreactive in solution, dimerising under incident UV light (scheme 5.1).
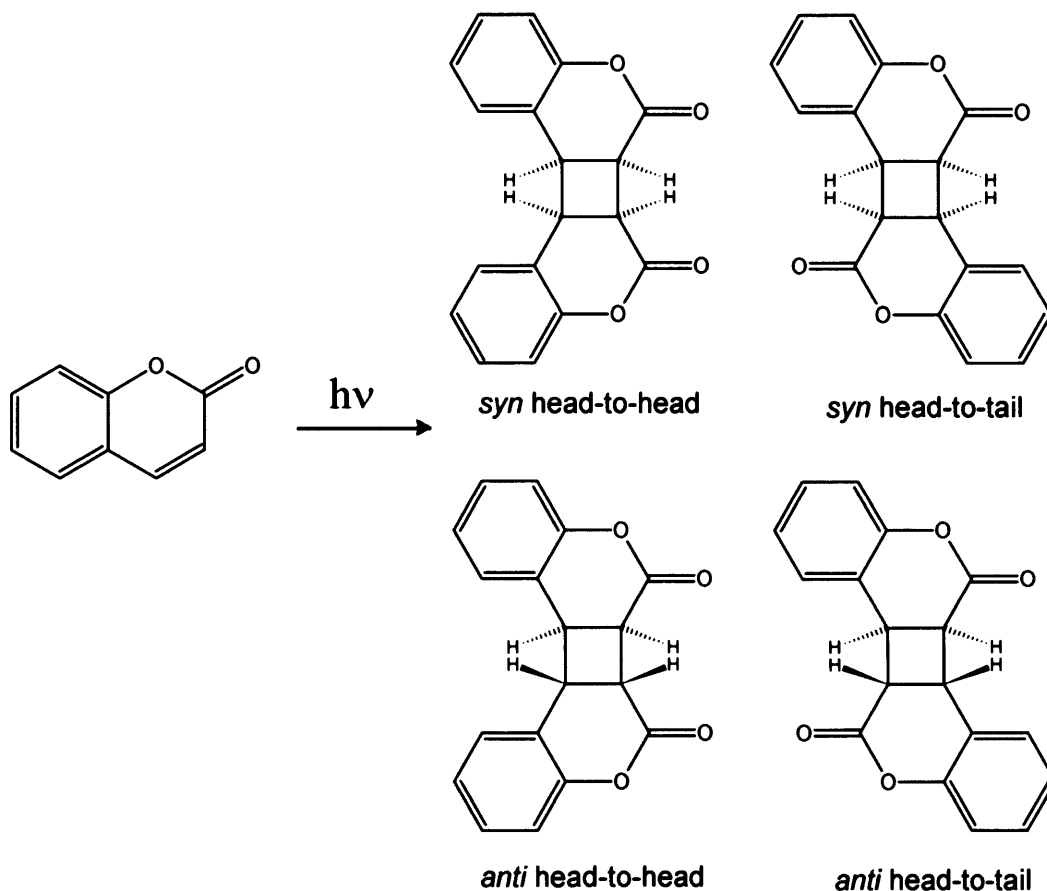
Scheme 5.1: Coumarin photodimerisation. Potential products include both head-to-head and head-to-tail monomer orientations with *syn* or *anti* configurations

The [2+2] photodimerisation of a coumarin in solution can conceivably yield any of four different reaction products, with the two molecules in the dimer orientated either head-to-head or head-to-tail in either *syn* or *anti* configurations (scheme 5.1). For example coumarin is found to give both *syn* and *anti* head-to-head dimers depending on the chosen solvent.[194]

In early crystal engineering work by Schmidt[195] [2+2] photodimerisation in the solid state of *trans*-cinnamic acids was found to proceed only if the molecular packing in the crystal structure allowed the dimerisation. Schmidt defined the 'topochemical principle'[196] that for double bonds to undergo photodimerisation they must be parallel and separated by less than approximately 4.2 Å. Many coumarins have been found to undergo photodimerisation in the solid state, but only if the crystal packing brings molecules into the correct orientation.[197] Indeed it has been noted[198] that many coumarins known to dimerise in solution are photostable in the solid state, such as 4-methoxycoumarin, 7-hydroxycoumarin and 4,7-dimethylcoumarin. In a solid state reaction, the orientation of each pair of molecules that react to form the dimer is the same and a single product will, in principle, exclusively form with the dimerisation product predictable from the crystal structure of the starting compound (figure 5.1). 'Diffusionless'[199] solid state reactions, based on the molecular orientation within the crystal, are an important target for crystal engineering – not only are these reactions solvent free, but because the molecular orientation is well defined in the crystalline state the reactions will be highly specific, in some cases providing 100% yields.[200] Photoreactions in the solid state have also been used to generate chiral products in high enantiomeric excess.[201]
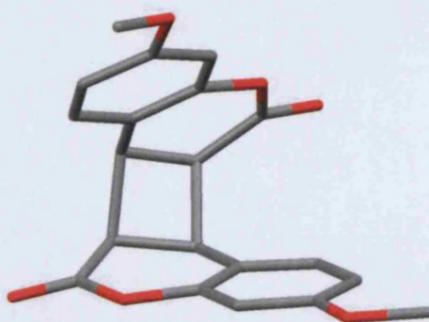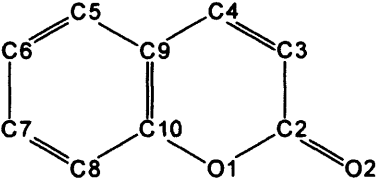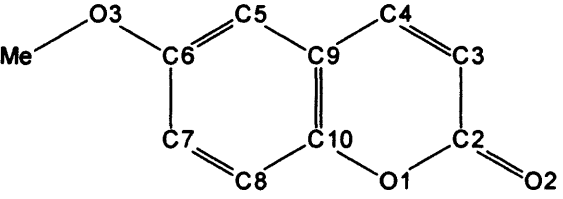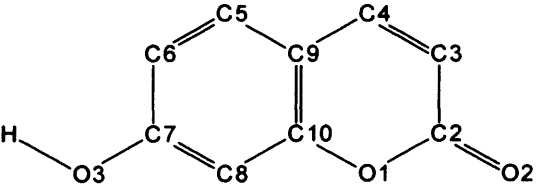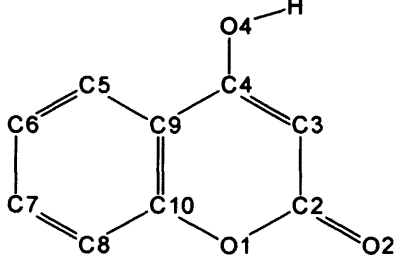
Figure 5.1: 7-methoxycoumarin photodimerises in the solid state to give exclusively the *syn* head-to-tail product[198]

Work was carried out in collaboration with Katharine Bowes (Cambridge University) to predict the possible crystal structures of coumarin and 6-methoxycoumarin as part of a study into the photo-crystallography of excited states of molecules induced by laser or UV light.[202] This collaborative work lead to further, individual investigations on a range of coumarins. 4-Hydroxycoumarin, where only a monohydrate was previously known,[203;204] was investigated using both crystal structure prediction and a manual crystallisation screen to discover anhydrous crystal structures. A limited crystallisation screen was carried out on 6-methoxycoumarin in an attempt to elucidate the full crystal structure of a second reported polymorph only previously identified by unit cell determination.[197] Crystal structure prediction was performed on 7-hydroxycoumarin to assess whether experimental screening would be likely to yield new polymorphs.

The molecular structures of all coumarins investigated in this work, their CSD reference codes and the research completed on them are summarised in table 5.1.

Table 5.1: (next page) Summary table of all work carried out on coumarin systems

Molecular Structure (aromatic hydrogen atoms omitted)	Crystal Structures and CSD reference codes	Research completed	Results
	Coumarin Form 1 COUMAR[01,02,10,11,12]	CSP to determine whether the known structure is the thermodynamic form MOLPAK: 18 space groups, 37 packing types, FIT + SCF-derived multipoles	CSP: known structure found at global energy minimum
	6-Methoxycoumarin Form 1 DAXBIN01 Form 2 DAXBIN (unit cell only)	CSP to determine whether the DAXBIN unit cell corresponds to that of a potential polymorph MOLPAK: 22 space groups, 47 packing types, FIT + SCF-derived multipoles Experimental: 20 crystallisations	CSP: DAXBIN01 found 22nd ranked. No cell found that corresponded to DAXBIN Experimental: DAXBIN crystal structure not found
	7-Hydroxycoumarin Form 1 DETFOX	CSP to determine whether the known structure is the thermodynamic form MOLPAK: 22 space groups, 47 packing types, FIT + MP2-derived multipoles	CSP: known structure found at global energy minimum
	4-Hydroxycoumarin Monohydrate HOQHAW[01] Forms 1–4 (this work, no reference codes)	CSP to predict possible the crystal structures of any anhydrous forms MOLPAK: 22 space groups, 47 packing types FIT + MP2-derived multipoles Experimental: crystallisation screen to find anhydrous forms	CSP: no predicted structures corresponded to the crystal structures of forms 2 &3 found in the experimental screen Experimental: four polymorphs discovered, two fully characterised by SXRD (forms 2 & 3)

5.2 Energy minimisations

Before crystal structure prediction was carried out on coumarin, 6-methoxycoumarin and 7-hydroxycoumarin the known crystal structures were energy minimised to assess the performance of the potential in each case, and for comparison purposes to the search results. The desired potential, the FIT[77;79;84] dispersion-repulsion potential combined with the electrostatic distributed multipole model[88] derived from the MP2/6-31G(d,p) charge density, did not perform well in some cases. Consequently energy minimisations for all structures were also performed with the FIT potential and SCF/6-31G(d,p)-derived multipoles, to assess whether the poor performance could be rectified by altering the electrostatic model. The rationalisation behind trying SCF-derived multipoles was that the FIT empirical parameters were originally derived using potential derived atomic point charges from SCF level calculations for each of the training set molecules. Some of the electron correlation effects explicitly modelled at the MP2 level may have been implicitly absorbed into the FIT empirical parameters at the point of their derivation. Therefore in some cases MP2-derived multipoles may not give superior results to SCF-derived multipoles, despite it being a more realistic representation of the isolated molecular charge distribution.

Summary tables for all energy minimisations discussed below are given at the end of this section.

Coumarin

The crystal structure of coumarin, determined from neutron data by Katharine Bowes,[202] was energy minimised with both the experimental and MP2/6-31G(d,p) *ab initio* optimised conformations using the FIT dispersion-repulsion potential and distributed multipoles derived from both the MP2/6-31G(d,p) charge density and the uncorrelated SCF/6-31G(d,p) charge density. Neither of the MP2-level minimisations, ExptMinExpt(MP2) and ExptMinOpt(MP2), gave satisfactory results yielding F-values of 62 and 126 respectively and having errors greater than 5% in the c axial length (table 5.2). An overlay of the unit cells of the experimental crystal structure and the ExptMinOpt(MP2) minimised structure shows that the molecular orientations and positions did not alter substantially with energy minimisation (figure 5.2). The change

to SCF-level calculation was both significant and beneficial, with the ExptMinExpt(SCF) and ExptMinOpt(SCF) minimisations giving good reproductions of the experimental crystal structure and with F-values less than 10.
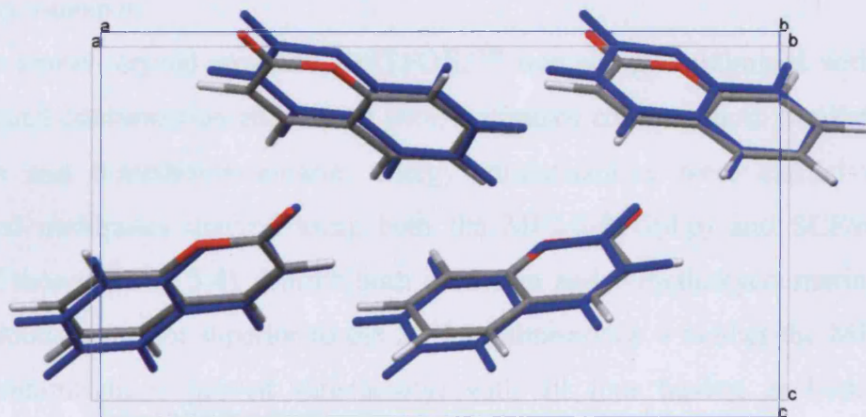
Figure 5.2: Overlay of the experimental coumarin crystal structure (coloured by element) and the ExptMinOpt(MP2) energy minimised structure (coloured blue)

6-Methoxycoumarin

For the purposes of energy minimisation the DAXBIN01[205] crystal structure was manually symmetry reduced from *Pnma* with $Z' = 0.5$ to the sub-group $Pn2_1a$ with $Z' = 1$. This structure was energy minimised using both the experimental and MP2/6-31G(d,p) *ab initio* optimised molecular conformations. The *ab initio* optimisation resulted in a conformation in which the molecular mirror symmetry was retained. Distributed multipoles were derived at both the MP2/6-31G(d,p) and SCF/6-31G(d,p) levels of theory and again the SCF-derived multipoles gave superior energy minimisation results (table 5.3), with the ExptMinOpt(SCF) minimisation returning an F-value of 29. The ExptMinExpt(MP2) minimisation required the removal of a symmetry operator to find a satisfactory energy minimum, thus reducing the space group to $P2_1/c$. The other three energy minimisations retained the *Pnma* symmetry, even though the energy minimisations were performed in $Pn2_1a$.

The *ab initio* optimised molecular conformation from the DAXBIN crystal structure was used as the MOLPAK search input but it is appreciated that there remains the possibility that the alternative molecular conformation, with methyl group rotated

180° with respect to the coumarin skeleton, could also potentially lead to low energy CSP crystal structures, but this was not investigated.

7-Hydroxycoumarin

The only known crystal structure, DETFOX,[206] was energy minimised with both the experimental conformation and the *ab initio* optimised conformation. Similarly to both coumarin and 6-methoxycoumarin, energy minimisations were carried out using distributed multipoles derived using both the MP2/6-31G(d,p) and SCF/6-31G(d,p) levels of theory (table 5.4). Unlike both coumarin and 6-methoxycoumarin, the SCF minimisations were not superior to the MP2 minimisations – neither the MP2 or SCF energy minimisations proved satisfactory, with all four having at least two cell parameters reproduced with an error greater than 5%, and all having F-values in excess of the usual range for successful reproduction. Qualitatively, overlay of the ExptMinOpt(MP2) energy minimised structure with the experimental structure (figure 5.3) showed that the hydrogen bonding had been retained and that the packing of the hydrogen bonded ribbons was similar.
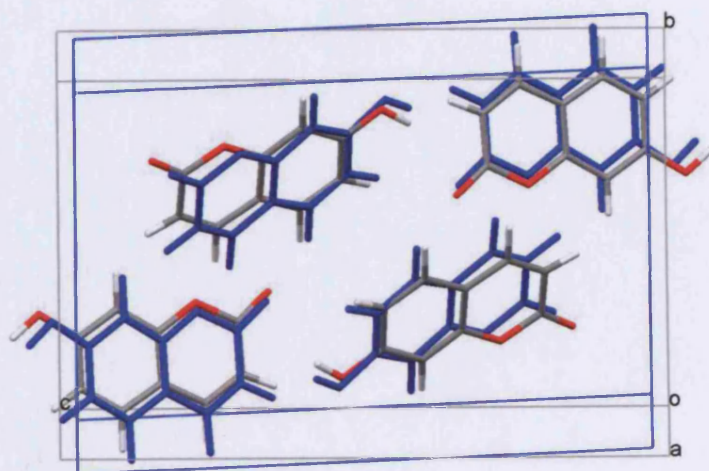
Figure 5.3: Overlay of the 7-hydroxycoumarin experimental structure (coloured by elements) and ExptMinOpt(MP2) (blue) energy minimised structure

For coumarin a MOLPAK search utilising 37 packing types covering 18 space groups was performed. For 6-methoxycoumarin, 7-hydroxycoumarin and 4-hydroxycoumarin, extended MOLPAK searches using 47 packing types spanning 22 space groups were carried out. In all of these searches the MP2/6-31G(d,p) *ab initio* optimised molecular conformation was used as the search input. The reproduction of the coumarin and 6-methoxycoumarin structures by lattice energy minimisation were substantially superior using SCF-derived multipoles and so this electrostatic model was used during the CSP searches on both of these molecules. For 7-hydroxycoumarin neither level of theory gave superior results, and so MP2-derived multipoles were used during the CSP searches because this utilises the higher level of quantum mechanical theory. Based on these results, MP2-derived multipoles were also used in the 4-hydroxycoumarin search. In all searches generated structures that initially energy minimised to a saddle point were symmetry reduced by up to two symmetry operations and if this did not lead to a true minimum, the structure was discarded. For all CSP searches a summary table of the low energy structures, analysis of their hydrogen bonding motifs and a diagram of the atom numbering used during CSP are provided in the supporting information.

Coumarin	Experimental	MP2 ExptMinExpt	% error	MP2 ExptMinOpt	% error	SCF ExptMinExpt	% error	SCF ExptMinOpt	% error
a (Å)	15.479	15.225	-1.64	15.072	-2.63	15.593	0.74	15.619	0.90
b (Å)	5.609	5.528	-1.44	5.464	-2.58	5.648	0.70	5.647	0.69
c (Å)	7.735	8.241	6.54	8.441	9.13	7.839	1.34	7.868	1.71
Volume (Å^3)	671.6	693.63	3.28	695.15	3.51	690.38	2.8	694.04	3.35
Density (g cm^{-3})	1.445	1.399	-3.18	1.396	-3.39	1.406	-2.72	1.399	-3.24
Final Energy (kJ mol^{-1})		-89.95		-88.41		-100.19		-98.18	
F		61.77		125.78		4.52		6.69	

Table 5.2: Energy minimisations of the crystal structure of coumarin

DAXBIN01	Experimental	MP2 ExptMinExpt*	% error	MP2 ExptMinOpt	% error	SCF ExptMinExpt	% error	SCF ExptMinOpt	% error
a (Å)	6.771	6.77	-0.02	6.842	1.06	6.734	-0.55	6.794	0.34
b (Å)	6.454	6.455	0.01	6.458	0.07	6.437	-0.27	6.441	-0.20
c (Å)	18.797	19.565	4.09	19.459	3.52	19.503	3.75	19.428	3.36
α (°)	90	92.18	2.42	90		90		90	
Volume (Å^3)	821.43	854.34	4.01	859.91	4.69	845.28	2.90	850.27	3.51
Density (g cm^{-3})	1.425	1.370	-3.85	1.361	-4.48	1.384	-2.82	1.376	-3.39
Final Energy (kJ mol^{-1})		-102.62		-99.96		-114.75		-112.33	
F		49.57		34.62		36.62		28.98	

Table 5.3: Energy minimisations of the crystal structure of 6-methoxycoumarin. * Structure was symmetry reduced from $Pn2_1a$ to $P2_1/c$ in this minimisation

DETFOX	Experimental	MP2 ExptMinExpt	% error	MP2 ExptMinOpt	% error	SCF ExptMinExpt	% error	SCF ExptMinOpt	% error
a (Å)	3.892	4.125	6.00	4.179	7.37	4.109	5.59	4.170	7.15
b (Å)	11.022	10.98	-0.38	11.093	0.65	10.958	-0.58	11.069	0.43
c (Å)	16.722	15.995	-4.35	15.865	-5.12	15.875	-5.07	15.736	-5.90
β (°)	90.58	86.02	-5.04	84.26	-6.97	85.83	-5.24	84.05	-7.21
Volume (Å^3)	717.30	722.74	0.76	731.78	2.02	712.97	-0.60	722.5	0.73
Density (g cm^{-3})	1.501	1.490	-0.75	1.472	-1.98	1.511	0.61	1.491	-0.72
Final Energy (kJ mol^{-1})		-118.66		-113.48		-136.48		-129.52	
F		117		190		127		206	

Table 5.4: Energy minimisations of the crystal structure of 7-hydroxycoumarin

5.3 Coumarin

5.3.1 Introduction

Coumarin has only been found to crystallise in one crystal form, although there are presently five independent determinations of this structure in the CSD,[207-211] along with the redetermination of this structure from powder neutron data, by Katharine Bowes, that formed the basis of this work. Coumarin crystallises in the orthorhombic space group $Pca2_1$ with one molecule in the asymmetric unit. There are no strong hydrogen bond donors but a ribbon motif is formed by two weak C-H···O hydrogen bonds, bifurcated at O2 (figure 5.4A). Coumarin is known to be photostable in this crystal structure as expected from the molecular orientation – the double bonds in the two molecules are situated 3.9-4.0 Å apart, but the torsion angle between the double bonds in the two molecules is 100° (figure 5.4B).[197]
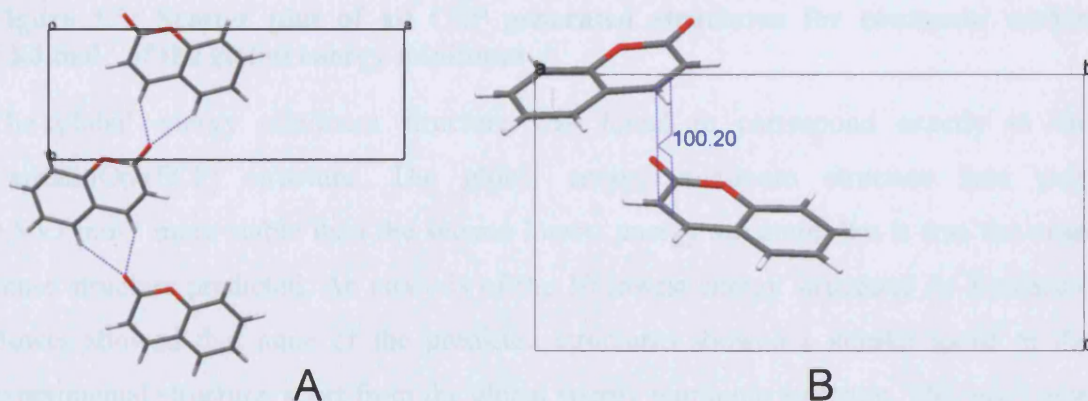
Figure 5.4: The crystal structure of coumarin. A: Weak C-H···O hydrogen bonding present in the crystal structure of coumarin; B: Orientation of the two molecules closest in orientation to a photodimerising pair

5.3.2 Crystal structure prediction

The CSP search yielded a large number of structures (260) within 10 kJ mol⁻¹ of the global energy minimum and 53 structures within 5 kJ mol⁻¹ (figure 5.5). All structures within 5 kJ mol⁻¹ were examined after the search for their true space group using the ADDSYM algorithm contained in PLATON.[139]
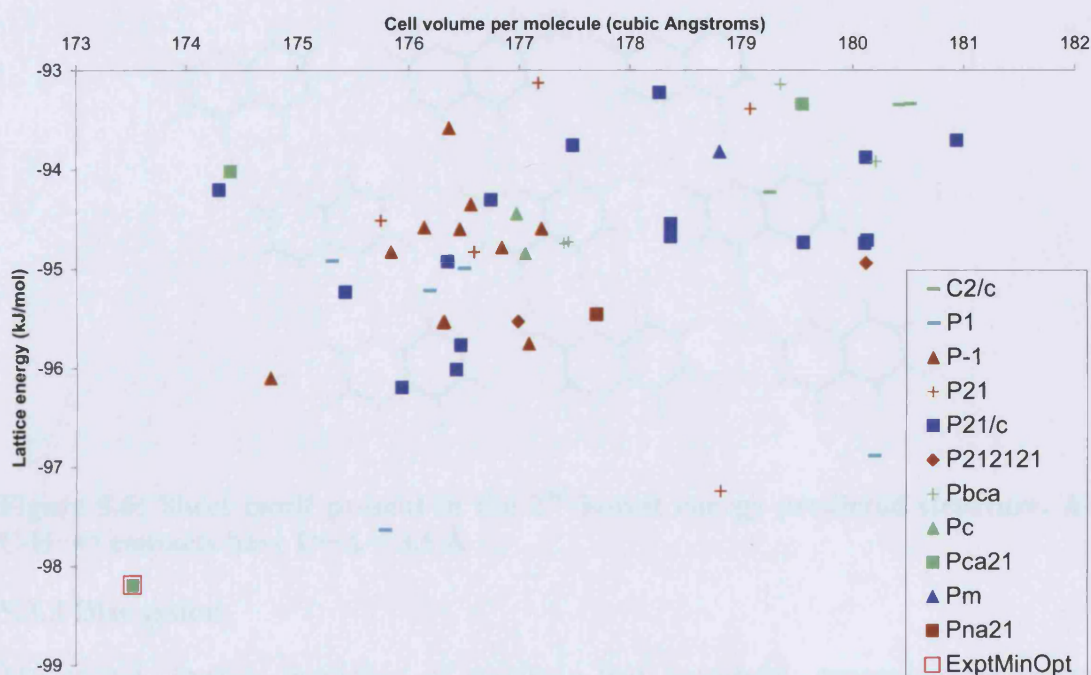
Figure 5.5: Scatter plot of all CSP generated structures for coumarin within 5 kJ mol^{-1} of the global energy minimum

The global energy minimum structure was found to correspond exactly to the ExptMinOpt(SCF) structure. The global energy minimum structure was only 0.5 kJ mol^{-1} more stable than the second lowest energy structure, but it was the most dense structure predicted. An analysis of the 10 lowest energy structures by Katharine Bowes showed that none of the predicted structures showed a similar motif to the experimental structure, apart from the global energy minimum structure. The other nine predicted structures all contained C-H⋯O weak hydrogen bonds, but in a variety of motifs, including sheets and ribbons that utilised the oxygen atom in the heterocyclic ring as a hydrogen bond acceptor as well as the carbonyl oxygen (figure 5.6).

Of the 10 lowest energy predicted structures the second, third and fourth lowest energy structures had the molecules orientated correctly for potential [2+2] photodimerisation, though the distances between the molecules in all three structures were between 4.125 and 4.469 Å. These distances are at Schmidt's upper limit for photodimerisation, though it has been noted that photodimerisation of 7-chlorocoumarin occurs at a molecular separation of 4.454 Å.[197]
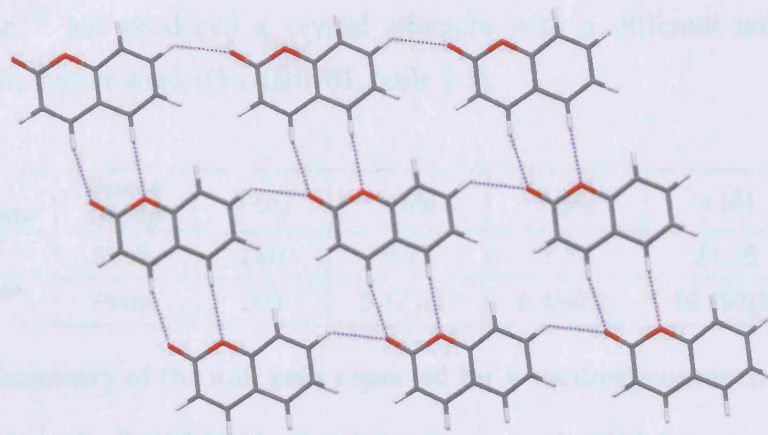
Figure 5.6: Sheet motif present in the 2nd lowest energy predicted structure. All C-H···O contacts have D···A < 3.6 Å

5.3.3 Discussion

The crystal structure prediction of coumarin was successful, generating the known crystal structure at the global energy minimum with the highest density of any predicted structure. The presence of five determinations in the CSD, plus the neutron diffraction redetermination used for this work, suggests that this crystal structure is the only easily obtainable crystal structure for coumarin, consistent with the CSP results.

The known coumarin crystal structure does not contain the constituent molecules in the correct orientation for photodimerisation. However, if a metastable crystal structure could be kinetically trapped that corresponded to one of the second, third or fourth lowest energy structures then photodimerisation to give the *syn* head-to-head dimer could be possible.

5.4 6-Methoxycoumarin

5.4.1 Introduction

Relatively little is known about the chemical and physical properties of 6-methoxycoumarin. It is known to photodimerise in solution, to produce the *anti* head-to-head photodimer.[212] A unit cell was reported for 6-methoxycoumarin in the course of a photodimerisation study of a range of coumarins in the solid state[197] (DAXBIN, table 5.5), though the full structure was not determined. This crystal structure was found to be photoreactive, producing *syn* head-to-head dimers. More recently an SXRD structure

determination[205] has produced a crystal structure with a different unit cell to that reported in the earlier work (DAXBIN01, table 5.5).

CSD reference code	Space Group	T (K)	a (Å)	b (Å)	c (Å)	V (Å³)
DAXBIN[197]	$Pca2_1$	298	6.73	7.17	21.26	1026
DAXBIN01[205]	$Pnma$	203	6.771(2)	6.454(2)	18.797(5)	821.4(4)

Table 5.5: Summary of the unit cells reported for 6-methoxycoumarin

Each molecule in the DAXBIN01 crystal structure occupies 205 Å³ compared to 256 Å³ for the DAXBIN unit cell. The extra volume in the latter unit cell could correspond to a solvate, but the solvent would have to be small, such as a dihydrate or a methanolate, or be a hemi-solvate with a larger solvent. Unfortunately no information is given in the original paper about the crystallisation that yielded the DAXBIN unit cell.

The DAXBIN01 crystal structure adopts the orthorhombic space group $Pnma$ with the molecule in the asymmetric unit located on the mirror plane (Z' = 0.5, four molecules in the unit cell). Like coumarin, 6-methoxycoumarin cannot form strong hydrogen bonds, but again weak C-H···O hydrogen bonds are present in the structure. The two long edges of the 6-methoxycoumarin molecule are complimentary for forming C-H···O interactions[205] with one edge presenting groups in the order donor-acceptor-acceptor and the other acceptor-donor-donor. This allows molecules to form triplets of approximately parallel C-H···O interactions. Two adjacent ribbons propagated by this triplet interaction form the sides of a ladder, with the rungs of the ladder consisting of a further C-H···O weak hydrogen bond (figure 5.7).
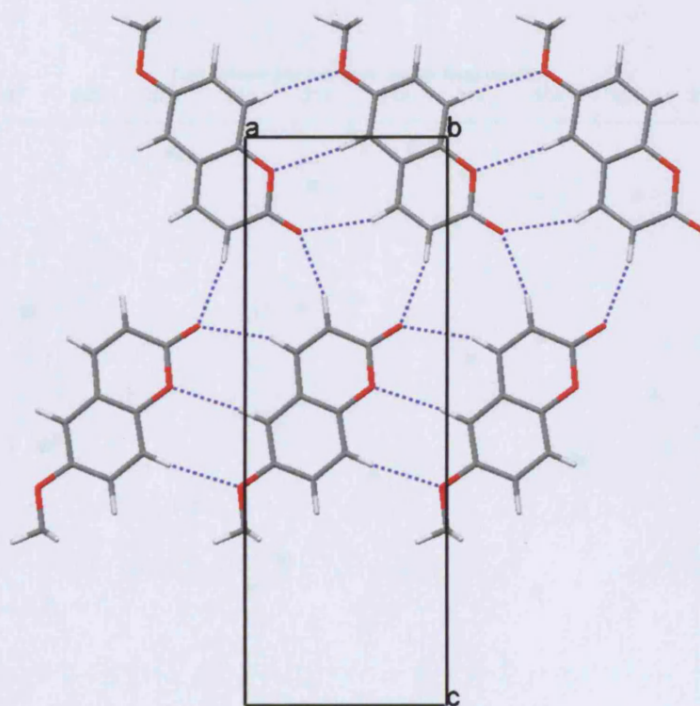
Figure 5.7: Ladder motif present in the 6-methoxycoumarin DAXBIN01 crystal structure. The ladder runs from left to right. All C-H···O hydrogen bonds have D···A < 3.5 Å

5.4.2 Crystal structure prediction

The CSP search produced 88 structures within 10 kJ mol^{-1} of the global energy minimum and 26 structures within 7 kJ mol^{-1} (figure 5.8). All structures within 7 kJ mol^{-1} were examined after the search for extra symmetry using the ADDSYM algorithm contained in PLATON.[139]
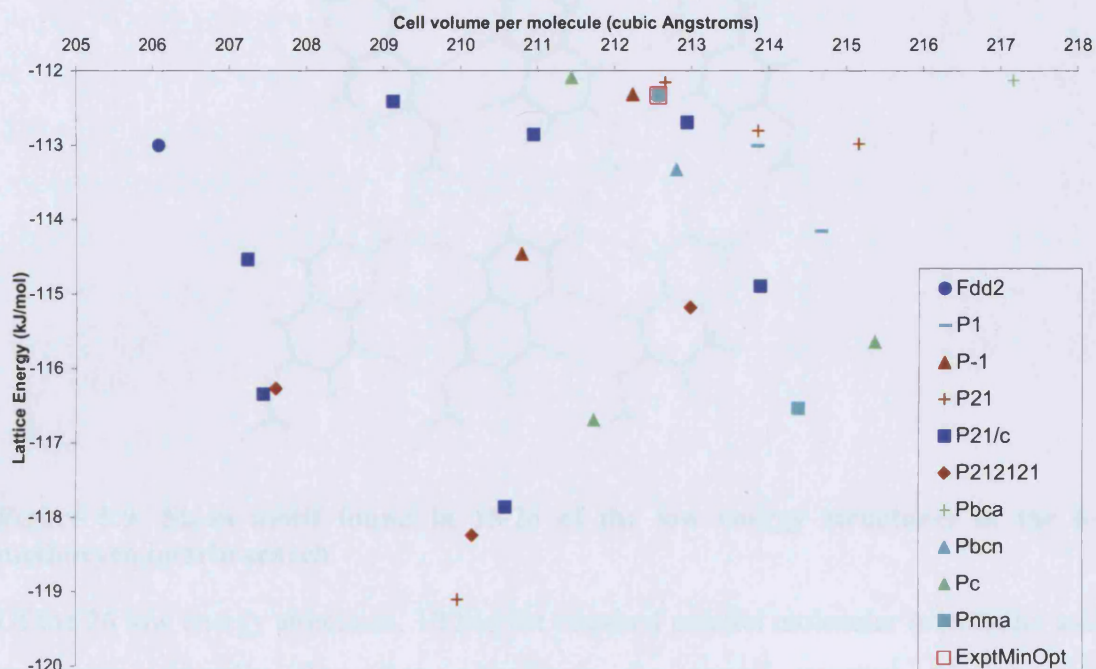
Figure 5.8: Scatter plot of all CSP generated structures for 6-methoxycoumarin within 7 kJ mol^{-1} of the global energy minimum

The CSP search generated a structure that corresponded exactly to the ExptMinOpt(SCF) minimised experimental crystal structure. However, this structure was only 22nd lowest in energy, 6.7 kJ mol^{-1} higher in energy than the global energy minimum. The structure was generated in the space group $P2_12_12_1$, and the symmetry corresponding to the space group *Pnma* was found after energy minimisation during the PLATON analysis.

All 26 structures within 7 kJ mol^{-1} were analysed for their C-H$\cdots$O hydrogen bonding motifs, and for the presence of molecular orientations that could potentially photodimerise. Of the 26 structures no less than 20 contained the same triplet interaction as that observed in the experimental structure, and in all cases the triplet interaction propagated the same ribbon. In 18 of the 20 structures that contained this ribbon, adjacent ribbons formed C-H$\cdots$O contacts *via* the methyl group, producing sheets in 15 instances (figure 5.9) and a 3D motif in the other 3 instances. The other two structures that contained triplet ribbons, but did not include methyl$\cdots$carbonyl interactions, were the structure that corresponded to the experimental structure and a structure closely related in energy and motif.

171
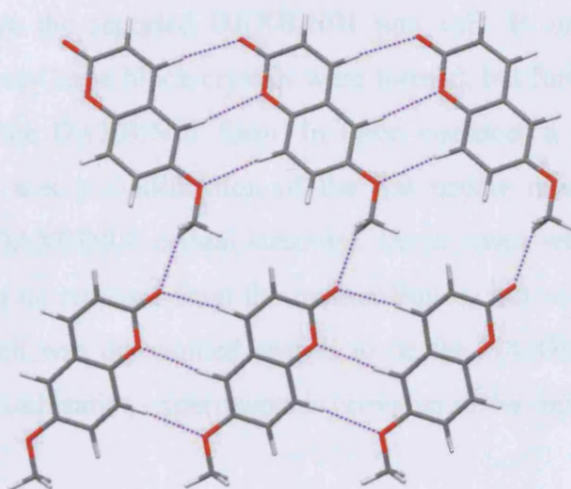
Figure 5.9: Sheet motif found in 15/26 of the low energy structures in the 6-methoxycoumarin search

Of the 26 low energy structures, 10 had the required parallel molecular orientation and molecular separation (less than 4.2 Å) in the crystal structure for potential photodimerisation, all of which would give the *syn* head-to-head dimer geometry. Interestingly there is no correspondence between the hydrogen bond motif present in the structure and the presence of molecules in the correct orientation for photodimerisation. The DAXBIN01 crystal structure does not have molecules in the correct orientation for photodimerisation, in contrast to DAXBIN, whose crystal structure is unknown, but which was reported to photodimerise.[197]

5.4.3 Experimental crystallisation screen

6-Methoxycoumarin was purchased from APIN Chemicals (Abingdon, 98% purity). The XRPD diffraction pattern of the commercial material was measured and compared to the simulated XRPD pattern from the DAXBIN01 single crystal structure, which it matched closely. A series of solvent evaporation crystallisations were carried out at room temperature using 20 common solvents chosen because any of them could have potentially been the solvent used for the crystallisation that yielded the DAXBIN crystals. In each crystallisation 10 mg of 6-methoxycoumarin was dissolved in 5 ml of each solvent. From each solution the solvent was allowed to evaporate slowly to give the 6-methoxycoumarin crystallisation product. A flat needle morphology was observed 14 times. In six of these cases the crystals were suitable for unit cell determination by

SXRD and all gave the reported DAXBIN01 unit cell. In one instance, from tert-butylmethylether, very large block crystals were formed, but further investigation again proved this to be the DAXBIN01 form. In three instances a plate morphology was observed, but this was a modification of the flat needle morphology, which again proved to be the DAXBIN01 crystal structure. In no cases were crystals grown that showed desolvation on removal from the mother liquor, and none of the nine samples for which a unit cell was determined proved to be the DAXBIN crystal structure. A summary of the crystallisation experiments is provided in the supporting information.

5.4.4 Discussion

The results of crystal structure prediction on 6-methoxycoumarin show that none of the predicted structures have a calculated volume per molecule close to that of the DAXBIN unit cell – the range from the predicted structures is 206-217 Å^3 compared to 256 Å^3 calculated for DAXBIN. This result lends weight to the assertion that DAXBIN corresponds to a solvate. The DAXBIN01 crystal structure was found in the search 6.7 kJ mol^{-1} higher in energy than the global energy minimum structure. However the majority of the predicted structures did contain the same ribbon hydrogen bond motif, comprised of the same triplet of interactions between adjacent molecules, although the ribbon was usually part of a hydrogen bonded sheet rather than the ladder observed in the DAXBIN01 structure.

The limited series of crystallisations carried out on 6-methoxycoumarin did not yield a crystal structure with a unit cell corresponding to that reported for DAXBIN, making it impossible to confirm the inference from the molecular volume comparison that it is a solvate. It is interesting to note the large difference in morphology with solvent. The flat needle or plate morphology found in the majority of 6-methoxcoumarin crystallisations would not be advantageous in pharmaceutical manufacturing processes such as filtration steps. However, if tert-butylmethylether was used as the crystallisation solvent of choice, it would produce block shaped crystals – the optimal crystal morphology.

5.5 7-Hydroxycoumarin

5.5.1 Introduction

7-hydroxycoumarin (umbelliferone) is found in a wide variety of plants and has found application as a fine chemical used in sunscreens to absorb ultraviolet radiation. One crystal structure has been determined for 7-hydroxycoumarin,[206] found in the space group $P2_1/c$ with one molecule in the asymmetric unit (CSD reference code DETFOX). The DETFOX crystal structure is made up of chains with a V-shaped profile, in which the constituent molecules form O-H⋯O hydrogen bonds. Comparing the C-O and C=O bonds directions across the hydrogen bond, adjacent molecules have an *anti* orientation (figure 5.10). The chains stack directly upon one another when viewed parallel to the *a* axis.
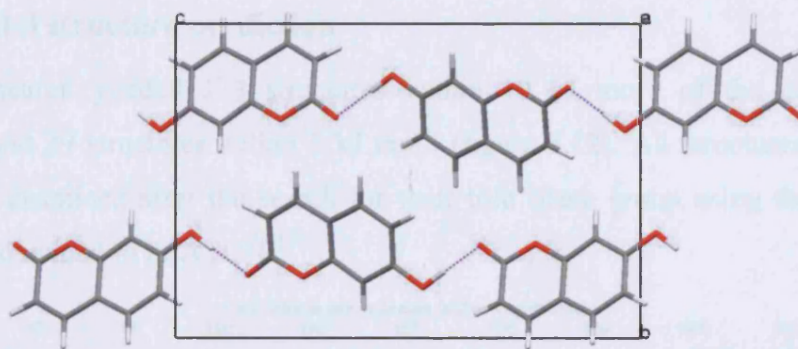
Figure 5.10: Hydrogen bonding motif present in 7-hydroxycoumarin. Adjacent molecules in the chain have an *anti* orientation

It has been reported that 7-hydroxycoumarin does not to photodimerise.[197] In the DETFOX crystal structure the length of the *a* axis (3.892(1) Å) is the separation between directly stacked parallel molecules (figure 5.11), well within Schmidt's upper limit.[195] The stacking of the molecules is offset to produce attractive π-stacking,[213] but this offset stacking could possibly cause the *p*-orbitals comprising the double bond between atoms C3=C4 to overlap insufficiently for [2+2] photodimerisation to occur. However the initial report of photostability does not identify the specific crystal structure tested, so it is conceivable that it was a different crystal structure to that of DETFOX. Consequently, the DETFOX 7-hydroxycoumarin crystal structure should not be discounted as incapable of photodimerisation, and the possibility that the photostable structure is a different polymorph should be considered.
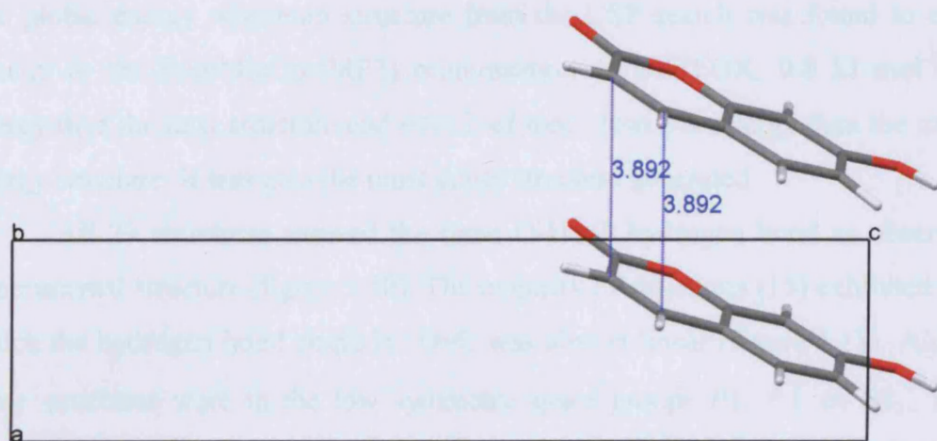
Figure 5.11: A pair of molecules in the 7-hydroxycoumarin structure. The C3=C4 double bonds are parallel and close enough for photodimerisation but the inclination of the molecules may not allow it. The intermolecular distance is shown

5.5.2 Crystal structure prediction

The CSP search yielded 173 structures within 10 kJ mol^{-1} of the global energy minimum and 29 structures within 5 kJ mol^{-1} (figure 5.12). All structures within 5 kJ mol^{-1} were examined after the search for their true space group using the ADDSYM algorithm contained in PLATON.[139]
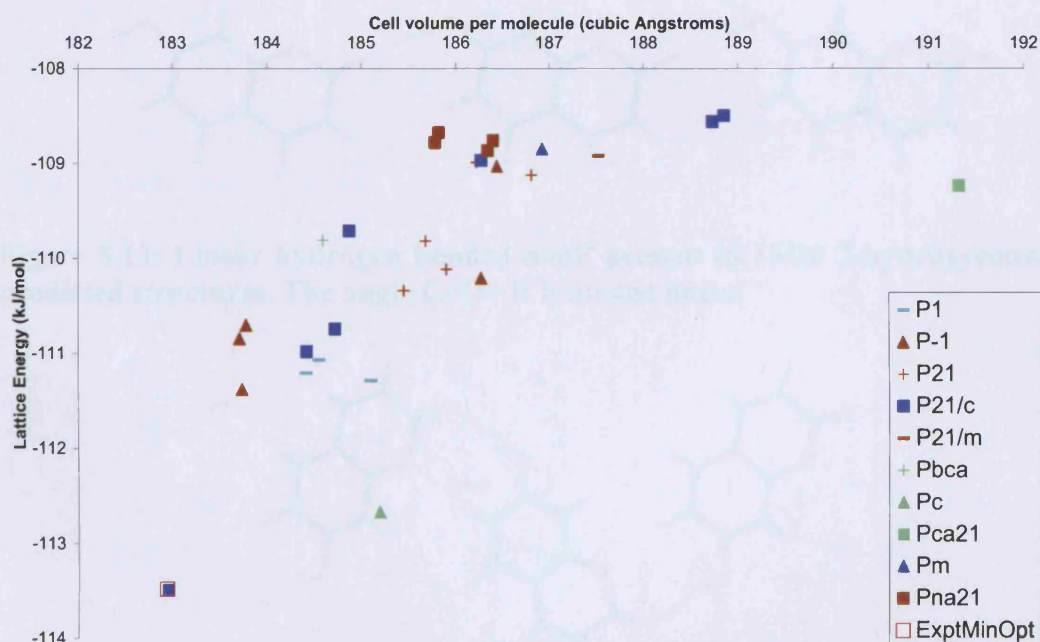
Figure 5.12: Scatter plot of all CSP generated structures for 7-hydroxycoumarin within 5 kJ mol^{-1} of the global energy minimum

The global energy minimum structure from the CSP search was found to correspond exactly to the ExptMinOpt(MP2) minimisation of DETFOX, 0.8 kJ mol^{-1} lower in energy than the next structure and over 2 kJ mol^{-1} lower in energy than the third lowest energy structure. It was also the most dense structure generated.

All 29 structures showed the same O-H···O hydrogen bond as observed in the experimental structure (figure 5.10). The majority of structures (15) exhibited ribbons in which the hydrogen bond angle H···O=C was almost linear (figure 5.13). Almost all of these structures were in the low symmetry space groups $P1$, $P\bar{1}$ or $P2_1$. The higher symmetry space groups such as $P2_1/c$ and $Pna2_1$ almost always formed chains with either *anti* (7) or *syn* (7) configurations. The global energy minimum structure was the only *anti* structure in the lowest 4.7 kJ mol^{-1}. Figure 5.14 shows a flat *syn* chain that takes advantage of the possibility of forming a C-H···O interaction as well as the strong hydrogen bond. However, only two *syn* structures form this flat chain, with the other five exhibiting rippled *syn* chains that preclude formation of the C-H···O hydrogen bond. Five of the six lowest energy predicted structures have offset π-stacking though there is no correlation within the set of low energy predicted structures between chain configuration and the presence or absence of offset π-stacking.
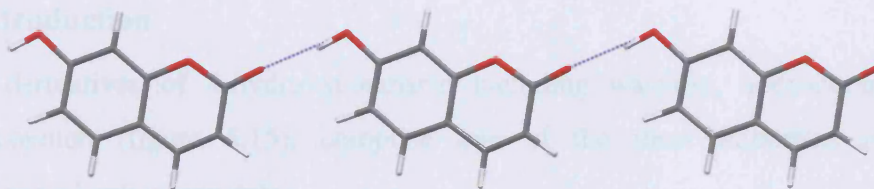
Figure 5.13: Linear hydrogen bonded motif present in 15/29 7-hydroxycoumarin predicted structures. The angle C=O···H is almost linear
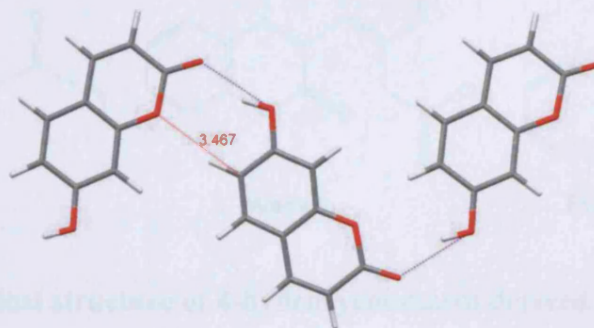
Figure 5.14: *syn* hydrogen bonded motif present in 7/29 7-hydroxycoumarin predicted structures. The C-H···O contact (3.467 Å) is shown once (red dotted line)

A total of 12 of the 29 low energy structures have molecules correctly orientated for potential photodimerisation, with nine producing the *syn* head-to-head product, two the *anti* head-to-tail product and one the *syn* head-to-tail product.

5.5.3 Discussion

The crystal structure prediction search on 7-hydroxycoumarin was successful with the known experimental structure found at the global energy minimum with highest density. Crystal structure prediction was carried out to assess whether 7-hydroxycoumarin was a molecule that could potentially exhibit polymorphism, and from the results of the search it can be concluded that the known form is probably the thermodynamically most stable structure, and that any other polymorphs that could be discovered would be metastable. Analysis of the predicted structures shows that the expected O-H···O chain hydrogen bonded motif is present in all structures, but with three different chain configurations possible, *syn/anti*/linear, with the linear disposition found twice as often as either *syn* or *anti*.

5.6 4-Hydroxycoumarin

5.6.1 Introduction

Several derivatives of 4-hydroxycoumarin including warfarin, acenocoumarol and phenprocoumon (figure 5.15), comprise one of the most important classes of pharmaceutical anti-coagulants.

Acenocoumarol   Warfarin   Phenprocoumon
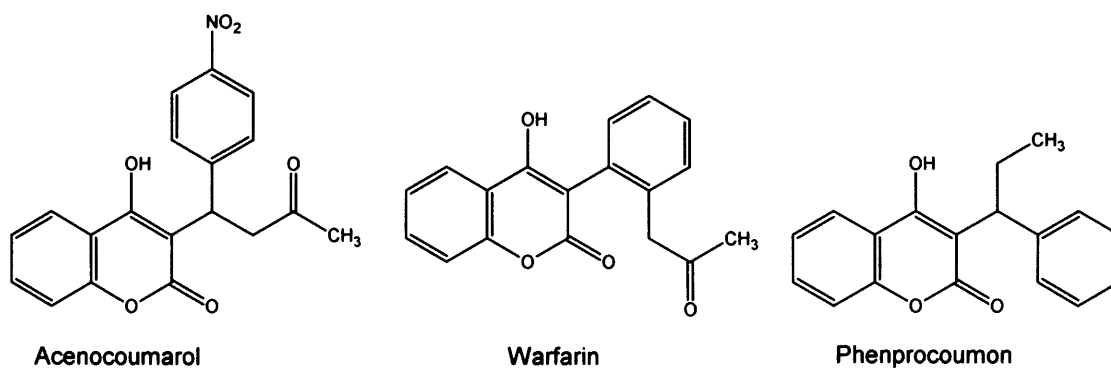
Figure 5.15: Chemical structure of 4-hydroxycoumarin derived anti-coagulants

The only known 4-hydroyxcoumarin crystal structure is a monohydrate that has been determined twice,[203;204] (CSD reference codes HOXCUM and HOXCUM01) but no

anhydrous crystal structure has been reported. The synthesis of 4-hydroxycoumarin was reported by Anschütz[214] in 1909 with a melting point of 204-206° C, while an alternative synthesis in 1943 by Stahmann *et al.*[215] reported a melting point of 214-216° C, perhaps indicating that these different syntheses yielded different crystal forms.

The monohydrate crystal structure has been reported to be photostable.[197] It crystallises in the orthorhombic space group $P2_12_12_1$, with a single 4-hydroxycoumarin molecule and a water molecule in the asymmetric unit. Water mediates all hydrogen bond contacts in the crystal structure, so there are no 4-hydroxycoumarin···4-hydroxycoumarin contacts (figure 5.16), and the hydrogen bonds produce a three dimensional motif. The 4-hydroxycoumarin hydroxyl group participates in one hydrogen bond and the carbonyl group acts as a hydrogen bond acceptor to two different water molecules.
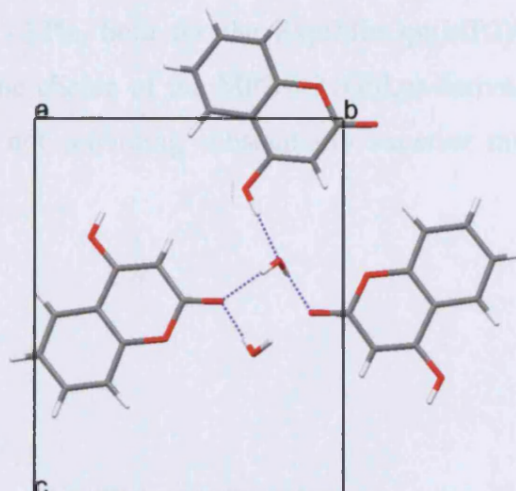
Figure 5.16: Hydrogen bonding present in 4-hydroxycoumarin monohydrate

Crystal structure prediction was undertaken alongside an experimental crystallisation screen to predict and discover anhydrous forms of 4-hydroxycoumarin. Four anhydrous forms were experimentally identified, with the crystal structures of two of them, forms 2 and 3, fully determined by SXRD. After determination of their crystal structures, both were energy minimised for comparison with the CSP results. Forms 1 and 4 were identified using XRPD, with form 4 successfully indexed.

5.6.2 Energy minimisation of forms 2 and 3

Forms 2 and 3 of 4-hydroxycoumarin were energy minimised with both their experimental molecular conformations and the MP2/6-31G(d,p) *ab initio* optimised molecular conformation (tables 5.6 and 5.7), and using the same potential as that used for the CSP search (FIT dispersion-repulsion potential plus MP2/6-31G(d,p) derived distributed multipoles). To enable the same comparison of electrostatic models as made in section 5.2, energy minimisations were also carried out with SCF/6-31G(d,p)-derived multipoles.

The energy minimisations of form 2 were not successful with either electrostatic model, with all minimisations having F-values greater than 75 and with errors greater than 5% in the *a* and *b* axes in all cases. In contrast the energy minimisations of form 3 were successful with both electrostatic models, with a largest F-value of 55 and largest cell parameter error of 3.32%, both for the ExptMinOpt(MP2) minimisation. These minimisations justified the choice of the MP2/6-31G(d,p)-derived multipoles with the SCF-derived multipoles not providing substantially superior minimisations in either case.

4-hydroxycoumarin Form 2	Experimental	MP2 ExptMinExpt	% error	MP2 ExptMinOpt	% error	SCF ExptMinExpt	% error	SCF ExptMinOpt	% error
a (Å)	9.355	8.795	-5.98	8.754	-6.43	8.915	-4.70	8.872	-5.16
b (Å)	10.975	11.808	7.69	11.979	9.14	11.599	5.68	11.764	7.18
c (Å)	14.817	14.908	0.62	15.007	1.28	14.868	0.35	14.985	1.14
β (°)	105.69	106.06	0.35	106.28	0.56	106.09	0.38	106.32	0.60
Volume (Å³)	1464.55	1487.91	1.59	1510.45	3.13	1477.29	0.87	1500.97	2.49
Density (g cm⁻³)	1.471	1.448	-1.57	1.426	-3.04	1.458	-0.86	1.435	-2.43
Final Energy (kJ mol⁻¹)		-112.5		-106.85		-127.08		-120.17	
F		131.2		178.2		76.7		112.64	

Table 5.6: Energy minimisations of the crystal structure of 4-hydroxycoumarin form 2

4-hydroxycoumarin Form 3	Experimental	MP2 ExptMinExpt	% error	MP2 ExptMinOpt	% error	SCF ExptMinExpt	% error	SCF ExptMinOpt	% error
a (Å)	21.201	21.687	-2.29	21.904	3.32	21.518	1.49	21.726	2.48
b (Å)	3.785	3.812	0.722	3.812	0.73	3.872	2.31	3.876	2.41
c (Å)	20.045	20.548	2.51	20.654	3.04	20.298	1.26	20.396	1.75
β (°)	115.68	117.73	1.77	117.71	1.76	117.68	1.73	117.68	1.73
Volume (Å³)	1449.64	1503.74	3.73	1526.93	5.33	1497.76	3.32	1521.06	4.93
Density (g cm⁻³)	1.486	1.432	-3.60	1.41	-5.06	1.438	-3.21	1.416	-4.70
Final Energy (kJ mol⁻¹)		-110.24		-105.98		-121.94		-117.51	
F		37.56		54.80		32.95		46.22	

Table 5.7: Energy minimisations of the crystal structure of 4-hydroxycoumarin form 3

5.6.3 Crystal structure prediction

The MP2/6-31G(d,p) *ab initio* optimised molecular conformation was the search input for the MOLPAK search which used 47 packing types covering 22 space groups. Distributed multipoles derived from the MP2/6-31G(d,p) charge density were used along with the FIT dispersion-repulsion potential. The search generated 147 structures within 10 kJ mol^{-1} of the global energy minimum and 50 structures within 5 kJ mol^{-1} (figure 5.17). All structures within 10 kJ mol^{-1} were examined after the search for their true space group using the ADDSYM algorithm contained in PLATON,[139] and all structures within 5 kJ mol^{-1} were analysed for their hydrogen bonding motifs. Predicted structures were not found that corresponded to the energy minimised structures of either form 2 or form 3 because in both structures the asymmetric unit contained two independent molecules. The energy minimised experimental structures of forms 2 and 3 were respectively 6.6 and 7.5 kJ mol^{-1} higher in energy than the global minimum.
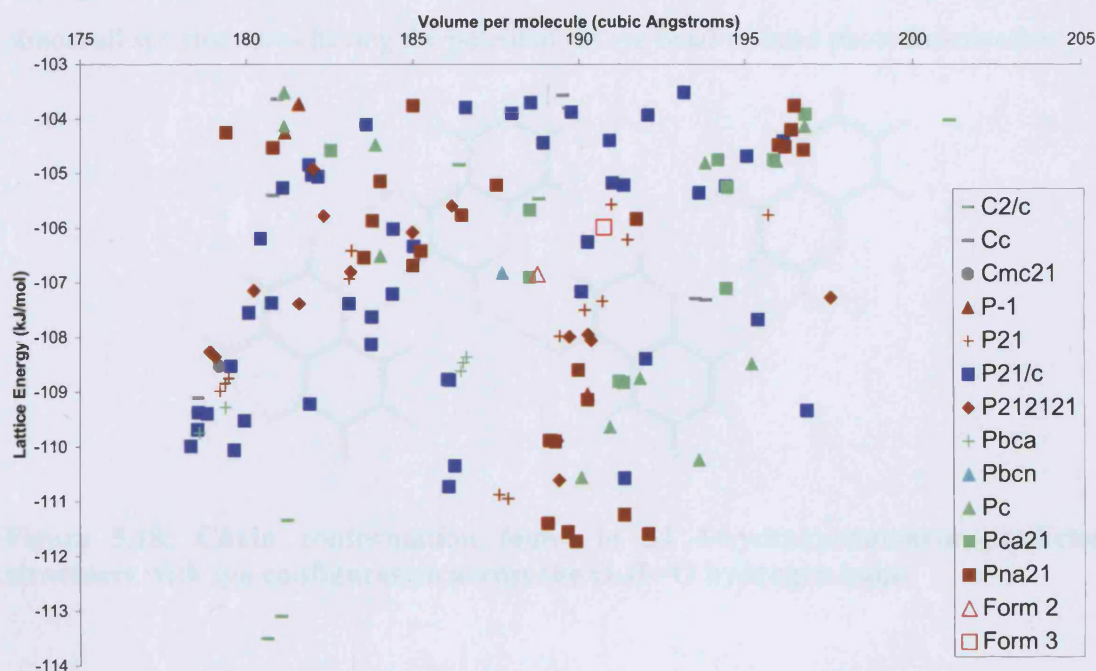
Figure 5.17: Scatter plot of all CSP generated structures for 4-hydroxycoumarin within 10 kJ mol^{-1} of the global energy minimum

In all predicted structures the hydrogen bond O4-H2$\cdots$O2 was present giving a chain motif, with $C_1^1(6)$ graph set. The two lowest energy structures both exhibited the space group *C2/c* and the crystal packing in both was found to be almost identical. The third

lowest energy structure was 1.8 kJ mol^{-1} higher in energy than the global minimum structure, with a substantially different chain conformation.

34 predicted structures had the molecules in a *syn* configuration across the hydrogen bond (figure 5.18), and none of these structures contained flat chains which would utilise the C3-H1···O1 and C8-H6···O2 weak hydrogen bonds. In 17 of these *syn* structures the chains had a rippled profile and the other 17 contained V-shaped chains (figure 5.19). In all structures containing V-shaped chains and all but one of the structures containing rippled chains, the chains stacked directly, with adjacent stacked chains have an offset π-π stacked geometry rather than having their aromatic rings in direct face-to-face contact. The almost exclusive observation of offset π-stacking in the *syn* structures follows the Hunter-Sanders rules[213] for π-π interactions that state that the face-to-face π-stacked geometry is repulsive whereas the offset π-stacked geometry has attractive π-σ interactions between the aromatic ring in one molecule and the aromatic hydrogen atoms of the next molecule.[216] This offset molecular stacking also lead to almost all *syn* structures having the potential for *syn* head-to-head photodimerisation.
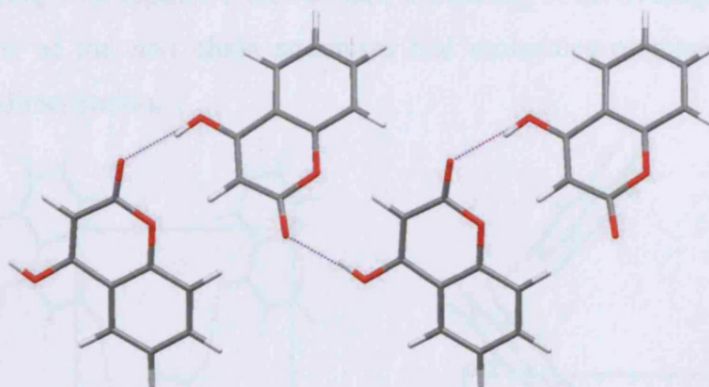
Figure 5.18: Chain conformation found in 34 4-hydroxycoumarin predicted structures with *syn* configuration across the O-H···O hydrogen bond
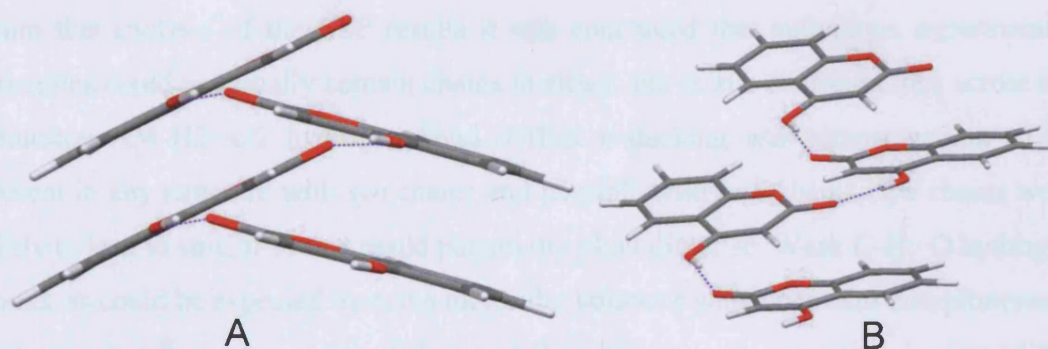
Figure 5.19: 4-Hydroxycoumarin *syn* **chain profiles. A: V-shaped profile; B: Rippled chain profile. Both views almost parallel to chain axes**

In 15 of the 50 lowest energy structures, including the two lowest energy predicted structures, the molecules are orientated *anti* across the hydrogen bond. In the two lowest energy structures the chain was in a stepped conformation (figure 5.20). In only the three lowest energy *anti* chain structures and one other structure were the molecules packed with the attractive offset π-stacking. The rest of the *anti* chain structures had molecules packing with repulsive face-to-face π-stacking or no overlap of the aromatic rings. Only two of the *anti* chain structures had molecules orientated correctly for potential photodimerisation.
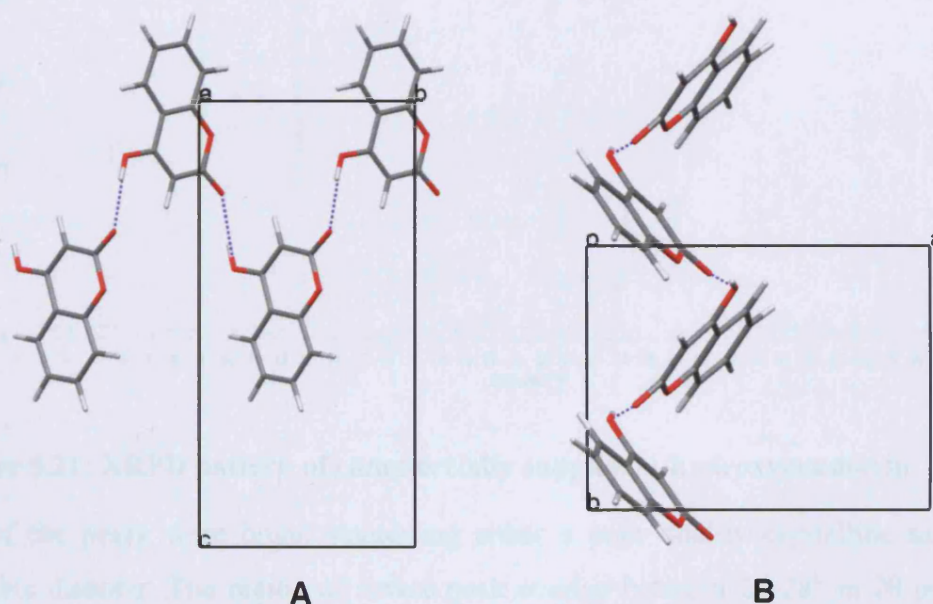
Figure 5.20: Chain conformation found in the two lowest 4-hydroxycoumarin predicted structures. A: the molecules have an *anti* **configuration across the O-H···O hydrogen bond; B: the chain has a stepped profile**

From this analysis of the CSP results it was concluded that anhydrous experimental structures could potentially contain chains in either *anti* or *syn* configurations across the ubiquitous O4-H2···O2 hydrogen bond. Offset π-stacking was almost certain to be present in any structure with *syn* chains and possible with *anti* chains. *Syn* chains were likely to lead to structures that could potentially photodimerise. Weak C-H···O hydrogen bonds, as could be expected from the molecular structure which contains complimentary hydrogen bonding edges, were not expected as they were not present in any of the predicted structures.

5.6.4 Experimental crystallisation screen

4-Hydroxycoumarin was purchased from Sigma-Aldrich (Poole, 98% purity). The material was supplied as an anhydrous microcrystalline powder and its XRPD pattern was recorded. As no anhydrous crystal structure had previously been reported for 4-hydroxycoumarin this crystal form was designated form 1 (figure 5.21).
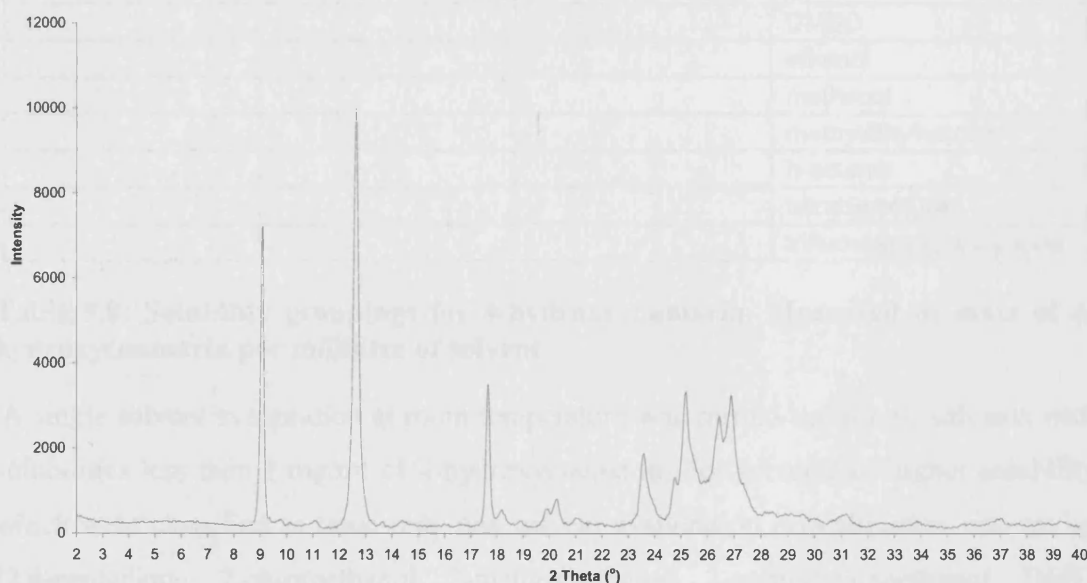
Figure 5.21: XRPD pattern of commercially supplied 4-hydroxycoumarin

All of the peaks were broad suggesting either a poor quality crystalline sample or possible disorder. The region of severe peak overlap between 24-28° in 2θ precluded indexing the structure.

The solubility of 4-hydroxycoumarin in 46 solvents at room temperature was crudely measured by adding measured quantities of 4-hydroxycoumarin to a specified volume of each solvent until the solid was in excess. The solvents were then grouped into four categories (table 5.8).

Insoluble	< 1 mg/ml	1-10 mg/ml	>10 mg/ml
1-octene	1,2-dichloroethane	1,4-dioxane	1-butanol
butylvinylether	1-bromobutane	acetonitrile	1-propanol
cyclohexane	benzonitrile	diethylether	2,2,2-trifluoroethanol
n-hexane	bromoethane	ethylacetate	2,4-pentadione
n-octane	chloroform	methylbenzoate	2-butanol
tetrachloroethylene	di-n-butylether	nitromethane	2-butoxyethanol
	dibromomethane	t-butylmethylether	2-chloroethanol
	dichloromethane		2-propanol
	isoprene		2-methoxyethanol
	isopropylether		2-methylaminoethanol
	o-xylene		acetic acid
	toluene		acetone
	water		DMF
			DMSO
			ethanol
			methanol
			methylethylketone
			n-octanol
			tetrahydrofuran
			trifluoroacetic anhydride

Table 5.8: Solubility groupings for 4-hydroxycoumarin. Measured in mass of 4-hydroxycoumarin per millilitre of solvent

A single solvent evaporation at room temperature was carried out for all solvents with solubilities less than 1 mg/ml of 4-hydroxycoumarin. For solvents of higher solubility which were classified as toxic only one solvent evaporation crystallisation was set up (2,4-pentadione, 2-chloroethanol, 2-methoxyethanol, 2-methylaminoethanol, DMF, DMSO, n-octanol, trifluoroacetic anhydride). For each of the remaining 21 solvents 5 crystallisations were set up (table 5.9): solvent evaporation of a 100% saturated solution at room temperature; solvent evaporation of a 50% saturated solution at room temperature; solvent evaporation of a 100% saturated solution at 5° C in a domestic refrigerator; vapour diffusions with both chloroform and toluene anti-solvents.

Solvent	EV 100 (RT)	EV 50 (RT)	EV 100 (5° C)	VD CHCl₃	VD TOL
1,4-Dioxane	x	x	x	x	x
1-Butanol	x	x	x	x	x
1-Propanol	x	x	x	x	x
2,2,2-Trifluoroethanol	x	x	x	x	x
2-Butanol	x	x	x	x	x
2-Butoxyethanol	x	x	x	x	x
2-Propanol	x	x	x	x	x
Acetic Acid	x	x	x	x	x
Acetone	x	x	x	x	x
Acetonitrile	x	x	x	x	x
Butylvinylether	x	x	x	x	x
Diethylether	x	x	x	x	x
Ethanol	x	x	x	x	x
Ethylacetate	x	x	x	x	x
Isopropylether	x	x	x	x	x
Methanol	x	x	x	x	x
Methylbenzoate	x	x	x	x	x
Methylethylketone	x	x	x	x	x
Nitromethane	x	x	x	x	x
t-Butylmethylether	x	x	x	x	x
Tetrahydrofuran	x	x	x	x	x
1,2-Dichloroethane	x				
1-Bromobutane	x				
2,4-pentadione	x				
2-Chloroethanol	x				
2-Methoxyethanol	x				
2-Methylaminoethanol	x				
Benzonitrile	x				
Bromoethane	x				
Chloroform	x				
Dibromomethane	x				
Dichloromethane	x				
Di-n-butylether	x				
DMF	x				
DMSO	x				
Isoprene	x				
n-Octanol	x				
o-Xylene	x				
Tetrachloroethylene	x				
Toluene	x				
Trifluoroacetic anhydride	x				
Water	x				

Table 5.9: Summary of all crystallisations of 4-hydroxycoumarin. x denotes that this crystallisation was set up; EV = evaporation of 100%, 50% saturated solutions; VD = vapour diffusion with chloroform or toluene

186

In 44 crystallisations the crystal morphology was evident, with fine needles observed 40 times which were usually too small for single crystal X-ray diffraction. In some of these crystallisations the needles formed ferns on the base and walls of the crystallising vessel. In the crystallisations in which a clear crystal morphology was not observed the product was a white microcrystalline precipitate.

Only six samples yielded crystals suitable for checking by SXRD, and five further samples were chosen for XRPD based on the quantity of precipitate and the crystal morphology.

Block crystals and a white precipitate were grown from dibromomethane and the blocks were found to be a new polymorph (form 2). This polymorph was also crystallised from toluene. The crystal structure (form 3) found in the common needle morphology was characterised by SXRD from crystals grown from a sub-saturated solution of 4-hydroxycoumarin in tert-butylmethylether. XRPD on samples with fern morphology found these to be form 3. A solvate was isolated by solvent evaporation from trifluoroacetic anhydride, identified by XRPD, and the desolvation of this sample led to the observation of the new unsolvated form 4, which was indexed from its XRPD pattern. 4-hydroxycoumarin was not particularly soluble in water at room temperature but the known monohydrate was grown from the hygroscopic solvents nitromethane and acetonitrile, and grew as long flat needles.

A crystallisation database, summarising all crystallisations and results, is provided on the supporting information CD (in Chapter_5_Coumarins) and a summary spreadsheet detailing crystallisation conditions for all of the forms of 4-hydroxycoumarin is also given (Crystallisation_Summary.xls).

5.6.5 4-Hydroxycoumarin form 2

Slow crystallisation from dibromomethane gave two products, a white microcrystalline material and several block crystals (figure 5.22). The two were manually separated and XRPD was used to determine that the white material was form 1, the same form as the commercially supplied material. The block crystals were of sufficient quality for single crystal X-ray diffraction and a full data set was collected and the structure determined. The simulated powder pattern of this new polymorph, form 2, was different to that of form 1. Repeating this crystallisation led to the observation that faster solvent

evaporation of dibromomethane gave no block crystals, and it was confirmed by XRPD that form 2 was completely absent.
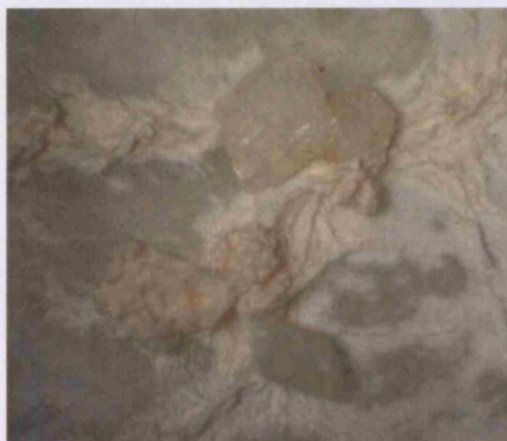
Figure 5.22: 4-Hydroxycoumarin grown from dibromomethane. Blocks are form 2 while the white microcrystalline material is form 1

4-Hydroxycoumarin form 2 crystallises in the space group $P2_1/c$ with two molecules in the asymmetric unit (figure 5.23, table 5.10).
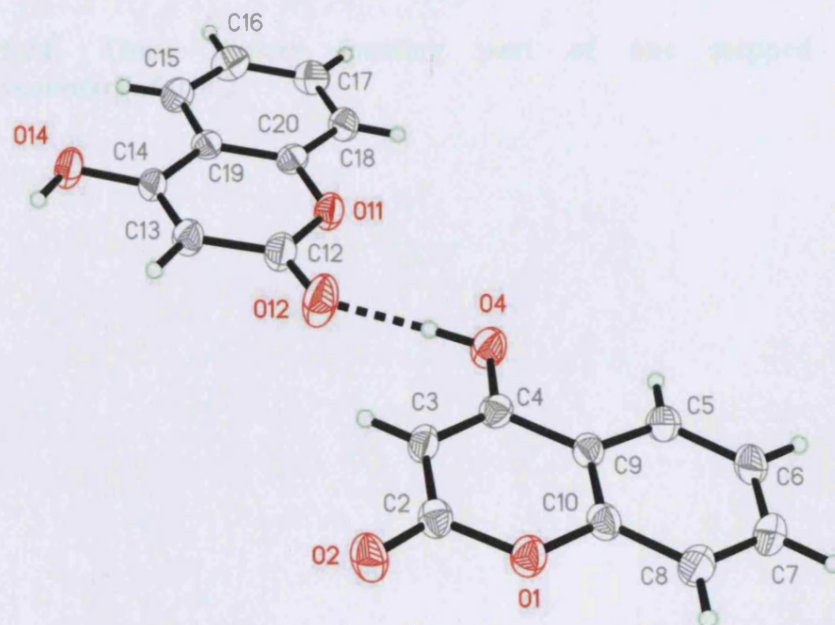
Figure 5.23: Asymmetric unit of 4-hydroxycoumarin form 2

In this crystal structure the molecules are not orientated correctly for photodimerisation and the aromatic rings are not stacked in the crystal structure. The hydrogen bond O4-H2···O12 is present with the two independent molecules forming an almost flat

dimer in an *anti* configuration across the hydrogen bond. These dimers then comprise the basic unit that form a stepped chain (figure 5.24) and adjacent chains intercalate (figure 5.25). The lowest energy predicted structure is made up of stepped chains that intercalate in a similar manner, but each unit in the stepped chain is a single molecule.
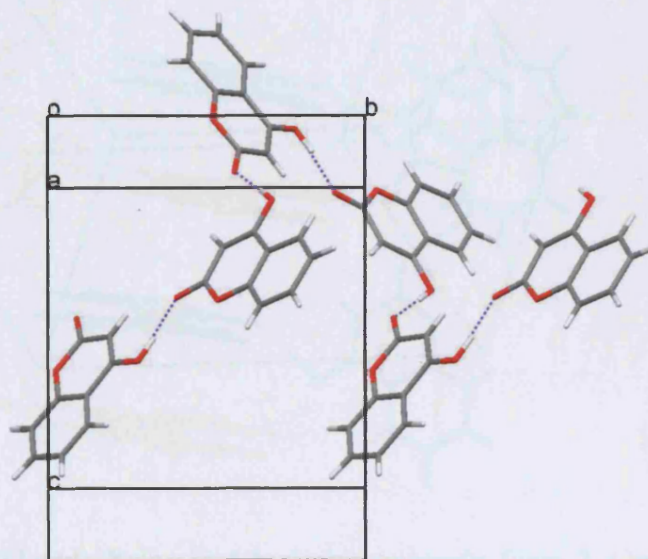
Figure 5.24: Three dimers forming part of one stepped chain in 4-hydroxycoumarin form 2
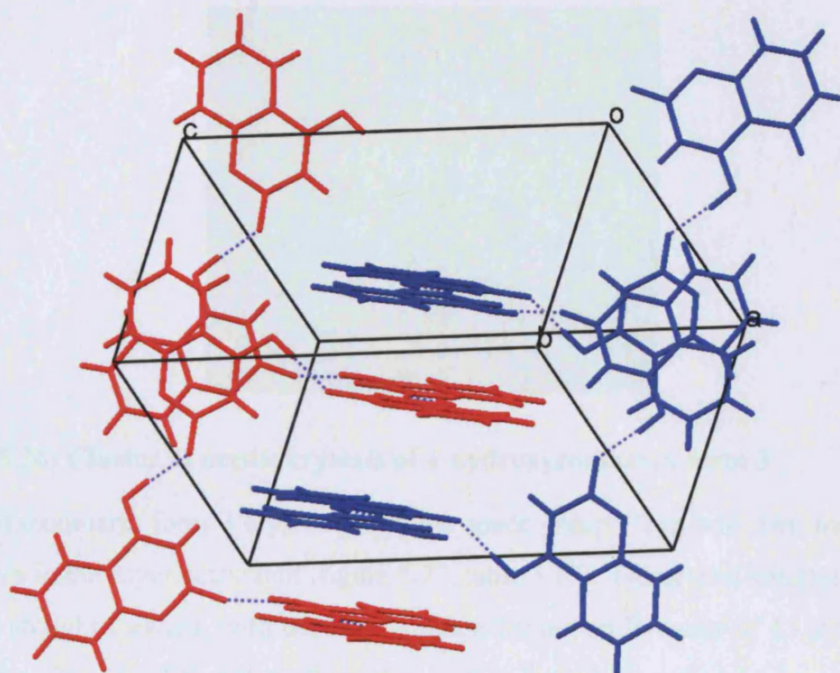
Figure 5.25: Two intercalated chains in 4-hydroxycoumarin form 2. One chain is coloured red and the other is coloured blue

5.6.6 4-Hydroxycoumarin form 3

Large crystals of the common needle morphology were grown from a sub-saturated solution of 4-hydroxycoumarin in tert-butylmethylether (figure 5.26). The crystals were found to be non-merohedrally twinned, and a SXRD data set was collected on a good quality twinned crystal.

The twin resolution program GEMINI[217] was used to index the two separate parts of the twin. The data set collected was integrated using the orientation matrix of both components separately and the TWINHKL component of GEMINI[217] was used to write two data files: one containing only the non-overlapping data for the major component and a second containing all data. This second data file was an hkl file in which each reflection was assigned as belonging to component 1 or component 2 or having contributions from both. The structure was solved on just the major component and subsequently the data file containing data from both components was used to refine against. The twin components were related by the twin law (-1 0 0, 0 -1 0, 0 -0.92 1) and were in the ratio 83:27.

Figure 5.26: Cluster of needle crystals of 4-hydroxycoumarin form 3

4-hydroxycoumarin form 3 crystallises in the space group $P2_1/c$ with two independent molecules in the asymmetric unit (figure 5.27, table 5.10). The crystal structure solution was not straight forward, with the final solution having an R-factor of 13.6%. Each of the atoms C6, C7, C8 and C10 in the first independent molecule have ratios of maximum to minimum atomic displacement parameters greater than four – such strong anisotropy usually indicates that there is disorder present in the crystal structure. Attempts to model disorder did not however lead to a superior structure solution and consequently the best ordered solution is presented here.
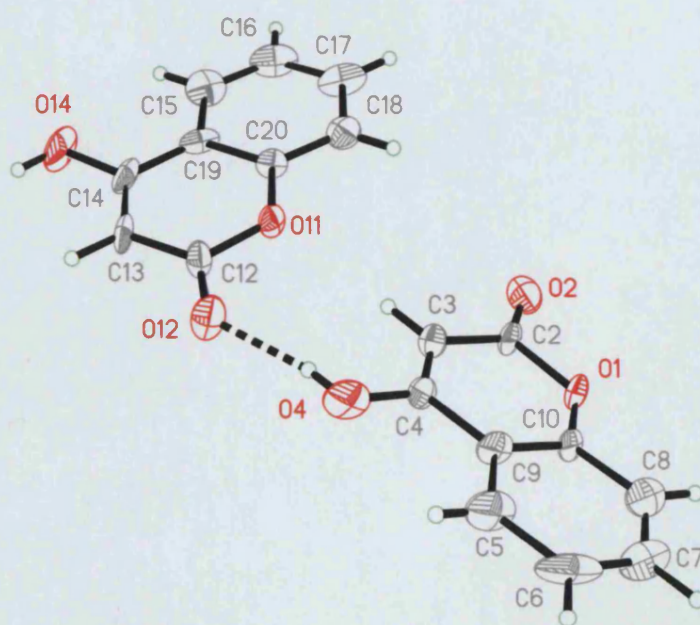
Figure 5.27: Asymmetric unit of 4-hydroxycoumarin form 3

In contrast to form 2, the independent molecules in this polymorph are *syn* orientated across the O4-H2···O12 hydrogen bond, forming a rippled chain. The chains stack directly in the crystal structure parallel to the *ac* plane, with the distance between the chains the *b* axial length, 3.785(2) Å (figure 5.28). The aromatic rings in the 4-hydroxycoumarin molecules have the attractive offset π-stacking as observed in almost all of the *syn*-chain based predicted structures and the molecular orientation, along with the short stacking separation, leads to this structure being potentially photoreactive.
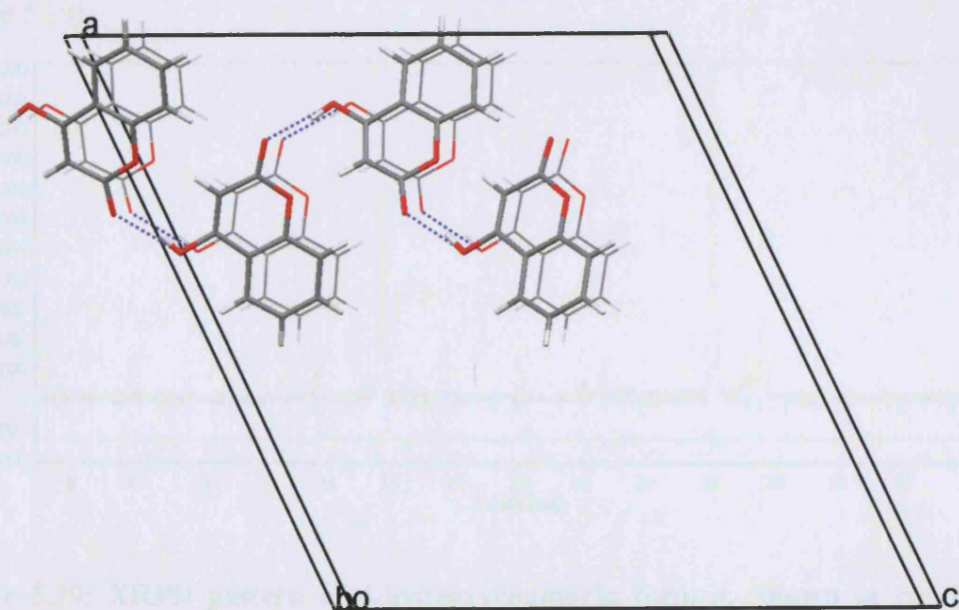
Figure 5.28: Chain stacking in 4-hydroxycoumarin form 3. The lower chain is shown in wireframe and the separation between the chains is 3.785(2) Å

5.6.7 4-Hydroxycoumarin trifluoroacetic anhydride solvate and form 4

Solvent evaporation from a saturated solution of 4-hydroxycoumarin in trifluoroacetic anhydride produced a microcrystalline material that was identified as a solvate by thermogravimetric analysis, with a 42% weight loss recorded at an onset temperature of 77° C. This compares with the calculated 56% weight loss that a 1:1 4-hydroxycoumarin : trifluoroacetic anhydride solvate would undergo upon desolvation. The solvate proved very unstable upon removal from the mother liquor and appreciable desolvation could be expected to have occurred prior to the weight loss measurement. XRPD of the solvate produced a novel diffraction pattern, but it was likely that there

was a mix of phases present, including both the solvate and one or more desolvated phases, and indexing was not possible. Upon rapid desolvation by mild heating an XRPD measurement run immediately showed that the initial desolvation product was predominantly amorphous. Within 12 hours the amorphous phase crystallised, usually to a mix of two phases: form 1 and a new form 4. In one particular slow desolvation experiment form 4 was successfully isolated and an accurate XRPD pattern measured between 2-40° in 2θ, with a 0.05° step size and 80 seconds of exposure at each step. This was analysed by Martin Vickers (University College London) using CrysFire[218] to index the data, Checkcell to assign a unit cell and Rietica[219] for Le Bail refinement (figure 5.29).
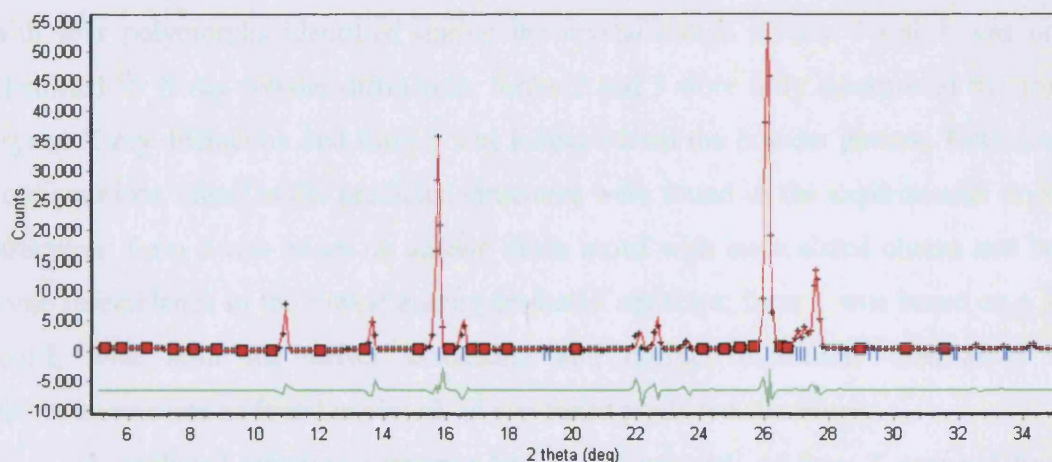
Figure 5.29: XRPD pattern of 4-hydroxycoumarin form 4. Shown in red is the recorded powder pattern; blue tick marks indicate the position of the peaks according to the determined unit cell; green line shows the agreement between the two

The unit cell was determined to belong to the orthorhombic space group *Fdd2*, and from the volume of the unit cell and multiplicity of the space group, this requires the asymmetric unit to contain one molecule (table 5.10). Proton NMR of the desolvation product of the trifluoroacetic anhydride solvate was unchanged from that of the commercially supplied material, discounting the possibility that this crystal structure is a reaction product or that solvent was still present.

5.6.8 Discussion

Crystal structure prediction carried out on 4-hydroxycoumarin gave a large number of low energy structures all based on the only strong hydrogen bond possible, O4-H2···O2. Two chain configurations were observed in the low energy predicted structures: 15 structures including the two lowest energy structures contained the *anti* chain configuration and 34 had a *syn* chain motif. Offset π-stacking was present in the majority of the *syn* structures, and the lowest energy *anti* structures. Within the 50 structures examined there was no preference among the lowest energy structures for either the *anti* or *syn* chain configuration.

The experimental search for anhydrous forms of 4-hydroxycoumarin was a success, with four polymorphs identified during the crystallisation screen. Form 1 was only identified by X-ray powder diffraction, forms 2 and 3 were fully determined by single crystal X-ray diffraction and form 4 was indexed from the powder pattern. Both chain configurations found in the predicted structures were found in the experimental crystal structures: form 2 was based on an *anti* chain motif with intercalated chains and bore some resemblance to the lowest energy predicted structure; form 3 was based on a *syn* motif, with both the offset π-stacking and correct molecular orientation for photodimerisation as found in almost all *syn*-based predicted structures.

A predicted structure corresponding to the unit cell of form 4 was not found because structures were not generated by MOLPAK in the space group *Fdd2*. Forms 2 and 3 were not predicted because they both contain two molecules in the asymmetric unit but the structures were energy minimised for comparison to the predicted structures. The global energy minimum structure was found to be over 6.6 kJ mol^{-1} lower in energy than forms 2 and 3, superficially suggesting that thermodynamically more stable polymorphs could be found. However the errors in the reproduction of forms 2 and 3 give little confidence in the energies of these structures relative to the predicted structures, and therefore no comment can be made on the plausibility of finding a thermodynamically more stable form.

The melting points of forms 1 and 3 were recorded and both were found to melt between 217-218° C, not providing any information on their relative stability. The production of form 2 concomitantly only under slow thermodynamically controlled crystallisation conditions, and its absence under faster kinetic crystallisation conditions,

indicates that it is more stable than form 1. The formation of form 4 was usually concomitant with form 1 and the phase change of the exclusively form 4 sample to a mix of forms 1 and 4 upon standing suggests that form 1 is more stable than form 4.

5.7 Conclusion

The crystal structure prediction of both coumarin and 7-hydroxycoumarin found the known structure in both searches as the global energy minimum structure suggesting that the known crystal structure of both molecules is the thermodynamically most stable form. These studies exemplify one of the key envisaged utilities of crystal structure prediction – namely the confirmation of a known polymorph as the thermodynamically most stable form. This information would be of considerable use during the physical characterisation stage of pharmaceutical development: further experimental investigation to confirm that the known crystal structure is the thermodynamically most stable form could be curtailed, and possibly refocused toward generating metastable forms.

The known structure of 6-methoxycoumarin was found during crystal structure prediction ranked 22nd. All of the predicted structures in this search exhibited the same fundamental triplet of C-H⋯O hydrogen bonds and differed only by the interactions between ribbons propagated by these triplet interactions. No predicted structure was found that corresponded to a second unit cell reported for 6-methoxycoumarin. From the range of cell volumes of the predicted structures is concluded that this unit cell must correspond to that of a solvate. A limited crystallisation screen did not produce a crystal structure with this unit cell.

The crystal structures of two anhydrous forms of 4-hydroxycoumarin were fully determined during the experimental investigation and two further polymorphs were identified by X-ray powder diffraction. The predicted structures of 4-hydroxycoumarin were found to exhibit chains in which the molecules were either *syn* or *anti* orientated across the hydrogen bond. In the *syn* chains offset π-stacking was observed which led to classification of almost all of these structures as potentially photoreactive. The *anti* chain based structures were found to be less likely to have offset π-stacking and consequently less likely to have the molecules orientated correctly for photodimerisation. While neither of the fully characterised crystal structures was

predicted in the search because of the limitations of the search method, they did confirm that the features identified in the predicted structures were realistic: form 2 was found to be based on an *anti*-chains and indeed showed neither offset π-stacking or molecular orientation correct for photodimerisation; form 3 contained *syn* configured chains and confirmed the prediction results by having both offset π-stacking and being potentially photoreactive.

The crystal structure prediction for all four molecules investigated in this chapter resulted in some hypothetical structures that could potentially photodimerise and some that could not. This observation was borne out by the experimental discovery of two 4-hydroxycoumarin polymorphs, one of which fell into each of these categories. This highlights the role crystal structure prediction could play, in combination with experiment, in identifying candidates for photodimerisation reactions and the potential for polymorphism to be exploited as a facet of crystal engineering towards the ultimate goal of solid state 'diffusionless' reactions. MacGillivray states that "Approaches to align olefins in molecular solids for a [2+2] photoreaction may be classified into two general categories: methods that employ (1) intramolecular substitution, or (2) auxiliary components"[220] – the work presented here suggests that the exploitation of polymorphism may provide a third pathway to photoreactive crystal structures.

5.8 Further work

The experimental study on 4-hydroxycoumarin remains incomplete: further investigation of the transiently stable form 4 could yield a structure determined by X-ray powder diffraction, either using the method of production detailed here, desolvation of the trifluoroacetic anhydride solvate, or potentially by an alternative method. Upon production of a pure sample of form 4, kinetic stabilisation by cooling to low temperature would be required to prevent the phase change to a mixed phase with form 1. Further analysis to determine the relative thermodynamic stabilities of the four polymorphs and their inter-relationships would be beneficial to understanding the crystallisation behaviour of 4-hydroxycoumarin.

Investigations into the photoreactivity of the polymorphs of 4-hydroxycoumarin could confirm the hypothesis that form 3 would photodimerise while form 2 would not. This, along with investigations of the DAXBIN01 form of 6-methoxycoumarin to

confirm the assertion that it is photostable, and of the DETFOX 7-hydroxycoumarin form to determine its photoreactivity, would add to our understanding of the crystal packing required for coumarins to photodimerise and how we can use crystal structure prediction to assess the likelihood that a coumarin will pack to give a photodimerising crystal structure.

Crystal Data

	4-hydroxycoumarin form 2	4-hydroxycoumarin form 3	4-hydroxycoumarin form 4*
Compound name			
Empirical formula	$C_9H_6O_3$	$C_9H_6O_3$	$C_9H_6O_3$
Formula weight	162.1	162.1	162.1
Crystal system, space group	Monoclinic, $P2_1/c$	Monoclinic, $P2_1/c$	Orthorhombic, $Fdd2$
a (Å)	9.3547(17)	21.20(1)	11.354(2)
b (Å)	10.975(2)	3.785(2)	32.12(4)
c (Å)	14.817(3)	20.05(1)	8.088(5)
α (°)	90	90	90
β (°)	105.688(3)	115.56(1)	90
γ (°)	90	90	90
V (Å^3)	1464.6(5)	1451.5(1.4)	2950(4)
Z', Z	2, 8	2, 8	1, 16
D(calc (g cm^{-3}))	1.471	1.484	1.460(2)
Data Collection			
Crystal size (mm)	0.45 x 0.27 x 0.13	0.57 x 0.12 x 0.10	0.3mm recessed flat plate
Temperature (K)	150(2)	150(2)	298(2)
hkl range (h, k, l)	-12→12, -14→14, -19→19	-28→23, -5→5, -26→26	-
Reflections measured, R_{int}	12284, 0.0404	12339, -	-
Independent reflections	3480	7938	-
Reflections I>2σ(I)	2699	4391	-
Refinement			
Parameters refined	265	225	-
R(F) (I>2σ(I))	0.050	0.136	-
wR(F^2) (all reflections)	0.137	0.353	3.61 (R_{wp})
Residual electron density (min, max (e Å^{-3}))	0.39, -0.19	0.69, -0.64	-

Table 5.10: Crystal structure summary for 4-hydroxycoumarin structures included in this chapter. * indexed from XRPD data, Cu Kα radiation, $\lambda = 1.54056$ Å

Chapter 6 – 3-Azabicyclo[3.3.1]nonane-2,4-dione

6.1 Introduction

3-Azabicyclo[3.3.1]nonane-2,4-dione (BQT, figure 6.1) is a small, rigid organic molecule that was one of the three test molecules used for the second CCDC international blind test of computational crystal structure prediction methods (CSP2001).[122]
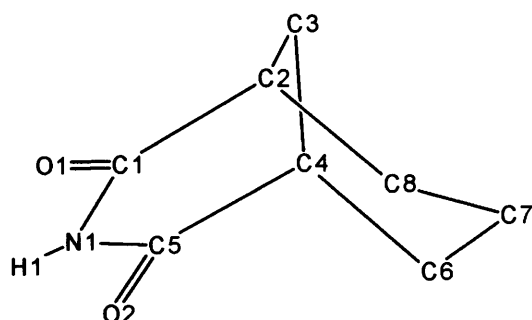
Figure 6.1: 3-Azabicyclo[3.3.1]nonane-2,4-dione (BQT) with numbering scheme

Computational crystal structure prediction progressed rapidly from its inception in the early 1960s[117] and has reached a stage where comparative testing of the different techniques in use today could give insights into the weaknesses of different methodologies. Three international blind tests of crystal structure prediction have been organised by the CCDC in 1999 (CSP1999),[121] 2001[122] and 2004 (CSP2004).[123] Each blind test comprised three test molecules of increasing complexity, with only the molecular diagrams released to the participating groups. Each group was limited to proposing three hypothetical crystal structures for each test molecule and the choice of which structures to submit from the total set of predicted structures was usually based on lattice energy, sometimes aided by crystallographic experience. One academic group also made predictions using a method exploiting the supramolecular synthons present in the crystal structures of structurally similar molecules,[221] which has the added advantage that the frequency of supramolecular synthons in the CSD[222] is influenced by kinetic factors as well as thermodynamics. Upon submission of all predicted hypothetical structures, comparisons were made with the experimental crystal structures, which until this point were held in secret. A predicted structure that corresponded closely to the

experimental structure within a tight tolerance limit based on nearest neighbour distances and orientation, was viewed as a successful prediction.[124]

6.2 2001 blind test results

BQT was the simplest molecule used for CSP2001, by virtue of being rigid, containing fewer than 25 atoms and comprising only C H N and O atoms. Participants were informed that the experimental crystal structure contained only one molecule in the asymmetric unit and that the space group was one of the ten most common space groups for organic molecules found in the CSD.[223] In spite of the relative simplicity of this molecule and its conventional crystallographic properties only 2 submissions from the total of 45 from the 15 participants correctly predicted the experimental structure (form 1[224]) within the allowed tolerance, though many participants subsequently found it within their expanded list of hypothetical structures. The two successful predictions were ranked second and third by their submitting participants. Analysis of the 45 submitted structures showed that there were two common recurring hydrogen bond motifs. 34 structures exhibited a hydrogen bonded dimer motif (figure 6.2A) and 8 exhibited hydrogen bonded chain motifs (figure 6.2B). The dimer-based structures were produced by a range of CSP methods using a variety of intermolecular potentials for energy minimisation of the crystal structures, leading to the conclusion that dimer-based structures are energetically competitive with the chain-based structures.
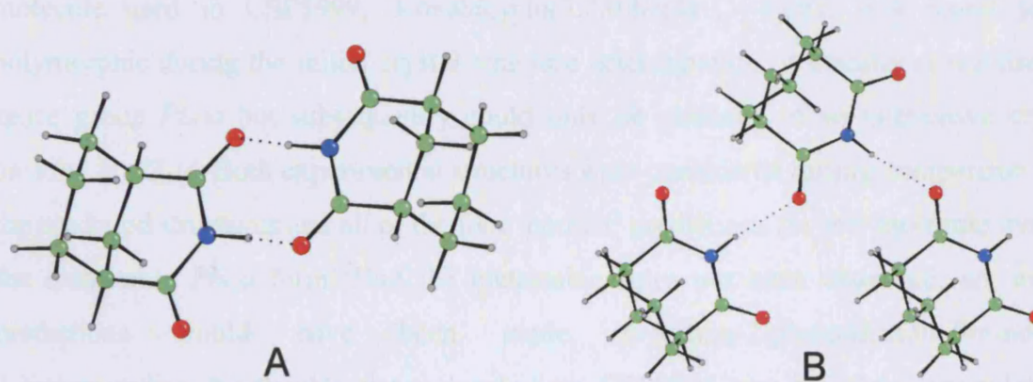
Figure 6.2: Common hydrogen bonded motifs found in BQT predicted structures. A: Dimer; B: Chain

Of the eight structures that exhibit a chain motif, only in the two successful predictions was the chain propagated by a glide plane; in the other six the chain was propagated by a 2_1 screw axis giving a different shape to the chain (figure 6.3).
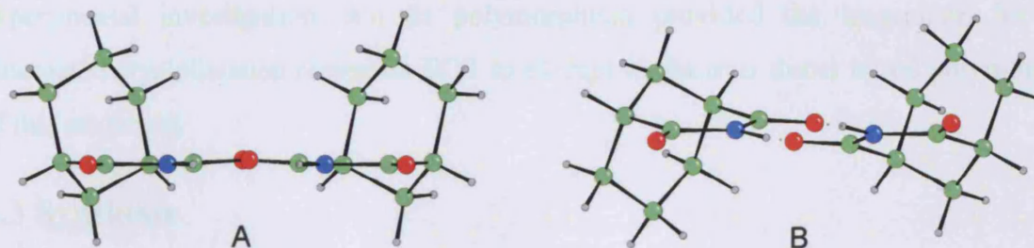
Figure 6.3: Hydrogen bonded chains viewed parallel to the chain axis. A: Experimental chain motif propagated by a glide plane; B: Predicted chain motif propagated by a 2_1 screw axis

Investigation of the extent of polymorphism in the test molecules was not explicitly addressed in any of the blind tests. Estimates that approximately one third of organic molecules exhibit polymorphism under normal temperature and pressure conditions,[225] makes potential for polymorphism in the blind test molecules a significant factor given the impact it would have on the outcome of the blind tests. For each of the blind test molecules a single crystal structure was determined and used for comparison with the hypothetical structures. Within the total test set of nine molecules used in the three blind tests, so far two have been found to exhibit polymorphism. The most simple test molecule used in CSP1999, 3-oxabicyclo(3.2.0)hepta-1,4-diene, was found to be polymorphic during the initial crystal structure determination. It initially crystallised in space group *Pbca* but subsequently could only be obtained in an alternative crystal packing in $P2_1/c$. Both experimental structures were considered during comparison with the predicted structures and all of the four 'correct' predictions for this molecule were of the metastable *Pbca* form. Had the metastable form not been observed, no correct predictions would have been made. 6-Amino-2-phenylsulfonylimino-1,2-dihydropyridine, the flexible test molecule from CSP2001, was found two years later to have a second polymorph[226] after one of the participants became convinced that there was a new polymorph with a different hydrogen bonding motif to be found. This new structure contained two molecules in the asymmetric unit, placing this outside the limits set by the CCDC on the predicted structures, and consequently did not lead to any further 'correct' predictions in retrospect.

The prevalence of hydrogen bonded dimer-based structures for BQT in the predicted structures submitted during CSP2001, which were found to be energetically competitive with the experimental chain-based crystal structure, and the lack of any previous experimental investigation into its polymorphism provided the inspiration for an automated crystallisation screen on BQT to attempt to discover dimer based polymorphs of this molecule.

6.3 Synthesis

BQT is not commercially available, and consequently it was synthesised during this study, specifically to facilitate the crystallisation screen. The synthesis was performed under the supervision of Dr. Colin Bedford (UCL Chemistry).

The procedure of Goodwin & Perkin[227] was used to convert the starting material, a mixture of cis/trans 1,3-cyclohexane dicarboxylic acid (Sigma-Adrich, Poole, UK), to the pure cis isomer. 75g of mixed cis/trans 1,3-cyclohexane dicarboxylic acid, was refluxed with 120ml of acetyl chloride for 1 hour at 80° C. Acetyl chloride and acetic acid were distilled off at 160° C, leaving a residue of pure 3-oxabicyclo[3,3,1]nonane-2,4-dione. Residual traces of acetic acid were removed by mild heating under strong vacuum (~1 mmHg). This anhydride was then dissolved in 150ml of hot distilled water to convert it to cis-1,3-cyclohexanedicarboxylic acid. The volume was reduced by approximately 20ml by heating and the mixture was cooled to room temperature, with constant stirring. The product was filtered off to leave pure cis-1,3-cyclohexane dicarboxylic acid. From 75g of the starting mixture of geometric isomers, 70.92g of pure cis acid was recovered, 94.6% yield.

The cis-1,3-cyclohexane dicarboxylic acid was converted to the final product after the method of Hall.[228] 40ml of 30% aqueous ammonium hydroxide was added to 25g of cis-1,3-cyclohexane dicarboxylic acid, with ammonia released by the reaction. The water was distilled off and the residual imide was distilled using a free flame under ~15 mmHg vacuum into an air condenser. The distilled product was a wet white solid, which was washed with water and mechanically recovered into a beaker. The water was strongly acidic due to the presence of cis-1,3-cyclohexane dicarboxylic acid from the hydrolysis of the minor by-product 3-oxabicyclo[3,3,1]nonane-2,4-dione. The acid solution was neutralised using 10% sodium hydroxide and the product was taken up in

50 ml of chloroform. The water layer was washed with 2x50ml portions of chloroform. 9.41g of BQT was recovered upon evaporation of the chloroform solution to dryness, 42% yield. 37.7g was made using this method.

6.4 Automated crystallisation screen

The Chemspeed Accelerator SLT100 (Augst, Switzerland) automated parallel crystallisation platform located at Strathclyde Institute of Biomedical Science was used in preference to manual crystallisation methods as it allowed a large number of crystallisations to be carried out in a short period of time along with giving a high degree of control over the exact crystallisation conditions.[229] This work was carried out under the guidance of Dr. Andrea Johnston (Strathclyde University). A solvent library containing 67 solvents with a wide range of physico-chemical properties[59] formed the principal basis of variation in the crystallisation screen and three different crystallisation methods were employed.

A schematic diagram of the crystallisation apparatus is shown in figure 6.4. Up to 32 crystallisations, incorporating different solvents, but the same crystallisation conditions could be performed in parallel. To allow for the possibility of a novel crystal structure based on a dimer motif, crystalline seeds of the starting form of BQT were rigorously excluded from all crystallisation vessels by a specifically designed filtration step between the dissolution and crystallisation stages. All solutions were filtered through a 10 μm filter at the same temperature as the dissolution stage. As soon as the solution had been filtered into the crystallisation vessel, it was subjected to a rapid heating cycle to dissolve any seeds smaller than 10 μm, before entering the crystallisation regime. This method, as far as it is possible, removed all crystalline seeds, forcing the solution to crystallise via primary nucleation or secondary nucleation only on the surface of the crystallisation vessel.

Vacuum to draw over solution

10 µm filter

Filtration followed by brief heat cycle to dissolve smaller seeds

Dissolution vessel with automatically dispensed solid and solvent

Crystallisation vessel

Manual transferral of crystallisation product to multi-sample tray for powder XRD

Heating and agitation
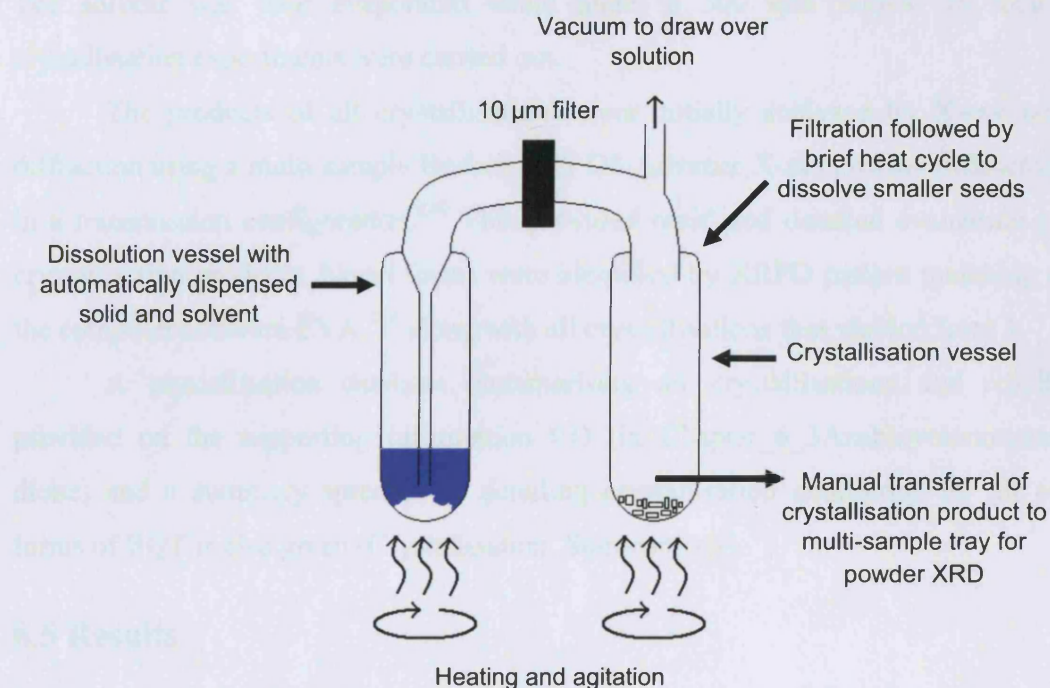
Scheme 6.1: Schematic representation of the automated crystallisation process. This apparatus is duplicated within the automated crystallisation platform

To facilitate high temperature dissolution in the range of solvents available, the 67 available solvents were divided into eight groups of eight and a group of three according to descending boiling point. For each group a temperature T_{max} was chosen just below the boiling point of the lowest boiling member of the group, at which the dissolution step could be carried out without significant solvent vaporisation. The solvent groups and associated T_{max} temperatures are given in the supporting information.

The first crystallisation method involved saturation of 2 ml quantities of each solvent with BQT at its T_{max}, followed by filtration of the solutions at T_{max} and controlled cooling to 15° C at >20 °C min^{-1}, while subject to a 1000 rpm vortex. The second crystallisation method involved dissolution of 100 mg of BQT in 2 ml of solvent at T_{max} to produce sub-saturated solutions, which were filtered at T_{max}, and subsequently cooled to 15 °C at >20 °C min^{-1} under an 850 rpm vortex. The third crystallisation method involved dissolution of 100 mg of BQT in 2 ml of solvent at 75 °C, followed by filtration and transferral of each solution to a crystallisation vial on a hot plate at 75 °C.

The solvent was then evaporated while under a 500 rpm vortex. In total 181 crystallisation experiments were carried out.

The products of all crystallisations were initially analysed by X-ray powder diffraction using a multi-sample Bruker-AXS D8-Advance X-ray powder diffractometer in a transmission configuration.[230] This provided rapid and detailed evaluation of the crystallisation products. Novel forms were identified by XRPD pattern matching using the computer software EVA,[231] along with all crystallisations that yielded form 1.

A crystallisation database, summarising all crystallisations and results, is provided on the supporting information CD (in Chapter_6_3Azabicyclononane-2,4-dione) and a summary spreadsheet detailing crystallisation conditions for all of the forms of BQT is also given (Crystallisation_Summary.xls).

6.5 Results

The synthesis of BQT gave a crystalline product which was identified as form 1 by XRPD. BQT was soluble in all 67 solvents at each solvent's T_{max} and the product of the vast majority of crystallisations in the screen was form 1. The crystallisation screen produced two solvates, containing acetic acid and 1-methylnaphthalene and a new metastable polymorph (form 2). Thermal analysis of the synthesised material also yielded form 2 and a high temperature plastic crystalline phase (form 3). A spreadsheet detailing all crystallisations and the identity of the products is provided in the supporting information.

3-Azabicyclo[3.3.1]nonane-2,4-dione acetic acid (1/1)[232]

The acetic acid solvate precipitated from a controlled cooling crystallisation as block crystals of sufficient size for SXRD and its crystal structure was fully determined. It crystallised in the triclinic space group $P\bar{1}$ with one molecule of BQT and one molecule of acetic acid in the asymmetric unit (figure 6.4 and table 6.2).
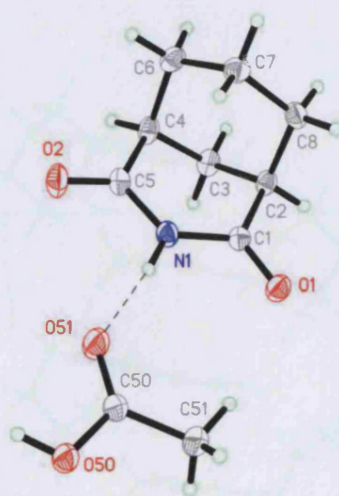
Figure 6.4: Asymmetric unit of BQT-acetic acid solvate

The crystal structure is comprised of hydrogen bonded chains of alternating BQT and acetic acid molecules. The hydrogen bonded back-bone of the chain (figure 6.5) is very similar to that found in BQT form 1 and both polymorphs of acetic acid.[233;234] The BQT-acetic acid chain is a modification of the BQT form 1 chain with the carboxylic acid group of the acetic acid mimicking the role of the BQT amide functionality in the hydrogen bonded chain.
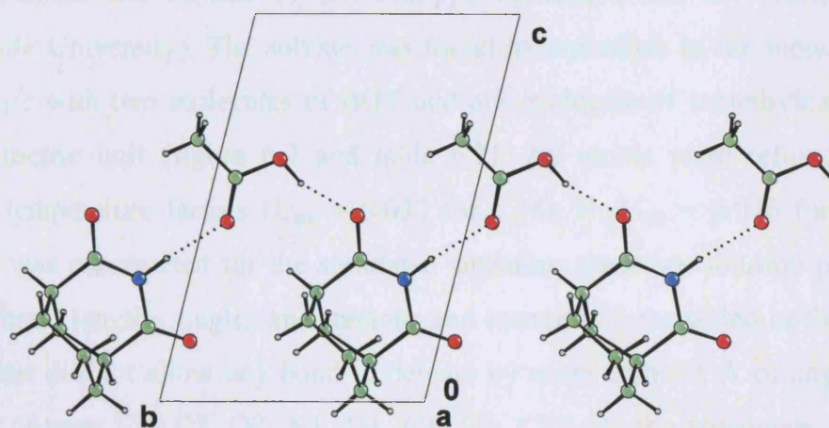
Figure 6.5: Hydrogen bonded chain present in the BQT acetic acid solvate

The chains propagate parallel to the *b* crystallographic axis and stack parallel to the *a* axis in an ABAB repeating pattern. This allows the bulky cyclohexane ring present in

the BQT molecules to alternate from side to side in the stack, allowing efficient packing of the chains (figure 6.6).
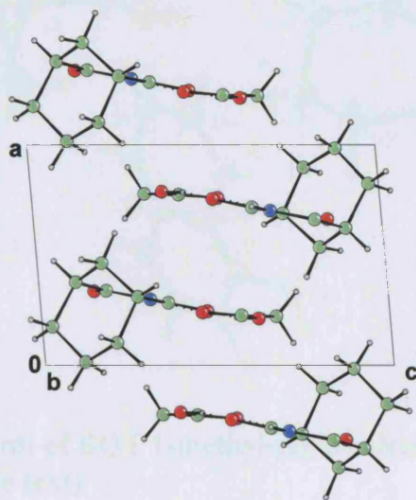
Figure 6.6: ABAB stacking of the BQT acetic acid chains

3-Azabicyclo[3.3.1]nonane-2,4-dione 1-methylnaphthalene (2/1)[235]

The 1-methylnaphthalene solvate crystallised with a fine needle morphology proving unsuitable for single crystal X-ray diffraction, and consequently an XRPD data set on a capillary transmission diffractometer was collected and from this data set the structure was determined and refined by Dr. Philippe Fernandes and Dr. Alastair Florence (Strathclyde University). The solvate was found to crystallise in the monoclinic space group $P2_1/c$ with two molecules of BQT and one molecule of 1-methylnaphthalene in the asymmetric unit (figure 6.7 and table 6.2). All atoms were refined with fixed isotropic temperature factors ($U_{iso} = 0.032$ for C, O, N; $U_{iso} = 0.076$ for all H). The molecule was constructed for the simulated annealing structure solution process using standard bond lengths, angles and torsions and restraints were added in the refinement process that did not allow any bond to deviate by more than 0.1 Å or angles by more than 0.8°. Atoms C4, C5, O2, N1, H1, C1, O1, C2 (and the analogous group in the second BQT molecule) were restrained to be planar. The 1-methylnapthalene molecule was also restrained to be planar.

207
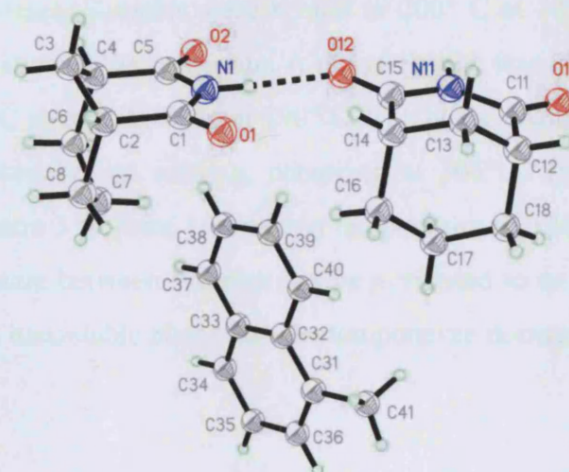
Figure 6.7: Asymmetric unit of BQT 1-methylnaphthalene. All temperature factors are fixed and isotropic (see text)

The two symmetry independent BQT molecules form a hydrogen bonded chain very similar to that of forms 1 and 2, and, similarly to form 2, the chain exhibits a pseudo-glide plane parallel to the *b* axis. The 1-methylnaphthalene molecules occupy channels with each channel surrounded by four BQT chains (figure 6.8). The plane of the 1-methylnapthalene molecules, parallel to the *ac* plane, is perpendicular to the direction of the channels, parallel to the *b* axis.
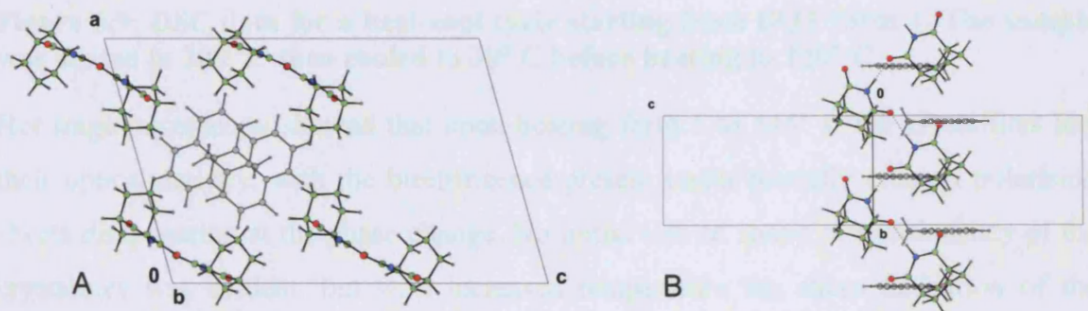
Figure 6.8: BQT 1-methylnaphthalene solvate. A: view parallel to the b axis showing the channels containing 1-methylanphthalene molecules; B: hydrogen bonded BQT ribbons shown coloured by element with 1-methylnaphthalene molecules coloured grey

3-Azabicyclo[3.3.1]nonane-2,4-dione form 3

Combined DSC and TGA were carried out on BQT form 1, as synthesised. The instrument used was a Netzsch STA449C, with a liquid nitrogen attachment to allow

208

control of cooling phases. Samples were heated to 200° C at 10° C min⁻¹ and cooled back to room temperature at the same rate. A phase change was observed with an onset temperature of 137° C prior to melting at 190° C. The phase change from form 1 to this new form 3 was reversible on cooling, occurring at 106° C (figure 6.9). A second sample exhibited a form 3 to form 1 transition temperature of 124° C and the variation in transition temperature between samples can be attributed to an undercooling effect - the persistence of the metastable phase into the temperature domain of the stable form.
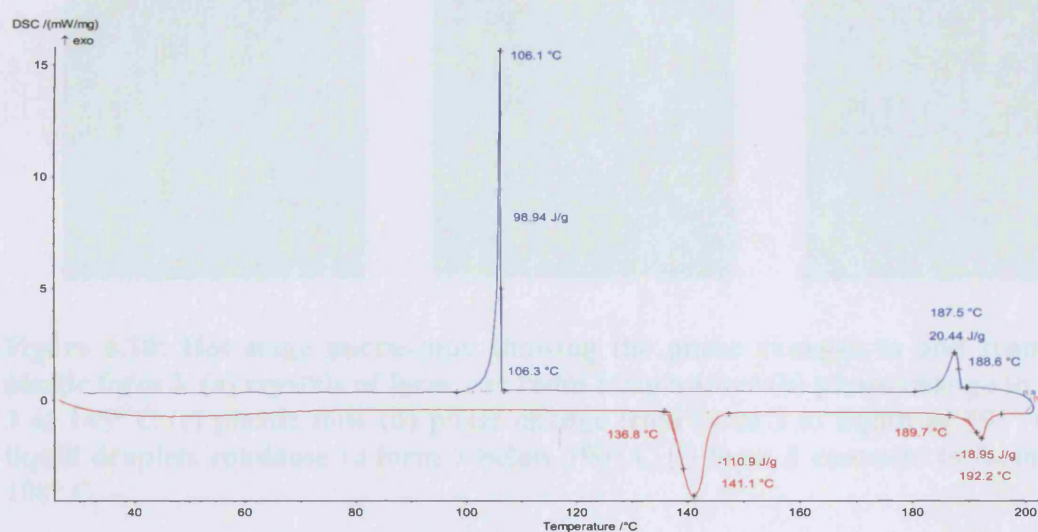
Figure 6.9: DSC data for a heat-cool cycle starting from BQT form 1. The sample was heated to 200° C then cooled to 30° C before heating to 220° C

Hot stage microscopy showed that upon heating form 1 to 145° C the crystallites lost their optical activity, with the birefringence present under partially crossed polarising sheets disappearing at the phase change. No initial loss of shape or translucency of the crystallites was evident, but with increased temperature the sharp definition of the separate crystallites was lost as the new phase began to slowly 'flow'. This plastic flow continued until the only features present in the sample were the boundaries between what were originally separate crystallites. At 195° C the plastic phase melted and droplets of liquid were formed. With cooling the liquid condensed into the plastic phase, which then crystallised sharply at 108° C (figure 6.10). On this evidence it was concluded that form 3 is a plastic crystalline phase.[236]
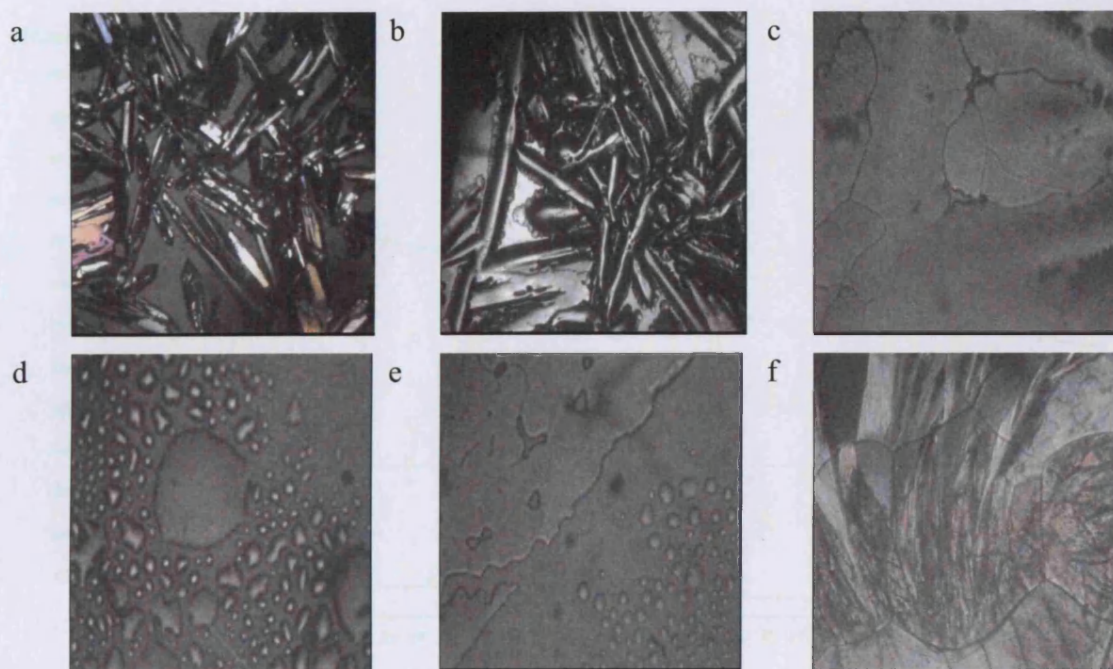
Figure 6.10: Hot stage microscopy showing the phase changes to and from the plastic form 3. (a) crystals of form 1 at room temperature (b) phase change to form 3 at 145° C (c) plastic flow (d) phase change from form 3 to liquid at 195° C (e) liquid droplets condense to form 3 below 190° C (f) form 3 converts to form 1 at 108° C

XRPD data were collected by Dr. Alastair Florence on form 3 above 145° C. The XRPD pattern contains seven distinct peaks (2θ (°) = 16.511(3), 23.424(3), 28.803(4), 33.385(1), 37.455(3), 41.200(3), 44.662(3), figure 6.11) which are present at all temperatures below the melting point of the plastic phase.

Intensity
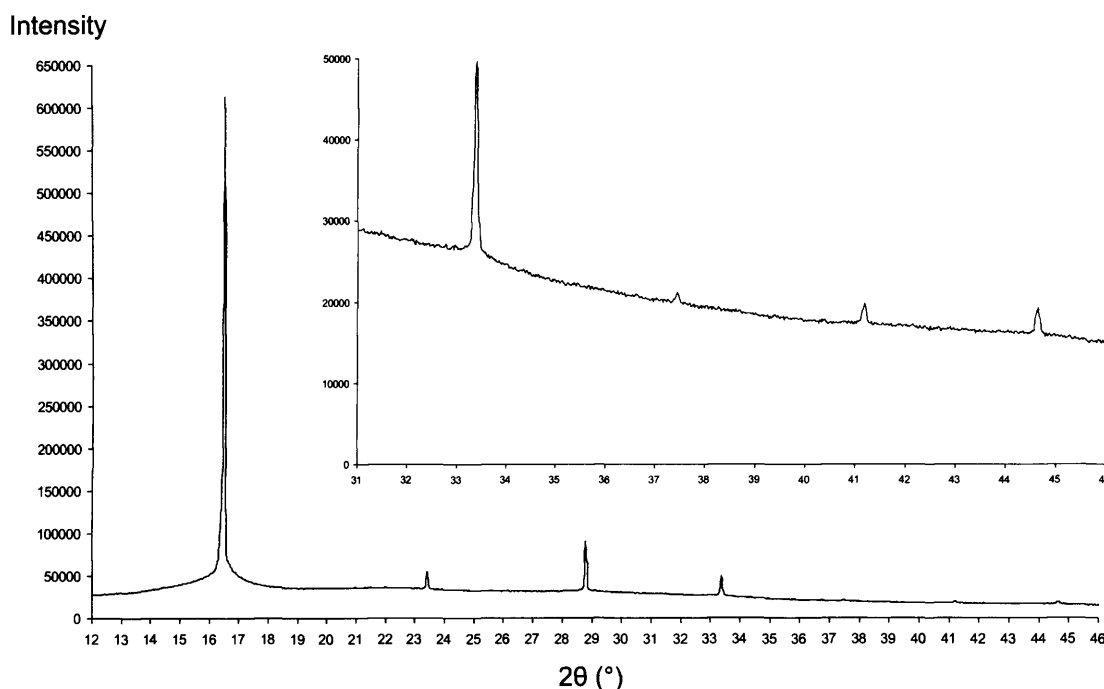
Figure 6.11: X-ray powder diffraction pattern of the form 3 plastic crystalline phase (12-46°). Insert shows high angle region (31-46°)

3-Azabicyclo[3.3.1]nonane-2,4-dione form 2[237]

Fast cooling of the plastic form 3 to –28° C produced a metastable polymorph, which was kinetically trapped for long enough to collect a XRPD data set. From this data set the structure was determined and refined by Dr. Philippe Fernandes and Dr. Alastair Florence. All atoms were refined with fixed isotropic temperature factors (U_{iso} = 0.038 for C, O, N; U_{iso} = 0.076 for all H). The molecule was constructed for the simulated annealing structure solution process using standard bond lengths, angles and torsions and restraints were added in the refinement process that did not allow any bond to deviate by more than 0.1 Å or angles by more than 1°. Atoms C4, C5, O2, N1, H1, C1, O1, C2 (and the analogous group in the second BQT molecule) were restrained to be planar. The unit cell dimensions of form 2 are closely related to those of form 1 (table 6.1), with the crystallographic *c* axis of the unit cell doubled with respect to the form 1 structure, with the cell now containing two molecules in the asymmetric unit (table 6.2 and figure 6.12).

211

	S. G.	a (Å)	b (Å)	c (Å)	β (°)
Form 1	$P2_1/a$	7.7046(5)	10.6062(6)	9.3384(2)	95.033(2)
Form 2	$P2_1/c$	7.6710(2)	10.5483(2)	18.8867(4)	95.580(1)

Table 6.1: Comparison of the unit cell dimensions of forms 1 and 2
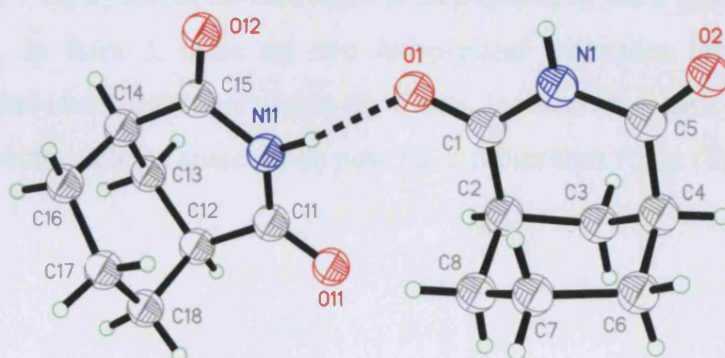
Figure 6.12: Asymmetric unit of BQT form 2. All temperature factors are fixed and isotropic (see text)

The powder pattern of form 2 is closely related to that of form 1, with a number of extra peaks (figure 6.13).
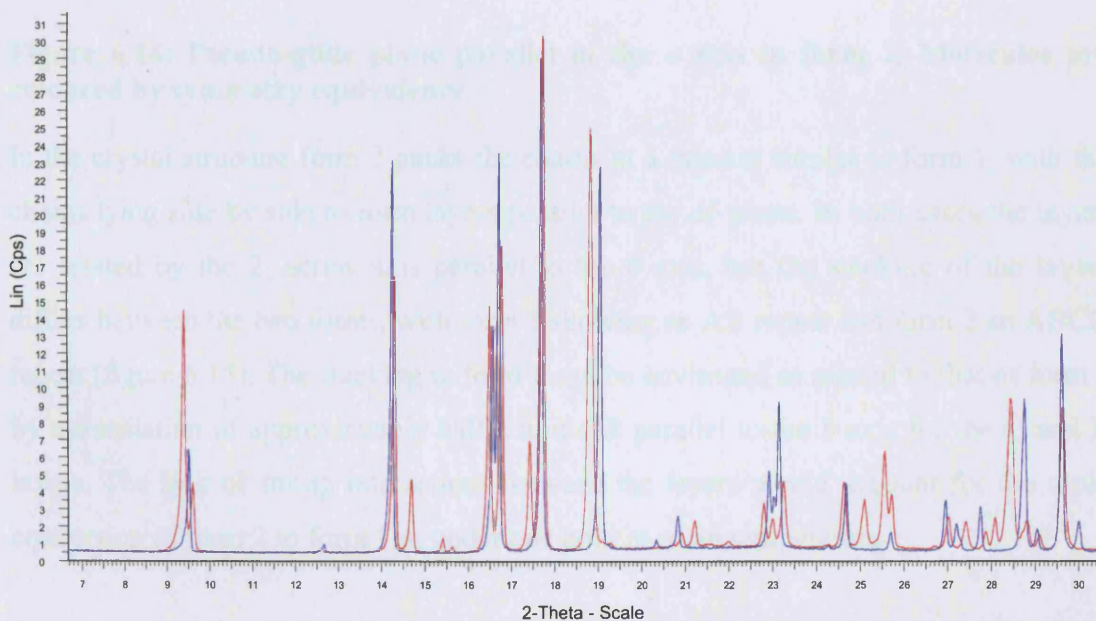
Figure 6.13: Overlay of experimental XRPD patterns for form 2 (red) and form 1 (blue)

Form 2 was also observed twice during the crystallisation screen, crystallising from chloroform and carbon tetrachloride, both by fast cooling of solutions saturated at 40° C and 70° C respectively. This highly metastable, kinetically stabilised form is an example of a trapped Ostwald phase,[238] a metastable phase initially formed on crystallisation that subsequently transforms to a more stable form.

In form 1 the hydrogen bonded chain is propagated, by the *a* glide plane, parallel to the *a* axis. In form 2, there are two independent molecules forming the same hydrogen bonded chain, again parallel to the *a* axis, but here the chain propagates by a pseudo-glide plane, with the space group now $P2_1/c$ rather than $P2_1/a$ (figure 6.14).
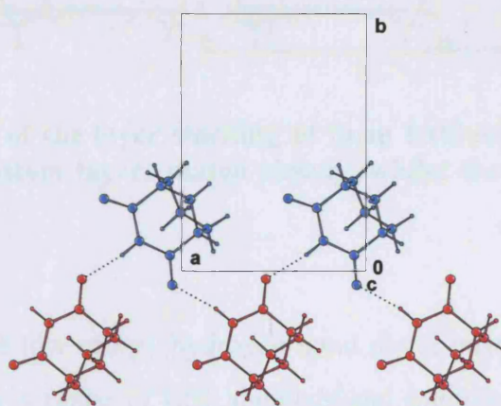
Figure 6.14: Pseudo-glide plane parallel to the *a* axis in form 2. Molecules are coloured by symmetry equivalence

In the crystal structure form 2 packs the chains in a manner similar to form 1, with the chains lying side by side to form layers parallel to the *ab* plane. In both cases the layers are related by the 2_1 screw axis parallel to the *b* axis, but the stacking of the layers differs between the two forms, with form 1 showing an AB repeat and form 2 an ABCD repeat (figure 6.15). The stacking in form 2 can be envisaged as related to that of form 1 by a translation of approximately half a unit cell parallel to the *b* axis for the C and D layers. The lack of strong interactions between the layers would account for the rapid conversion of form 2 to form 1 in under one hour at room temperature.
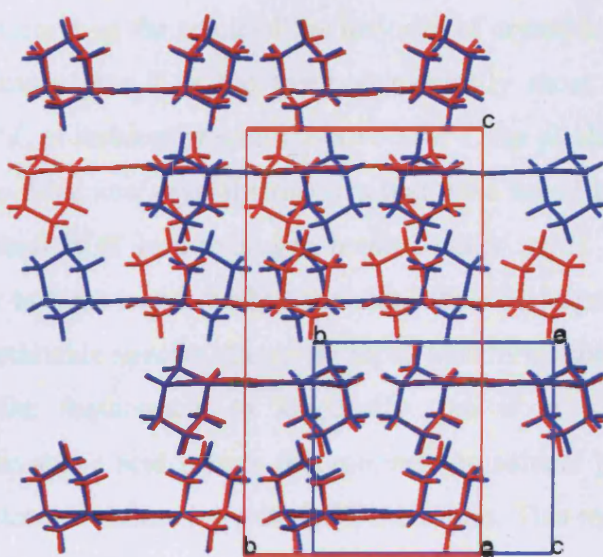
Figure 6.15: Overlay of the layer stacking of form 1 (blue) and form 2 (red). The top layer and two bottom layers match closely, whilst the intermediate layers do not

6.6 Discussion

The predicted plausible low energy hydrogen bond dimer-based hypothetical structures for BQT, produced by a range of CSP methods and intermolecular potentials during CSP2001, has not been realised by experimental research. Rather a new metastable chain-based polymorph has been discovered.

The crystallisation methods employed in this study provided for a range of conditions under which a new polymorph could have the opportunity to form. The filtration step removed, in so far as it is possible, all pre-existing crystalline seeds from the crystallisation process. The high temperature dissolution and the heat pulse after filtration removed the possibility of 'crystal memory' – small clusters of undissolved BQT molecules which retained the form 1 crystal structure. The wide range of physico-chemical properties exhibited by the solvent range used provided the opportunity for different self-assembled units at the pre-nucleation stage, possibly leading to different crystallisation outcomes. In sub-saturated non-polar solvents it could be envisaged that the first association between BQT molecules would be the formation of hydrogen bonded dimers. However, even from non-polar solvents in a sub-saturated regime, the form 1 crystal structure was produced. The formation of chain based crystal structures from all solvents using the different crystallisation methods requires an explanation.

The production of form 1 as the result of the majority of crystallisations in this screen leads to the conclusion that it is the thermodynamically most stable phase below approximately 140° C at ambient pressure. Above 140° C the plastic form 3 is the most stable phase. Of the three new crystal structures that were found by the crystallisation screen, none contained BQT in a hydrogen bonded dimer motif. Form 2 has a close structural similarity to form 1 and exhibits the same chain hydrogen bond motif. It can be considered a metastable species due to its rapid transformation to form 1 at room temperature, and the requirement to kinetically trap it at low temperature for characterisation. The acetic acid solvate incorporates the solvent into the chain motif, with acetic acid molecules alternating with BQT molecules. This result is not surprising given that the supramolecular synthons present in BQT form 1 and in both of the polymorphs of acetic acid are very similar and the hydrogen bonded chain found in this solvate can be envisaged as an 'average' of the two parent supramolecular synthons. The 1-methylnapthalene solvate contains the same hydrogen bonded chains of BQT as found in forms 1 and 2.

Plastic crystalline phases such as form 3 are well known for globular molecules, and they commonly express cubic symmetry.[236] The most well known molecules that form plastic crystalline phases are adamantane and camphor and plastic phases have been found in molecules structurally similar to BQT such as bicyclo[3.3.1]nonane,[239] bicyclo[3.3.1]nonan-9-one[240] and 3-oxabicyclo[3.3.1]nonane-2.4-dione. The plastic phase has been contrasted with liquid crystals by Timmermans.[236] He envisaged the melting process as consisting of two parts both of which are a result of thermal motion: the free rotation of the molecule and the loss of coherence of the crystal (the translational repeat order that forms the lattice). In liquid crystals the coherence of the crystal is lost before isotropic rotation of the molecules can occur, due to the strongly anisotropic shape of the molecules. In plastic crystals the free rotation is facilitated first, due to the globular conformation of the molecules, with the breaking of the crystalline lattice occurring at higher temperature as a separate event. The large librational motion of the molecules in a plastic crystal accounts for the rapid decrease in the intensity of X-ray scattering with angle. The Debye factor relating the observed intensity to the intensity for non-vibrating atoms is

$$I = I_R \exp(\frac{-B\sin^2\theta}{\lambda^2}) \text{ where } B = 8\pi^2\overline{\mu^2}$$

where μ^2(bar) is the mean square amplitude of vibration normal to the reflection plane.[241] At higher angles of diffraction the intensity decreases rapidly, leading to a small number of observed Bragg peaks, consistent with the observation of only seven peaks in the XRPD pattern of form 3. The seven peaks in the XRPD pattern are consistent with a cubic cell, space group $I2_3$, $a = 7.5856(1)$ Å, $R_{wp} = 0.016$). The unit cell has a volume of 436.49(1) Å^3 and when compared to the volume of 190 Å^3 per molecule for BQT form 1, it can be assumed that the unit cell contains two molecules, with the approximate 15% increase in volume per molecule consistent with the introduced disorder. Two molecules in the unit cell would be consistent with body centring.

Two models for the disorder in plastic crystals have been proposed.[242] In the Pauling model[243] the molecules are free rotors with no well defined energy minima. In the Frenkel model[244] the crystal is considered disordered, with the molecules distributed over a number of discrete energy minima. The molecules show significant libration at each energy minimum and can step between the discrete minima. It has been noted that X-ray diffraction would not be able to distinguish between free rotation and a range of discrete orientations.[245] The largest diameter that can be swept out by a BQT molecule rotating around its centre of mass is approximately 8.6 Å (calculated from the distance from the centre of mass to H9 = 3.10 Å and the van der Waals radius for hydrogen = 1.2 Å). Comparison of the diameter of a BQT molecule to the unit cell size suggests that each molecule does not have sufficient space to freely rotate and the Frenkel model is appropriate in this case.[241] Guthrie and McCullough[245] have attempted to relate the entropy of transition to the number of discrete molecular orientations based on the Frenkel model. In this model it is assumed that the entropy of transition is caused solely by the disordering of the molecules across the discrete orientations and that the lattice modes and molecular vibrational modes do not contribute significantly to the entropy of transition. The relation is:

$$\Delta S_{trs} \approx R \ln\left(\frac{N_{plastic}}{N_{ordered}}\right)$$

where $N_{plastic}$ is the number of discrete orientations in the plastic phase and $N_{ordered}$ is the number of discrete orientations in the low temperature fully crystalline phase, in this case 1. Seven DSC heating measurements of the form 1 to plastic phase transition and plastic melting events were recorded. The averages of these data gave an onset temperature for the form 1 to plastic transition of 135.6 ± 0.4° C, with ΔH_{trs} = 16.3 ± 0.7 kJ mol^{-1}, leading to ΔS_{trs} = 39.9 ± 1.9 J K-1 mol^{-1}. For the plastic melting event, the average onset temperature from six measurements was 190.6 ± 0.3 ° C with ΔH_{trs} = 3.27 ± 0.2 kJ mol$_{-1}$ and ΔS_{trs} = 7.05 ± 0.43 J K-1 mol^{-1}. The entropy of the form 1 to plastic transition therefore suggests that there are 120 discrete orientations in the plastic phase, but this is very much an approximation as some of the entropy of transition should be apportioned to the librational motion and consideration of the errors suggests a range of 100 to 150 orientations. This figure is similar to that calculated for heptacyclotetradecane, using a modified version of the above method, of 114 discrete orientations.[246] BQT is a rare example of an organic compound that has a high temperature plastic phase and also has hydrogen bond donors and acceptors that can form conventional hydrogen bonds.[247] Calculation of the solvent-accessible electrostatic surface of BQT by Gareth Welch (University College London) has shown the N-H group to be a weak hydrogen bond donor, with the electrostatic potential around this hydrogen not particularly more positive than around the hydrogen atoms on the cyclohexane ring. This weakness of the hydrogen bonds is consistent with the formation of the plastic phase, in which these hydrogen bonds must be broken. The related compound, 3-oxabicyclo[3,3,1]nonane-2,4-dione also shows a plastic transition, but at lower temperature, onset 74° C, consistent with the absence of even the weak hydrogen bonds seen in BQT.

6.7 Conclusion

A metastable polymorph, two solvates and a high temperature plastic phase have been discovered for the molecule 3-azabicyclo[3.3.1]nonane-2,4-dione, all based on hydrogen bonded chains. This is an example where our understanding of polymorphism has been complicated by the experience of applying CSP methods – the proposal of nearly equi-energetic hypothetical structures built from hydrogen bonded dimers

contrasts with the experimental situation where only crystal structures based on hydrogen bonded chains have been observed.

A greater knowledge of the processes of crystal growth – specifically the assembly of molecules in solution prior to nucleation and the competition between nuclei of different crystal structures early in the pre-critical nucleus stage – could provide the crucial reasoning as to why hydrogen bond dimer based structures have not been discovered. The weakness of the N-H$\cdots$O hydrogen bond, as implied by the formation of the plastic phase and the calculations of the electrostatic potential around the molecule, and the globular shape of the molecule provides the basis for a hypothesis why dimer-based structures are not observed: both of these factors would allow the molecules in crystal nuclei to reorientate easily, allowing rearrangement into the more thermodynamically stable chain-based structure that is observed (form 1, and form 2 as a modification of form 1). The ease of this conversion would make the trapping of a dimer-based form very unlikely. This selectivity of the crystallisation pathway is not considered in the computational crystal structure prediction method. Those structures that are thermodynamically plausible but are selected against by the crystallisation process will remain in the set of hypothetical structures generated by CSP for as long as lattice energy is the only factor used to discriminate between hypothetical structures.

Further polymorphs may be waiting to be discovered, possibly by using novel crystallisation conditions such as growth on polymer substrates[25] or templating,[53] though it should be noted that high pressure crystallisation techniques[66] which can yield novel structures, did not produce a new polymorph in this instance.[248]

Crystal Data

Compound name	BQT acetic acid	BQT 1-methylnaphthalene*	BQT form 2*
Empirical formula	$C_8H_{11}NO_2$, $C_2H_4O_2$	$C_8H_{11}NO_2$, $\frac{1}{2}(C_{11}H_{10})$	$C_8H_{11}NO_2$
Formula weight	213.2	224.28	153.2
Crystal system, space group	Triclinic, $P\bar{1}$	Monoclinic, $P2_1/c$	Monoclinic, $P2_1/c$
a (Å)	6.6224(7)	15.0236(2)	7.6710(2)
b (Å)	7.3580(8)	7.3230(1)	10.5483(2)
c (Å)	10.7995(12)	22.5164(3)	18.8867(4)
α (°)	103.598(2)	90	90
β (°)	93.378(2)	106.0201(6)	95.580(1)
γ (°)	97.272(2)	90	90
V (Å³)	505.2(1)	2380.99(6)	1521.00(6)
Z', Z	1, 2	2, 8	2, 8
D(calc (g cm⁻³))	1.402	1.251	1.338
Data Collection			
Crystal size (mm)	0.35 x 0.29 x 0.17	XRPD capillary	XRPD capillary
Temperature (K)	150(2)	295(2)	250(2)
hkl range (h, k, l)	-8→8, -9→9, -14→13	0→10, -4→4, -14→14	0→5, -6→-6, -12→12
Reflections measured, R_{int}	4424, 0.0126	736, -	448, -
Independent reflections	2313	-	-
Reflections I>2σ(I)	2121	-	-
Refinement			
Parameters refined	196	211	142
R(F) (I>2σ(I))	0.039	0.044 (R_p)	0.054 (R_p)
wR(F²) (all reflections)	0.098	0.053 (R_{wp})	0.070 (R_{wp})
Residual electron density (min, max (e Å⁻³))	0.35, -0.18	-	-

Table 6.2: Crystal structure summary of all BQT crystal structures included in this chapter. * Crystal structures determined by XRPD by Dr. Philippe Fernandes and Dr. Alastair Florence (Strathclyde University), [235;237] **Cu Kα radiation, λ = 1.54056 Å**

Chapter 7 – Crystal structure prediction of monohydrates

7.1 Introduction

7.1.1 The significance of hydrates

The potential of a compound to form crystalline hydrates is of particular significance in fine chemical development because water is commonly used in many chemical manufacturing processes and the water molecule is a versatile hydrogen bond donor and acceptor which can often lead to the formation of a hydrate. It has been reported that approximately one third of organic molecules are capable of forming hydrates,[19] and figures as high as 75% have been suggested for pharmaceutical compounds.[187]

No conceptual difference exists between hydrates and other solvates – all solvates incorporate the solvent into the crystal structure in addition to the parent molecule. Some molecules can form hydrates in different hydration states, each with different stoichiometric ratios of water to the parent molecule, including hemihydrates (1:2), monohydrates (1:1) and dihydrates (2:1). Each different hydration state is classified as a different solvate of the parent molecule, and where two hydrates have the same stoichiometric ratio, but different crystal structures, the two forms can be classified as polymorphic hydrates.

In fine chemical development, the likelihood and potential implications of producing hydrates must be thoroughly understood. Factors that can affect the state of hydration of a drug substance include manufacturing processes, changes in environmental humidity and time.[11] Many manufacturing processes such as aqueous granulation, spray drying, crystallisation and aqueous film-coating may bring the drug substance into contact with water, providing the opportunity for a hydrate to form.[11] Even after the drug substance has been formulated into a drug product, there may exist the opportunity for a phase change to a hydrate by incorporating water from hydrated excipients by water redistribution within the dosage form or by incorporating water from the air in humid environments.[249]

Byrn[2] has summarised why solvates require investigation:

- They are commonly the penultimate form of a drug, prior to desolvation to the drug form used in manufacture, and consequently control over the formation of the solvate is required.

- They may be chosen for recovery or purification processes.

- They may exhibit physical properties, such as morphology, that are important during particular manufacturing steps.

- They may be the only form of a drug substance that is crystalline in the solid state, allowing determination of the drug substance structure by X-ray diffraction.

- The desolvated form may be used in the drug product because of attractive physical properties such as dissolution rate and bioavailability.

- They may be a route to increased patent protection of the drug substance.

The different chemical composition and crystal packing in a hydrate compared to an anhydrous form of the same molecule will cause them to have different physical and chemical properties. The internal energy of the hydrate *versus* the anhydrate is altered by the different intermolecular interactions, and the entropy can be altered significantly if the water is disordered. These changes alter the free energy of the crystal structure which can ultimately affect the solubility and dissolution rate of the hydrate.[249] A hydrate will always be less soluble in water than the corresponding anhydrate: in the hydrate crystal structure the parent molecule already has water contacts and the dissolution process yields fewer new parent molecule···water contacts, leading to a lower free energy of dissolution.[249]

7.1.2 The formation of hydrates

Water is a common solvate-forming solvent because of its versatile hydrogen bonding capability, enabling improved intermolecular interactions in the hydrate crystal structure compared to the anhydrate, commonly by addressing hydrogen bond donor/acceptor imbalances. Its size allows water to fill small voids in crystal structures and its ability to form up to two donor hydrogen bonds and two acceptor hydrogen bonds allows it to simultaneously hydrogen bond to several groups from the parent molecule or other water molecules. Water commonly forms a tetrahedron of hydrogen bonds with the oxygen lone pairs acting as hydrogen bond acceptors and the two hydrogen atoms acting as hydrogen bond donors (figure 7.1). The charge density of the oxygen atom is more accurately represented as a single diffuse lobe of electron density[250] and the tetrahedral geometry, as found in all polymorphs of ice, is an artefact of donor/acceptor balancing to give two donors and two acceptors.
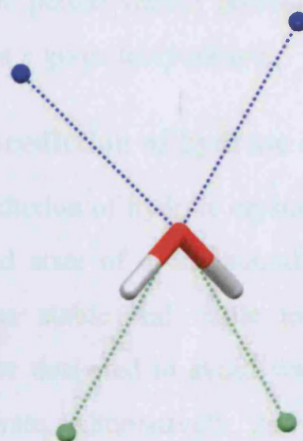
Figure 7.1: Tetrahedral water hydrogen bond geometry. The water molecule forms four hydrogen bonds, acting as a hydrogen bond donor twice (green) and as a hydrogen bond acceptor twice (blue)

Water does not have to exhibit this tetrahedral hydrogen bond geometry in hydrate crystal structures. A recent study[251] has identified a range of water hydrogen bonding geometries found in hydrate structures present in the CSD, from a single hydrogen bond up to the maximum of four. 2-, 3-, and 4- co-ordinated geometries were all present in significant numbers of the organic hydrate structures investigated. It has been noted[252] that hydrate crystal structures are more prevalent for organic molecules in which there is

a donor/acceptor imbalance particularly when there are fewer donors than acceptors and the inclusion of the water addresses this imbalance.

In solution the activity of the water is critical in determining whether a hydrate or anhydrate forms and if a hydrate is favoured, which level of hydrate is formed. The equilibrium between the hydrate and anhydrate is given by the equation:[11]

$$A_{(solid)} + m.H_2O \overset{K_h}{\Longleftrightarrow} A.mH_2O_{(solid)}$$
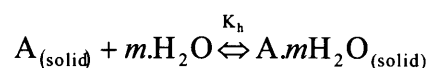
with the equilibrium constant defined as:

$$K_h = \frac{a[A.mH_2O_{(solid)}]}{a[A_{(solid)}]a[H_2O]^m}$$

The equilibrium constant depends on the activity of the water in the solution, which can be expressed as the ratio of the partial vapour pressure of water in the hydrate to the vapour pressure of pure water at a given temperature.

7.1.3 Terminology for the prediction of hydrate crystal structures

The reliable computational prediction of hydrate crystal structures would be a powerful tool in understanding the solid state of a compound. For cases where one or more hydrates were predicted to be stable and liable to be found experimentally, the manufacturing process could be designed to avoid water as a solvent so negating the possibility of forming the hydrate. Alternatively, the prior knowledge of the possible hydrates that could form could be exploited to help control the formation of the hydrate and aid in its identification.

The crystal structure prediction of hydrates is not technically distinct from the crystal structure prediction of other solvates, co-crystals or single molecule systems with $Z' > 1$. In all cases the prediction requires generation of hypothetical structures with more than one molecule (identical or non-identical) in the asymmetric unit. Van Eijck and Kroon[253] have provided several definitions for the formulation of this problem:

- $Z' = Z/M$ where Z is the number of residues (e.g. the unit [A.H$_2$O] for a monohydrate) in the unit cell and M is the multiplicity of the space group.

- Z'' = the number of crystallographically inequivalent molecules. An asymmetric unit containing two half molecules at special positions (eg: mirror plane or inversion centre) will have $Z' = 1$ but $Z'' = 2$.

- G = the number of independent molecules used for the CSP search. For a system where two symmetry independent chemically identical molecules are used as the basis for the search (G = 2), a sub-set of the generated structures may gain extra symmetry, and can be expressed in a space group incorporating the extra symmetry with $Z' = 1$. These structures have G = 2, but $Z' = Z'' = 1$.

A CSP search for a 1:1 molecular adduct will require G = 2 (one molecule of species 1 plus one molecule of species 2), and will commonly generate $Z' = 1$ structures, where one unit is defined as containing one molecule of each species. Even assuming that the molecules are rigid, the dimensionality of the crystal structure prediction of systems with G > 1 is significantly higher than for the case where G = 1. For each additional G, six extra variables are added to define the position and orientation of the extra molecule. For the prediction of hydrates, with no prior information about the stoichiometric ratio, several different values of G would have to be investigated and for some values of G different ratios would have to be considered. A search with G = 2 (1 parent molecule : 1 water molecule) would only generate monohydrates. G = 3 (1 parent molecule : 2 water molecules) would generate dihydrates but G = 3 (2 parent molecules : 1 water molecule) would commonly generate hemihydrates with $Z' = 2$ (where $Z = A.\frac{1}{2}H_2O$).

In the only published work on the prediction of hydrate crystal structures, Van Eijck and Kroon[253] used a random search method to generate possible crystal structures for polyalcohols and carbohydrates. Seven of these molecules were chiral, limiting the search procedure to the four most common enantiomorphic space groups and in five of these searches the experimental structure was found. For two non-chiral molecules searches were carried out in the experimental space group plus the four most common enantiomorphic space groups, but neither search found the experimental structure. The presence of the experimental structure in the list of generated structures is the primary requirement for CSP, with the structure ranking and energy difference with the global minimum indicators of the performance of the intermolecular potential and the impact

of neglect of thermal effects. This work highlighted the requirement for the search procedure to thoroughly explore the crystal packing landscape in light of the extra complexity introduced by including more than one molecule in the search procedure.

This present study is the first use of the MOLPAK/DMAREL method to attempt to predict the crystal structure of a monohydrate system. Predicted monohydrate structures were generated for 5-azauracil, which has a single known monohydrate crystal structure. This search required prior testing of the FIT dispersion-repulsion potential to assess its ability to reliably reproduce intermolecular interactions involving water. It was tested, in conjunction with a MP2/6-31G(d,p)-derived distributed multipole electrostatic model, for its ability to reproduce four of the six ordered structures of ice and a range of 22 experimental hydrate crystal structures.

7.2 Validation of the intermolecular potential using the polymorphs of ice

7.2.1 The polymorphs of ice

The phase diagram of ice presently contains 14 distinct crystalline phases. In the majority of the phases the hydrogen atoms are disordered, with partial hydrogen occupancy along each of the four tetrahedrally arranged hydrogen bonds around each water molecule. Disordered phases include ices Ih (common ice), III, IV, V, VI, VII and XII. Ices II[254;255] and VIII[256] are high pressure ordered phases with no disordered analogues. Ice IX[257] is a nearly ordered modification of ice III. Ice XI[258;259] is a low temperature, ambient pressure modification of ice Ih that is proton ordered. Very recently low temperature ordered versions of ices V and XII have been reported.[260]

The ability of the FIT dispersion-respulsion potential to reproduce water⋯water interactions was a pre-requisite for its application to the energy minimisation of hydrate structures. Successful reproductions of the crystal structures of ice would satisfy this condition. The disordered polymorphs of ice were discounted from the set of ice structures chosen for this potential validation because DMAREL cannot energy minimise disordered structures. Four of the ordered structures of ice were chosen: ices II, VIII, IX and XI. The reproduction of ice XI was of most interest, as its stability domain (stable under 73 K at ambient pressure) most closely matched the conditions of the DMAREL energy minimisation process. Ices II and VIII were chosen because their

structures were known to be unique phases and fully ordered. Ice IX was considered using the fully ordered major component while ignoring the minor (*ca.* 4%) disordered component.[257]

7.2.2 The crystal structures of ices II, VIII, IX and XI

The crystal structure of ice II has been determined by neutron single crystal diffraction[255] using D_2O, crystallising in the trigonal space group $R\bar{3}$ with two complete independent molecules in the asymmetric unit. Each of the water molecules is tetrahedrally hydrogen bonded with the crystal structure primarily comprised of hydrogen bonded six-member water rings (of which ice Ih is exclusively comprised) but also contains smaller rings caused by collapsing of the six-fold rings at higher pressure.[250]

The crystal structure of D_2O in the ice VIII structure has been determined by Kuhs *et al.*,[256] and was found to crystallise in the tetragonal space group $I4_1/amd$, with the oxygen atom located on the mirror plane ($Z' = 0.5$). The high pressure required to generate this phase causes Ice VIII to exhibit interpenetrating sub-lattices of the six-member rings, with each water molecule tetrahedrally co-ordinated by four identical hydrogen bonds.[250]

The ice IX structure was determined using D_2O and single crystal neutron diffraction.[257] It crystallises in the space group $P4_12_12$, with one water molecule on a general position and a second molecule present on the twofold axis. This structure has $Z' = 1.5$ but $Z'' = 2$.

The structure of D_2O in the Ice XI modification has been determined by neutron powder diffraction.[259] Ice XI crystallises in the orthorhombic space group $Cmc2_1$ with two independent half molecules in the asymmetric unit. For one of the independent molecules each atom has half occupancy because the molecule lies on the mirror plane. For the second independent molecule, the oxygen lies on the mirror plane and has half occupancy but the single hydrogen it is bonded to in the asymmetric unit is located on a general position with the second hydrogen generated by the mirror plane. This leads to a structure with $Z' = 1$, but $Z'' = 2$. Both independent molecules are tetrahedrally hydrogen bonded, and the structure is comprised exclusively of six-member hydrogen bonded rings.

The bond lengths, H-O-H angles and the hydrogen bonding parameters for these structures of ice are given in table 7.1. For comparison, the average water geometry of an isolated molecule is also given.[250] Deviations of water molecules in ice crystal structures from the average water geometry occur because the water molecules distort upon incorporation into the crystal structure to optimise the hydrogen bonding geometries. An *ab initio* energy scan varying the H-O-H angle and using the high quality MP2/AUG-cc-pVTZ basis set, showed that deformations up to $104 \pm 10°$ incurred a maximum 6.7 kJ mol^{-1} energy penalty. The O-H bonds are longer in ice structures than in the average water geometry value because of their participation in hydrogen bonding.

Structure	Intramolecular		Hydrogen bonds	
	O-H lengths	H-O-H angle	O···O length	O-H···O angle
Ice II	0.958	103	2.805	166
	0.972	107	2.767	167
	0.942	-	2.779	178
	1.014	-	2.845	168
Ice VIII	0.968	106	2.879	178
Ice IX	0.977	106	2.75	167
	0.971	105	2.797	175
	0.979	-	2.763	165
Ice XI	0.976	108	2.74	177
	1.054	114	2.803	178
	0.947	-	2.737	176
Average Water Geometry	0.9572	104.52	n/a	n/a

Table 7.1: Summary of water geometries and hydrogen bond values for ice II, VIII, IX and XI. The average water geometry[250] is given for comparison

Three of the ice structures detailed above contain half molecules in the asymmetric unit and required symmetry reduction to lower symmetry settings containing full molecules prior to energy minimisation. The ice XI structure was symmetry lowered to the subgroup $C 1 1 2_1$ (unconventional setting of $P2_1$) by removal of the mirror and glide symmetry operators. The asymmetric unit of this symmetry lowered structure contains two fully occupied water molecules ($Z' = 2$, $Z'' = 2$). The ice IX structure was symmetry lowered to $P2_1$ (with consequent change in setting), with six molecules in the asymmetric unit, four of which are identical and the remaining two are identical to each other ($Z' = 6$, $Z'' = 2$). The ice VIII structure was symmetry lowered to space group $P\bar{1}$

with four identical molecules in the asymmetric unit ($Z' = 4$, $Z'' = 1$). After energy minimisation, the retention of the higher experimental symmetry by the energy minimised structures was confirmed by the PLATON[139] ADDSYM algorithm.

7.2.3 Intermolecular potentials optimised for water

A range of potentials have been developed over many years to model the physical properties of water, dating back to the work of Bernal and Fowler in 1933[261] which used a negative point charge located on the H-O-H angle bisector, positive charges on the hydrogen atoms and dispersion-repulsion terms only on the oxygen. The majority of water potentials developed since this first work are empirically based, with point charges either located on the atomic nuclei, or on slightly off-nuclear positions and the dispersion-repulsion terms commonly represented by a Lennard-Jones 12-6 potential[76] only on the oxygen atom. The potentials that take this form, e.g. BF,[261] SPC,[262] TIP3P,[263] differ in the dispersion-repulsion parameterisation and the magnitude of the partial charges placed on the atomic sites, or at off-nuclear positions. In an attempt to gain further accuracy more complex potentials have been developed with water modelled as flexible rather than rigid (SPC/F), or with empirically based polarisability added to the water molecule (SPCP, PTIP4P, SPC/FQ), or a combination of these variables (SPC/FP).[264] Much more complex force fields with more accurate treatment of the electrostatic, dispersion and repulsion parts of the intermolecular potential and with explicit calculation of polarisability and charge transfer terms based on the results from *ab initio* calculations have recently been developed.[265] Such further advancements beyond the earlier rigid body/point charges/Lennard-Jones model are too complex for consideration here, but clearly indicate the upper limit of the type of models that are required to model water accurately.

It is worth appreciating the physical properties against which the more simple water potentials were parameterised, and therefore aim to reproduce. Guillot[264] has summarised the most common physical properties used, which include the density of liquid under ambient conditions, the heat of vaporisation, the self-diffusion coefficient, the atom-atom pair distribution functions in liquid water, the temperature of maximum density and the critical temperature. Properties of ice are not usually incorporated into the parameterisations of these simple potentials, though an explicit attempt to modify a

water potential for use with ice is available,[266] which reparameterises the TIP4P potential to reproduce the density of several forms of ice.

7.2.4 Ice energy minimisation method

All four of the experimental ice crystal structures were determined by neutron diffraction. Neutron diffraction produces accurate hydrogen (or deuteron) locations, leading to true O-D distances, as opposed to the foreshortened O-H lengths reported by X-ray diffraction. Consequently the water molecular conformations as found in the crystal structures were used in the energy minimisations, with no optimisation of their geometries. Consequently all of the results are ExptMinExpt minimisations.

DMAREL can use dispersion-repulsion potentials based on both Lennard-Jones and Buckingham forms, and a range of water potentials were tested alongside with the FIT potential for their ability to reproduce the four chosen structures of ice. In the FIT potential the H_p polar hydrogen parameters were employed for the water hydrogen atoms. The Lennard-Jones based potentials included SPC,[262] its derivatives SPC/E[267] and MSPC/E,[268] and the two related potentials TIP3P[263] and TIP4P.[263] The NSPC/E[269] potential was also included as it is a Buckingham form potential. All of these potentials have point charges on the three atomic sites (with TIP4P having an off-site charge near the oxygen, rather than on the atomic site) and a dispersion-repulsion potential placed only on the oxygen atom. A derivative of the FIT potential was also included, FIT(COOH), which used the H_p dispersion-repulsion terms derived from the hydrogen in carboxylic acid groups during a CSP study[270] of small organic molecules containing carboxylic acid groups. The repulsion term for the carboxylic acid hydrogen was reduced compared to the standard value for H_p in the FIT potential, achieved by an approximate halving of the pre-exponential factor. Oxygen dispersion-repulsion potential parameters were unaltered from the FIT values. The potentials used are summarised in table 7.2.

The FIT/FIT(COOH) oxygen potential has a much shallower form compared to all of the Lennard-Jones based potentials because it uses explicit dispersion-repulsion terms sited on the hydrogen atoms as well as the oxygen, whereas the Lennard-Jones based potentials incorporate the hydrogen dispersion-repulsion into a single term on the oxygen atom (figure 7.2). The NSPCE potential has a single Buckingham form potential

sited on the oxygen and has a markedly deeper well situated at greater intermolecular separation than the other potentials perhaps because it was optimised for vapour-liquid coexistence properties over a wide temperature range, a quite specific task different from the properties against which the Lennard-Jones potentials were parameterised.

Lennard-Jones							
Oxygen Potential	r(OH) (Å)	<H-O-H (°)	q(O) (e)	q(H) (e)	A (eV mol^{-1})		C (eV mol^{-1})
SPC	1.0	109.47	-0.82	0.41	27293		27.12
SPC/E	1.0	109.47	-0.8476	0.4238	27446		27.20
MSPC/E	0.9839	109.47	-0.8216	0.4108	21568		23.56
TIP4P	0.9572	104.52	-1.04*	0.52	26018		26.45
TIP3P	0.9572	104.52	-0.834	0.417	25237		25.80
Buckingham							
Oxygen Potential	r(OH) (Å)	<H-O-H (°)	q(O) (e)	q(H) (e)	A (eV mol^{-1})	B (Å)	C (eV mol^{-1})
NSPCE	1.0668	109.5	-0.7374	0.3687	2240.97	3.29	64.78
FIT/ FIT(COOH)	-	-	-	-	2384.47	3.96	11.65
Hydrogen Potential					A (eV mol-1)	B (Å)	C (eV mol-1)
FIT					2240.97	3.96	11.65
FIT(COOH)					1072.98	3.96	11.65

Table 7.2: Summary of water potentials used in testing. * Oxygen charge located 0.15 Å off nuclear site
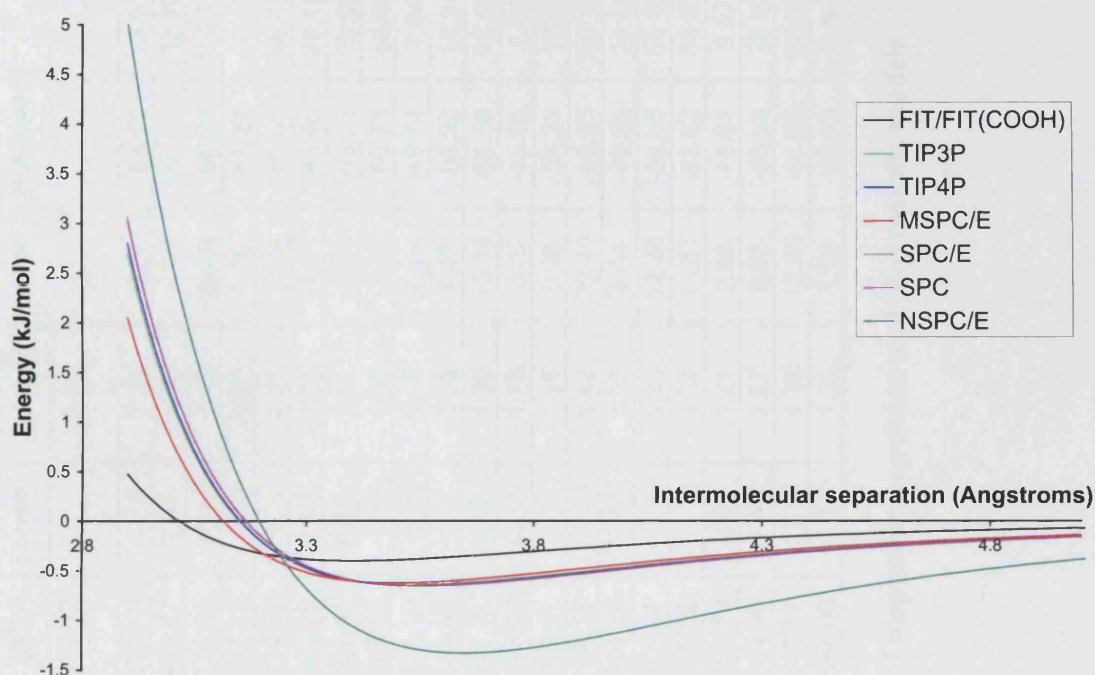
Figure 7.2: O⋯O intermolecular dispersion-repulsion potential for a range of common water potentials. The FIT oxygen homoatomic potential is plotted for comparison

For all potentials minimisations were carried out with the electrostatic intermolecular potential modelled using both MP2/6-31G(d,p)-derived distributed multipoles and electrostatic potential derived charges (CHELPG charges,[271] denoted ESP). For each of the potentials that included nuclear site charges in their definition, an additional energy minimisation was carried out using their published potential charges (only for SPC derivatives and TIP3P, denoted Potential Charges).

7.2.5 Results and discussion

The results of the energy minimisations of all four ice structures with all of the dispersion-repulsion/electrostatic model combinations are summarised in tables 7.3-7.6. For each individual structure wide variations in the quality of reproduction with different potential combinations were observed. For some combinations the success of reproduction varied substantially between the ice polymorphs – for example the TIP4P/multipoles combination reproduced the structure of ice VIII with an F-value of 1.23 whereas it reproduced the structure of ice XI with an F-value of 161 and with a 10% error in the *b* lattice length and an over-estimation of the density by 16%.

Dispersion/ Repulsion	Electrostatic	a (Å)	% error	b (Å)	% error	c (Å)	% error	Vol. (A³)	% error	Density (g/cm³)	% error	Final E (kJ/mol)	F
Experimental		12.983		12.983		6.254		912.82		1.18			
FIT	Multipoles	12.773	-1.61	12.773	-1.61	6.190	-1.02	874.59	-4.18	1.23	4.37	-53.61	13.51
	ESP	13.343	2.78	13.343	2.78	6.443	3.03	993.47	8.84	1.08	-8.12	-40.13	32.76
FIT(COOH)	Multipoles	12.025	-7.38	12.025	-7.38	5.802	-7.22	726.53	-20.41	1.48	25.64	-66.04	182.77
	ESP	12.734	-1.92	12.734	-1.92	6.104	-2.39	857.12	-6.10	1.26	6.50	-47.21	19.45
NSPCE	Multipoles	12.612	-2.85	12.612	-2.85	5.984	-4.32	824.32	-9.70	1.31	10.74	-60.82	46.43
	ESP	13.275	2.25	13.275	2.25	6.246	-0.12	953.25	4.43	1.13	-4.24	-46.46	20.18
	Potential	13.382	3.07	13.382	3.07	6.264	0.16	971.41	6.42	1.11	-6.03	-43.82	29.93
SPC	Multipoles	12.622	-2.78	12.622	-2.78	5.984	-4.31	825.57	-9.56	1.30	10.57	-60.73	45.08
	ESP	13.021	0.29	13.021	0.29	6.207	-0.75	911.32	-0.16	1.18	0.16	-45.74	7.64
	Potential	12.847	-1.05	12.847	-1.05	6.118	-2.17	874.38	-4.21	1.23	4.40	-54.92	14.31
SPCE	Multipoles	12.631	-2.71	12.631	-2.71	5.988	-4.25	827.27	9.37	1.30	10.34	-60.58	43.67
	ESP	13.029	0.36	13.029	0.36	6.211	-0.69	913.02	0.02	1.18	-0.02	-45.65	7.64
	Potential	12.777	-1.59	12.777	-1.59	6.087	-2.67	860.51	-5.73	1.25	6.08	-59.28	19.99
MSPCE	Multipoles	12.261	-5.56	12.261	-5.56	5.823	-6.89	758.13	-16.95	1.42	20.40	-66.95	127.01
	ESP	12.690	-2.25	12.690	-2.25	6.058	-3.13	844.89	-7.44	1.28	8.04	-49.28	28.48
	Potential	12.512	-3.63	12.512	-3.63	5.966	-4.60	808.81	-11.39	1.33	12.86	-59.65	58.46
TIP3P	Multipoles	12.503	-3.69	12.503	-3.69	5.931	-5.15	803.02	-12.03	1.34	13.67	-62.62	66.63
	ESP	12.912	-0.54	12.912	-0.54	6.158	-1.52	889.19	-2.59	1.21	2.66	-46.81	9.97
	Potential	12.698	-2.19	12.698	-2.19	6.051	-3.24	844.96	-7.43	1.27	8.03	-58.59	28.63
TIP4P	Multipoles	12.547	-3.36	12.547	-3.36	5.950	-4.85	811.17	-11.14	1.33	12.53	-61.99	58.23
	ESP	12.951	-0.24	12.951	-0.24	6.175	-1.25	897.00	-1.73	1.20	1.76	-46.49	8.64

Table 7.3: Results of energy minimisations of Ice II using different combinations of dispersion-repulsion and electrostatic models

232

Dispersion Repulsion	Electrostatic	a (Å)	% error	b (Å)	% error	c (Å)	% error	Vol. (Å^3)	% error	Density (g/cm^3)	% error	Final E (kJ/mol)	F
Experimental		4.656		4.656		6.775		146.87		1.63			
FIT	Multipoles	4.583	-1.56	4.583	-1.56	6.635	-2.06	139.39	-5.09	1.72	5.36	-55.53	10.45
	ESP	4.373	-6.07	4.373	-6.07	8.013	18.27	153.26	4.35	1.56	-4.17	-40.65	464.68
FITcooh	Multipoles	4.394	-5.63	4.394	-5.63	6.240	-7.89	120.48	-17.97	1.99	21.90	-66.00	138.66
	ESP	4.255	-8.61	4.255	-8.61	7.571	11.74	137.06	-6.68	1.75	7.16	-46.55	316.66
NSPCE	Multipoles	4.820	3.51	4.820	3.51	7.107	4.90	165.08	12.40	1.45	-11.03	-52.41	54.50
	ESP	4.732	1.62	4.732	1.62	8.161	20.46	182.70	24.40	1.31	-19.61	-40.70	489.25
	Potential	4.751	2.04	4.751	2.04	8.216	21.26	185.43	26.25	1.29	-20.79	-38.60	530.97
SPC	Multipoles	4.669	0.29	4.669	0.29	6.835	0.89	149.03	1.47	1.61	-1.45	-53.05	2.17
	ESP	4.555	-2.18	4.555	-2.18	7.837	15.67	162.57	10.69	1.47	-9.65	-39.03	294.40
	Potential	4.511	-3.12	4.511	-3.12	7.724	14.00	157.16	7.01	1.52	-6.55	-45.67	247.81
SPCE	Multipoles	4.672	0.34	4.672	0.34	6.841	0.97	149.32	1.67	1.60	-1.64	-52.93	2.42
	ESP	4.557	-2.13	4.570	-2.13	7.842	15.75	162.86	10.88	1.47	-9.82	-38.96	296.94
	Potential	4.493	-3.49	4.493	-3.49	7.678	13.33	155.02	5.55	1.54	-5.26	-48.95	231.98
MSPCE	Multipoles	4.559	-2.07	4.559	-2.07	6.613	-2.39	137.48	-6.39	1.74	6.83	-57.87	16.64
	ESP	4.459	-4.22	4.459	-4.22	7.612	12.35	151.37	3.06	1.58	-2.97	-41.64	214.79
	Potential	4.414	-5.20	4.414	-5.20	7.496	10.64	146.02	-0.58	1.64	0.58	-49.12	188.76
TIP3P	Multipoles	4.633	-0.49	4.633	-0.49	6.763	-0.17	145.19	-1.14	1.65	1.16	-54.49	1.61
	ESP	4.524	-2.84	4.524	-2.84	7.763	14.59	158.87	8.17	1.51	-7.55	-39.80	263.63
	Potential	4.469	-4.01	4.469	-4.01	7.624	12.53	152.29	3.69	1.57	-3.56	-48.37	216.12
TIP4P	Multipoles	4.646	-0.21	4.646	-0.21	6.789	0.20	146.54	-0.22	1.63	0.22	-54.06	1.23
	ESP	4.534	-2.62	4.534	-2.62	7.789	14.97	160.12	9.02	1.50	-8.28	-39.62	273.98

Table 7.4: Results of energy minimisations of Ice VIII using different combinations of dispersion-repulsion and electrostatic models

233

Dispersion Repulsion	Electrostatic	a (Å)	% error	b (Å)	% error	c (Å)	% error	Vol. (Å³)	% error	Density (g/cm³)	% error	Final E (kJ/mol)	F
Experimental		6.73		6.83		6.73		309.35		1.16			
FIT	Multipoles	6.639	-1.35	6.701	-1.89	6.639	-1.35	295.40	-4.51	1.22	4.72	-53.27	16.01
	ESP	7.312	8.64	6.112	-10.51	7.312	8.64	326.74	5.62	1.1	-5.32	-39.20	311.63
FITcooh	Multipoles	6.169	-8.33	6.672	-2.32	6.169	-8.33	253.92	-17.92	1.41	21.83	-66.42	170.04
	ESP	6.916	2.76	5.964	-12.68	6.916	2.76	285.27	-7.78	1.26	8.44	-46.19	209.1
NSPCE	Multipoles	6.352	-5.62	6.774	-0.82	6.352	-5.62	273.3	-11.65	1.31	13.19	-63.19	75.34
	ESP	7.058	4.88	6.406	-6.20	7.058	4.88	319.17	3.17	1.12	-3.08	-45.80	100.41
	Potential	7.114	5.70	6.413	-6.10	7.114	5.70	324.52	4.9	1.11	-4.67	-43.28	119.15
SPC	Multipoles	6.485	-3.63	6.985	2.26	6.485	-3.63	293.77	-5.04	1.22	5.3	-61.80	39.27
	ESP	6.999	4.00	6.387	-6.48	6.999	4.00	312.91	1.15	1.15	-1.14	-44.98	86.11
	Potential	6.880	2.22	6.379	-6.60	6.880	2.22	301.91	-2.41	1.19	2.46	-54.47	62.28
SPCE	Multipoles	6.490	-3.56	6.987	2.29	6.490	-3.56	294.32	-4.86	1.22	5.11	-61.64	38.14
	ESP	7.004	4.07	6.39	-6.44	7.004	4.07	313.48	1.33	1.15	-1.32	-44.89	86.9
	Potential	6.836	1.57	6.377	-6.64	6.836	1.57	297.95	-3.68	1.2	3.82	-58.88	57.01
MSPCE	Multipoles	6.277	-6.74	6.919	1.31	6.277	-6.74	272.58	-11.89	1.32	13.49	-68.68	111.43
	ESP	6.802	1.07	6.296	-7.81	6.802	1.07	291.33	-5.83	1.23	6.19	-48.63	73.2
	Potential	6.681	-0.73	6.287	-7.95	6.681	-0.73	280.6	-9.29	1.28	10.25	-59.38	74.11
TIP3P	Multipoles	6.416	-4.67	6.97	2.04	6.416	-4.67	286.9	-7.26	1.25	7.82	-63.90	58.61
	ESP	6.934	3.03	6.360	-6.89	6.934	3.03	305.79	-1.15	1.17	1.16	-46.09	76.49
	Potential	6.790	0.89	6.348	-7.05	6.790	0.89	292.68	-5.39	1.23	5.7	-58.23	59.54
TIP4P	Multipoles	6.442	-4.28	6.969	2.04	6.442	-4.28	289.23	6.51	1.24	6.96	-63.19	50.22
	ESP	6.958	3.39	6.366	-6.79	6.958	3.39	308.2	-0.37	1.16	0.37	-45.75	80.29

Table 7.5: Results of energy minimisations of Ice IX using different combinations of dispersion-repulsion and electrostatic models

Dispersion Repulsion	Electrostatic	a (Å)	% error	b (Å)	% error	c (Å)	% error	Vol. (Å³)	% error	Density (g/cm³)	% error	Final E (kJ/mol)	F
Experimental		4.5019		7.7979		7.328		257.25		0.93			
FIT	Multipoles	4.696	4.32	7.484	-4.02	7.194	-1.82	252.87	-1.70	0.95	1.73	-55.47	43.95
	ESP	4.323	-3.97	8.997	15.37	7.465	1.87	290.32	12.86	0.82	-11.39	-37.88	331.87
FITcooh	Multipoles	4.412	-2.00	6.829	-12.42	6.704	-8.52	201.98	-21.49	1.19	27.37	-71.02	274.77
	ESP	4.202	-6.66	8.416	7.93	7.109	-2.98	251.42	-2.27	0.95	2.32	-44.92	153.14
NSPCE	Multipoles	4.441	-1.34	6.299	-19.23	6.763	-7.71	189.19	-26.46	1.27	35.97	-71.16	515.17
	ESP	4.425	-1.72	8.531	9.40	6.897	-5.88	260.34	1.20	0.92	-1.18	-43.80	205.41
	Potential	4.444	-1.30	8.735	12.02	6.921	-5.56	268.63	4.42	0.89	-4.24	-39.20	279.59
SPC	Multipoles	4.581	1.76	7.036	-9.77	7.009	-4.35	225.92	-12.18	1.06	13.87	-68.46	140.35
	ESP	4.421	-1.80	8.276	6.13	7.153	-2.38	261.71	1.73	0.91	-1.70	-45.02	79.05
	Potential	4.367	-2.99	8.167	4.73	7.097	-3.16	253.12	-1.61	0.95	1.63	-52.29	72.17
SPCE	Multipoles	4.584	1.83	7.042	-9.69	7.014	-4.28	226.43	-11.98	1.06	13.61	-68.27	138.08
	ESP	4.423	-1.76	8.282	6.21	7.158	-2.32	262.20	1.92	0.91	-1.89	-44.92	79.93
	Potential	4.349	-3.41	8.103	3.91	7.062	-3.62	248.84	-3.27	0.96	3.38	-56.71	69.06
MSPCE	Multipoles	4.454	-1.07	6.786	-12.98	6.805	-7.13	205.68	-20.05	1.16	25.07	-77.17	263.06
	ESP	4.333	-3.75	8.002	2.62	6.984	-4.69	242.19	-5.85	0.99	6.22	-49.12	68.88
	Potential	4.278	-4.97	7.893	1.22	6.923	-5.53	233.76	-9.13	1.02	10.05	-57.46	83.88
TIP3P	Multipoles	4.539	0.83	6.953	-10.84	6.942	-5.26	219.10	-14.83	1.09	17.41	-71.14	174.45
	ESP	4.392	-2.44	8.185	4.96	7.098	-3.13	255.18	-0.80	0.94	0.81	-46.28	69.89
	Potential	4.328	-3.87	8.041	3.12	7.021	-4.19	244.32	-5.03	0.98	5.29	-56.15	70.26
TIP4P	Multipoles	4.555	1.18	6.984	-10.44	6.967	-4.92	221.62	-13.85	1.08	16.08	-70.19	161.05
	ESP	4.403	-2.21	8.219	5.40	7.117	-2.87	257.53	0.11	0.93	-0.11	-45.86	72.91

Table 7.6: Results of energy minimisations of Ice XI using different combinations of dispersion-repulsion and electrostatic models

The average F-values summed over all four structures for each potential combination are summarised in table 7.7, ranked by the average F-value.

Dispersion-repulsion potential	Electrostatic potential	Average F-value
FIT	Multipoles	20.98
SPC/E	Multipoles	55.58
SPC	Multipoles	56.72
TIP4P	Multipoles	67.68
TIP3P	Multipoles	75.33
TIP3P	Potential	93.64
SPC/E	Potential	94.51
MSPC/E	ESP	96.34
SPC	Potential	99.14
MSPC/E	Potential	101.30
TIP3P	ESP	105.00
TIP4P	ESP	108.96
SPC	ESP	116.80
SPC/E	ESP	117.85
MSPC/E	Multipoles	129.54
NSPC/E	Multipoles	172.86
FIT(COOH)	ESP	174.59
FIT(COOH)	Multipoles	191.56
NSPC/E	ESP	203.81
NSPC/E	Potential	239.91
FIT	ESP	285.24

Table 7.7: Summary of the average F-value for each dispersion-repulsion/electrostatic combination averaged over all four structures

A given dispersion-repulsion potential in combination with the multipole electrostatic model potential generally proved superior to the corresponding combination with ESP or potential charge electrostatic models, and the five most successful potential combinations used the multipole electrostatic model. Of the five potentials which were tested with their published potential charges three (TIP3P, SPC/E and SPC) gave slightly superior results compared to the ESP derived charges. The NSPC/E and FIT(COOH) dispersion-repulsion potentials proved inadequate for reproducing these ice structures, in combination with both the point charge and multipole electrostatic models. The NSPC/E potential was derived to reproduce vapour-liquid co-existence properties, and so was an inappropriate choice to use for ice. The FIT(COOH) potential only differed from the FIT potential by a reduction in the repulsion at the hydrogen atoms, but this change proved critical. It appears that the H_p parameters[82] derived by fitting to a

range of predominantly N-H containing molecules (FIT) approximates the hydrogen in water better than the H_p parameters derived from carboxylic acids (FIT(COOH)).[270] The most successful potential was the FIT/multipoles combination, which was expected as the potential was more sophisticated than the others tested, containing hydrogen dispersion-repulsion terms as well as those located on oxygen. The FIT potential was also the only potential derived from properties of the crystalline solid state, rather than the liquid state of water. Interestingly the sensitivity of the FIT potential to the combined electrostatic potential was very large, with the FIT/ESP charges proving the least successful potential combination of any, only reproducing the structure of ice II adequately and introducing large errors in the reproductions of the other three polymorphs. The range of final lattice energies varied considerably depending on the dispersion-repulsion/electrostatic model combination, but the FIT/multipoles calculated energies for the four polymorphs fell into a small range, -55.53 to -53.27 kJ mol^{-1}, realistic of the typical relative energy differences between polymorphs, and compares with the sublimation enthalpy of ice Ih at 0 K, calculated to be -47.34(2) kJ mol^{-1}.[272]

This work has proved that within the limitations of the simple potential forms considered here, the FIT dispersion-repulsion potential in combination with the distributed multipoles electrostatic potential can successfully model water-water interactions in the solid state, giving good reproductions of all four of the structures of ice considered. The accuracy with which this combination modelled the four structures of ice suggests that attempts to optimise the FIT H_p dispersion-repulsion potential further specifically for ice would not lead to a significantly improved potential.

7.3 Validation of the intermolecular potential using hydrate crystal structures

7.3.1 Hydrate crystal structures test set and method

The FIT empirical dispersion-repulsion parameters for oxygen were originally derived from a series of oxohydrocarbons and the H_p parameters from molecules containing hydrogen atoms in N-H functional groups. It was not certain that the combining rules would produce heteroatomic parameters to model the interactions between the atoms in water and the atoms in the parent molecule that would be adequate to successfully reproduce hydrate crystal structures. To assess the performance of the cross-terms,

testing against a range of hydrate structures was undertaken. MP2/6-31G(d,p)-derived distributed multipoles were used to model the electrostatic terms, both for the organic molecule and the water molecule in the hydrate structures in all cases.
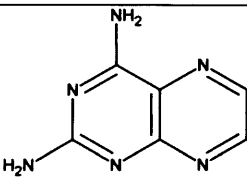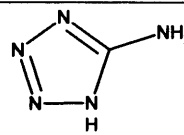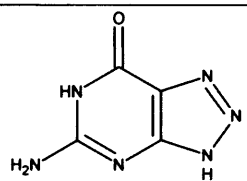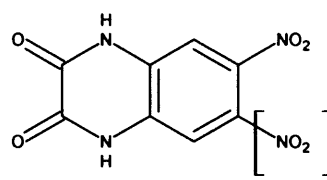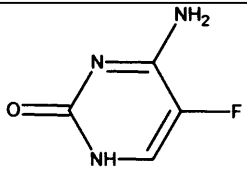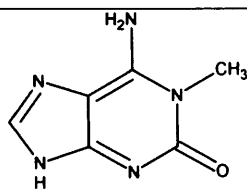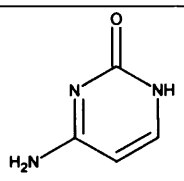
The CSD[23] (May 2004) was searched for suitable hydrate crystal structures. The search was limited to structures containing only the atomic species C, H, N, O, and F. Filters were applied to remove structures containing ions, polymers, disordered structures, structures without three dimensional co-ordinates and structures determined from XRPD. Each structure in the subsequent set was examined manually and those with undetermined water hydrogen positions were discounted along with those in which the parent molecule was too flexible. The most flexible groups included were methyl, nitro and amino substituents.
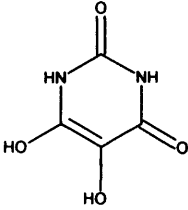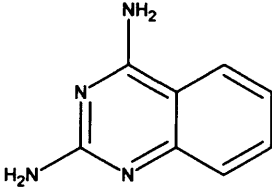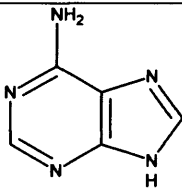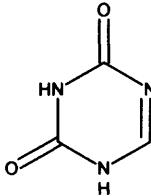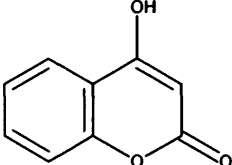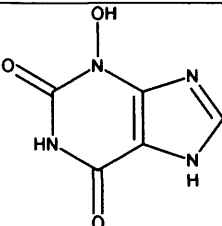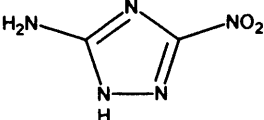
Seven of the monohydrate compounds were found to have one or more corresponding anhydrous crystal structures reported on the CSD. Thymine, cytosine, 3-amino-5-nitro-1,2,4-triazole, 5-azauracil have one reported anhydrous structure, 5-fluorocytosine (chapter 4) and 4-hydroxycoumarin (chapter 5) have two anhydrous structures each and 5-nitrouracil has three. An anhydrous structure was also found for a close structural analogue of 6-nitro-2,3-dihydroxyquinoxaline, with a second nitro group at the 7 position, and was also included in testing. These eight hydrate/anhydrous pairs were specifically identified to allow further analysis of the performance of the FIT dispersion-repulsion potential. The performance of the potential in energy minimising each anhydrous crystal structure can be compared to the results of energy minimising the corresponding hydrated structure to assess the degradation in potential performance with addition of water molecules.
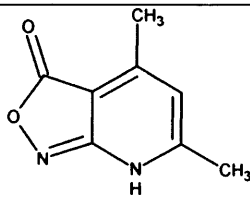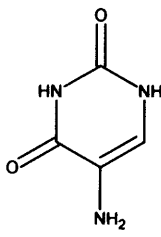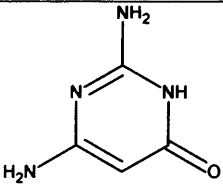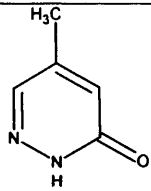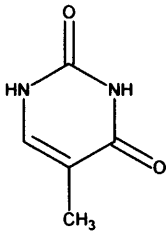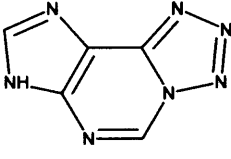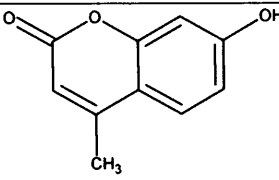
For all structures both ExptMinExpt (with neutron standardised bond lengths to hydrogen) and ExptMinOpt minimisations were carried out. For the ExptMinOpt minimisations the water molecule was optimised at the MP2/6-31G(d,p) level and then O-H bond lengths fixed to 1.0 Å to address the bond lengthening that occurs when water is hydrogen bonded. For several of the structures which contain flexible torsion angles these torsions were fixed to the experimental values during the molecular optimisation, with all other bond lengths and angles allowed to fully optimise. These structures were the 5-nitrouracil monohydrate and the three anhydrous 5-nitrouracil

structures (torsion to the NO_2 group fixed), dialuric acid monohydrate (one hydroxyl group) and anhydrous 5-fluorocytosine form 1 (ring atom positions fixed).

This gave a final set of 22 hydrate crystal structures and eight anhydrous structures that were energy minimised, as summarised in table 7.8.

Compound Name	CSD code	Molecular Structure	F-value*
2,4-Diaminopteridine monohydrate	AMPTRA10		56.39
5-Aminotetrazole monohydrate	AMTETZ01		114.70
8-Azaguanine monohydrate	AZGUAN		295.66
6-Nitro-2,3-dihydroxyquinoxaline monohydrate	BAKGOJ01		56.79
6,7-Dinitro-2,3-dihydroxyquinoxaline	HIHZUT		56.51
5-Fluorocytosine monohydrate	BIRMEU		60.05
anhydrate form 1	MEBQEQ01		6.88[†]
anhydrate form 2	MEBQEQ		10.77
1-Methyl-isoguanine dihydrate	CIMMEQ		1366.3
Cytosine monohydrate	CYTOSM		811.54
anhydrate	CYTSIN		10.16

Dialuric acid monohydrate	DIALAC02		236.36[†]
2,4-Diaminoquinazoline monohydrate	DUPYIW		42.92
Adenine trihydrate	FUSVAQ01		106.11
5-Azauracil monohydrate anhydrate	HOQHAW XERBEB		99.42 71.97
4-Hydroxycoumarin monohydrate anhydrate form 2 anhydrate form 3	HOXCUM01 - -		89.36 178.21 54.76
3-Hydroxyxanthine dihydrate	HXANTH10		83.90
3-Amino-5-nitro-1,2,4-triazole monohydrate anhydrate	 JIYWET JOWWIB		 109.28 51.20

4,6-Dimethyl-isoxazolo-(3,4-b)pyridin-3-one monohydrate	MIOZPO		208.26
5-Nitrouracil monohydrate	NURAMH		$71.88^{\dagger}$
anhydrate form 1	NIMFOE		$50.52^{\dagger}$
anhydrate form 2	NIMFOE01		$98.42^{\dagger}$
anhydrate form 3	NIMFOE02		$20.02^{\dagger}$
2,6-Diamino-4-pyrimidinone monohydrate	SEYDIJ		184.05
5-Methylpyridazin-3-one monohydrate	TEKVIO		43.53
Thymine monohydrate	THYMMH		1632.6
anhydrate	THYMIN01		15.60
Tetrazolo(5,1)-purine monohydrate	TRZPUR		117.81
7-Hydroxy-4-methyl-chromen-2-one monohydrate	WIKDAV		35.50
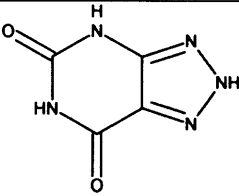
Xanthazole monohydrate	XANAZH01		94.96

Table 7.8: Hydrate structures used for potential validation. Compound name, CSD reference code and molecular structure are given. * ExptMinOpt F-value, †ExptMinConOpt

7.3.2 Results and discussion

The final column in table 7.8 includes the corresponding ExptMinOpt F-values for all structures in the test set. Table 7.9 summarises the cell parameter variations and F-values for all 22 hydrate minimisations ranked by F-value. Table 7.10 summarises the most changed hydrogen bond donor acceptor distance between experimental and ExptMinOpt structures to assess the reproduction of the hydrogen bonding. To identify the hydrogen bond patterns, the hydrogen bonds for all experimental and minimised structures were identified using the standard PLUTO[141] hydrogen-bond criteria: $O\cdots O$ MAXD (maximum distance) ≤ 2.52 Å, $O\cdots N$ MAXD ≤ 2.55 Å, $N\cdots O$ MAXD ≤ 2.52 Å, $N\cdots N$ MAXD ≤ 2.55 Å with hydrogen bonds slightly above this range manually identified only in minimised structures if a corresponding experimental hydrogen bond occurred. Summary tables detailing both the ExptMinExpt and ExptMinOpt reproductions of all 22 hydrate structures and the hydrate/anhydrate pairs are given on the supporting information CD (Chapter_7_5Azauracil_Monohydrate).

REFCODE	F	a (Å)	% error	b (Å)	% error	c (Å)	% error	α (°)	% error	β (°)	% error	γ (°)	% error	Volume (Å³)	% error
WIKDAV	35.5	6.932	-2.45	11.536	1.77	12.001	1.56	90	-	104.95	-0.34	90	-	927.15	0.99
DUPYIW	42.92	21.219	-1.7	21.219	-1.7	7.738	1.53	90	-	90	-	90	-	3484.27	-1.9
TEKVIO	43.53	6.705	-2.86	6.404	1.61	14.9	3.43	90	-	93.87	1.9	90	-	638.31	1.92
AMPTRA10	56.39	6.778	0.25	7.373	0.78	8.459	-0.97	96.82	-2.01	101.21	2.46	112.66	2.64	373.81	-2.49
BAKGOJ01	56.79	7.695	2.95	11.724	-1.7	10.831	3.24	90	-	96.99	-0.53	90	-	969.81	4.6
BIRMEU	60.05	7.511	1.68	9.63	2.52	17.89	1.76	90	-	99.83	1.23	90	-	1274.93	5.70
HXANTH10	83.9	7.717	-1.54	7.963	-5.26	7.631	2.56	111.7	-2.21	106.04	-2.54	67.4	2.36	397.32	-0.88
HOXCUM01	89.36	6.870	-0.75	9.628	-4.15	12.902	6.24	90	-	90	-	90	-	853.29	1.06
XANAZH01	94.96	9.635	-0.36	10.733	0.22	5.417	3.38	98.05	-2.79	145.08	2.24	87.89	1.00	314.51	-3.13
NURAMH	71.88	5.317	3.5	22.639	3.11	9.595	0.08	90		146.29	1.95	90		640.84	-0.36
HOQHAW	99.42	6.584	2.37	5.775	-7.01	7.065	3.59	90	-	101.8	2.08	90	-	262.98	-2.07
FUSVAQ01	106.11	6.841	4.97	7.675	-3.23	8.515	-1.71	94.44	-2.42	98.27	6.28	99.65	0.30	433.8	-0.92
JIYWET	109.28	10.397	-6.58	14.942	-1.87	7.389	3.64	90	-	98.35	-2.90	90	-	1135.76	-4.15
AMTETZ01	114.7	6.301	-1.42	6.891	-5.3	9.722	-0.83	90	-	86.55	-4.09	90	-	421.37	-7.59
TRZPUR	117.81	11.688	-0.35	17.13	0.7	3.8	0.53	90	-	90	-	90	-	760.83	0.88
SEYDIJ	184.05	16.321	-2.39	3.898	-8.11	19.082	4.31	90	-	90	-	90	-	1213.92	-6.44
MIOZPO	208.26	8.077	-4.19	9.321	7.51	6.943	2.1	71.44	-8.32	76.1	-1.25	62.51	-3.12	436.85	0.74
DIALAC02	236.36	13.205	3.86	3.967	7.92	12.262	-5.31	90	-	91.54	-3.03	90	-	642.13	6.42
AZGUAN	295.66	3.671	2.83	11.2	-1.84	16.196	-2.02	90	-	95.33	0.03	90	-	663.03	-1.11
CYTOSM	811.54	8.832	13.22	9.317	-5.36	7.260	-5.51	90	-	86.58	-13.16	90	-	596.34	2.54
CIMMEQ	1366.27	4.513	9.82	8.755	-7.39	13.254	7.73	62.97	-15.01	69.75	-17.47	78.53	-1.79	437.16	-3.31
THYMMH	1632.6	4.956	-18.45	26.282	-5.67	4.941	29.48	90	-	89.93	-4.65	90	-	643.55	-0.11

Table 7.9: Summary of the energy minimisations of all 22 hydrate structures

REFCODE	Most poorly reproduced hydrogen bond		
	Experimental (Å)	Minimised (Å)	% error
WIKDAV	2.790	2.875	3.05
DUPYIW	3.051	2.890	-5.28
TEKVIO	2.759	2.877	4.28
AMPTRA10	3.066	2.926	-4.57
BAKGOJ01	2.768	2.863	3.43
BIRMEU	2.755	2.994	8.68
HXANTH10	2.817	2.948	4.65
HOXCUM01	2.783	2.93	5.28
XANAZH01	2.883	2.814	-2.39
NURAMH	2.924	2.845	-2.70
HOQHAW	2.733	2.819	3.15
FUSVAQ01	2.993	2.843	-5.01
JIYWET	No longer within hydrogen bond criteria		
AMTETZ01	2.978	2.917	-2.05
TRZPUR	3.017	2.857	-5.30
SEYDIJ	2.789	3.081	10.47
MIOZPO	2.722	2.852	4.78
DIALAC02	No longer within hydrogen bond criteria		
AZGUAN	No longer within hydrogen bond criteria		
CYTOSM	No longer within hydrogen bond criteria		
CIMMEQ	No longer within hydrogen bond criteria		
THYMMH	2.551	2.864	12.27

Table 7.10: Most poorly reproduced hydrogen bonds between experimental and minimised structures. In five structures the hydrogen bonding pattern had been substantially altered by the energy minimisation procedure

Energy minimised structures that had a low F-value and retained the hydrogen bonding present in the corresponding experimental structure with minimal discrepancy in the most poorly reproduced hydrogen bond length were classified as successful. This was the case for all but one (JIYWET) of the 15/22 ExptMinOpt minimisations with an F-value of less than 120. Of the four ExptMinOpt structures within the F-value range 180-300 range, two retained the corresponding experimental hydrogen bond set (MIOZPO and SEYDIJ) and in the other two the hydrogen bonding was altered (DIALAC02 and AZGUAN). Three ExptMinOpt structures had F-values greater than 300, with two having altered hydrogen bonding patterns (CYTOSM and CIMMEQ). All but one of the structures with F greater than 180 had at least one cell parameter that had altered by more than 7.5%. In only five of the 22 minimisations was the hydrogen bonding altered in the minimised structure: cytosine monohydrate (CYTOSM), thymine monohydrate (THYMMH), 8-azaguanine monohydrate (AZGUAN), 1-methyl-isoguanine dihydrate (CIMMEQ) and dialuric acid (DIALAC02), while the remaining 17 structures

successfully retained the original hydrogen bonding pattern on minimisation. From these results it can be concluded that an upper F-value limit of 120 can be taken to be indicative of a successful energy minimisation of a monohydrate crystal structure.

For the eight hydrate/anhydrous pairs studied, in four cases the hydrate structure was reproduced with approximately equivalent accuracy to that of the corresponding anhydrate structures (for hydrates NURAMH, HOQHAW and BAKGOJ01, HOXCUM01). For BIRMEU the hydrate structure was adequately reproduced, but appreciably less accurately than the corresponding anhydrate crystal structure. For the remaining three pairs, JIYWET, CYTOSM and THYMMH, energy minimisation of the anhydrate structures gave very good reproductions, while the hydrate structures were significantly altered, with significant changes in both unit cell parameters and molecule placement in the unit cell. Figure 7.3 shows the least successful hydrate reproduction, overlaying the experimental thymine monohydrate structure with the ExptMinOpt structure, matched on the thymine molecules. It is clear that the thymine hydrogen bonded chains remain intact, but the positions of the water molecules are very different and are the main contributory factor.

From the successful reproduction of the majority of hydrates structures tested it can be concluded that the FIT potential is adequate for use in a CSP search for monohydrates of a small organic molecule.
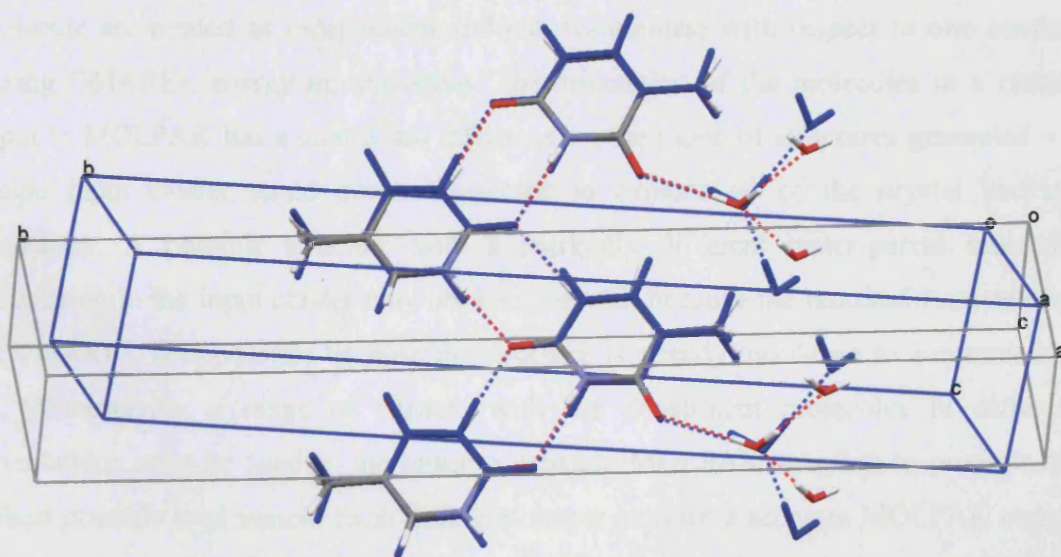
Figure 7.3: Overlay of crystal structure of thymine monohydrate (coloured by element) with that of the ExptMinOpt energy minimised structure (coloured blue)

7.4 Crystal structure prediction of 5-azauracil monohydrate

7.4.1 Modification of the CSP method

5-Azauracil monohydrate was chosen from the 22 hydrates used in the hydrate potential testing to be used as the monohydrate CSP test subject. It was reasonably well reproduced by the FIT potential, with an F-value of 99. The known structure was found to crystallise in the space group $P2_1/m$ with both the 5-azauracil and water molecules lying on the mirror plane, giving $Z' = 0.5$. For the energy minimisation process, the space group was reduced to $P2_1$ with $Z' = 1$. MOLPAK does not generate structures in the space group $P2_1/m$ but it was expected that this structure would be generated in the sub-groups of $P2_1/m$, such as $P2_1$ or possibly $P\bar{1}$.

When generating hypothetical crystal structures for a single molecule MOLPAK requires a rigid conformation of the molecule to be used as the basic unit which it packs into different packing types with different orientations to generate crystal structures. For monohydrates the MOLPAK search required modification to include a water molecule along with the parent molecule to produce a cluster which MOLPAK can then use to generate possible crystal structures. The orientation of the two molecules in this cluster remains fixed during the MOLPAK structure generation procedure, but once the most densely packed structures are passed to DMAREL the parent molecule and water molecule are treated as independent and can re-orientate with respect to one another during DMAREL energy minimisation. The orientation of the molecules in a cluster input to MOLPAK has a significant influence on the range of structures generated – a single input cluster could not be expected to explore all of the crystal packing landscape. A possible structure with a markedly different water-parent molecule orientation to the input cluster may not be generated because the required reorientation in DMAREL is impossible because the structure is already too dense to accommodate it. Consequently a range of clusters with the constituent molecules in different orientations must be used as the input to separate MOLPAK searches to produce the fullest possible total search. Each different cluster requires a separate MOLPAK search with the sum of all of these individual searches producing the total search, making the crystal structure prediction of monohydrates a computationally expensive undertaking.

For 5-azauracil (scheme 7.1) clusters were not be generated in a random way, but by exploiting crystallographic knowledge of the hydrogen bond directionality and intermolecular separation that could be expected in hydrate crystal structures. In different clusters water molecules were placed at sites to hydrogen bond to one of the three 5-azauracil hydrogen bond acceptors. The three acceptor groups were denoted A1, A2 and A3 (scheme 7.1). The CSD was searched to provide common values for some the orientational variables between the water and 5-azauracil molecules, to limit the total number of clusters required. Clusters were not considered in which water hydrogen bonded to the two amino hydrogen bond donors on 5-azauracil because it was expected that this cluster space would be automatically sampled by the generation of full crystal structures from the acceptor clusters and for practical considerations to limit the number of individual MOLPAK searches required. It should be noted that the methodology described here to define the set of clusters is specific to 5-azauracil, though the general method is applicable to any other molecule.

A1

O2

H3   C1   H4

N1   N2

C3   C2

H5   N3   O3

A3   **A2**
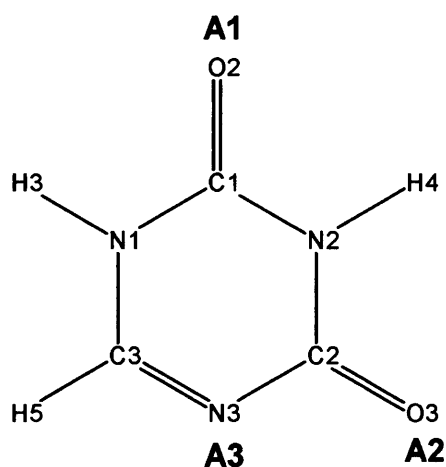
Scheme 7.1: Definition of the numbering scheme for 5-azauracil used for CSP with the acceptor labelling also shown

7.4.2 CSD analysis

Water molecules in predicted monohydrate crystal structures of 5-azauracil were expected to be hydrogen bonded and the directionality of hydrogen bonds, combined with typical hydrogen bond length and angle statistics derived from the CSD, were exploited to define common values for some orientational parameters between water and 5-azauracil to limit the total number of clusters. The orientation between water and

another molecule across a hydrogen bond can be defined completely in terms of one distance (b1···a3), two angles (a2-a3···b1, a3···b1-b2) and three torsions (a1-a2-a3···b1, a2-a3···b1-b2, a3···b1-b2-b3), as shown in scheme 7.2,[273] with the water atoms H-O-H defined as a3-a2-a1.
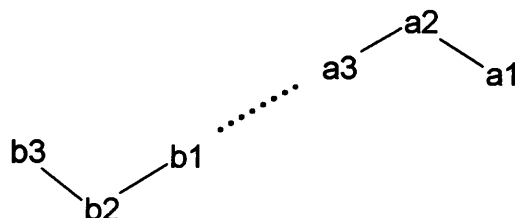
Scheme 7.2: Definition of the orientation of two fragments. Distance b1···a3, angles a2-a3···b1, a3···b1-b2 and torsions a1-a2-a3···b1, a2-a3···b1-b2 and a3···b1-b2-b3

The CSD[136] (May 2005) was searched for close contacts between water and the two types of hydrogen bond acceptor present in 5-azauracil (figure 7.4). Statistics were determined for the distance, angles and torsions defined above, between the water molecule and hydrogen bond acceptor fragments. The CSD searches were limited to returning fully determined structures with R factors < 0.075, no disorder, no errors, no polymers, no ions, no XRPD structures, and only organic molecules. All bonds to hydrogen were normalised to neutron values during the CSD search. For both A and B (figure 7.4) the results from the searches were sorted into 0.1 Å bins for the distances and 10° bins for the angles and torsion angles (tables 7.11 and 7.12).
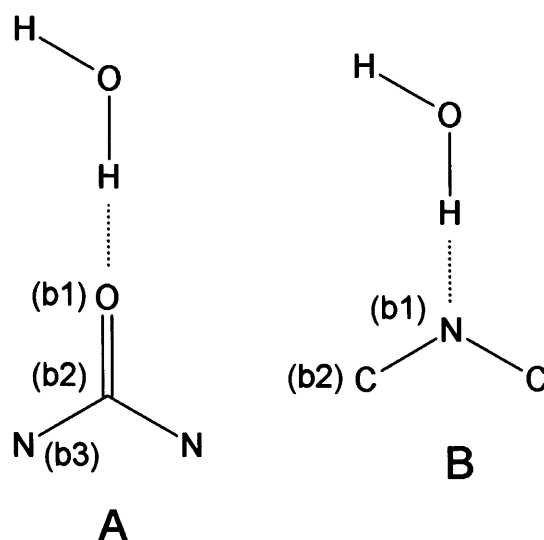
Figure 7.4: Interactions used to search the CSD. A: hydrogen bond to carbonyl group; B: hydrogen bond to aza group

For A (figure 7.4A) 205 matches were returned. The distance $H_w\cdots O$ (a3$\cdots$b1) showed a maximum frequency in the range 1.8-1.9 Å (63/205 matches) with the range 1.9-2.0 Å almost as populated (59). Both of the angles showed single maxima, with the angle O_w-$H_w\cdots O$ (a2-a3$\cdots$b1) range 160-170° having maximum frequency (74). For the angle H_w-O=C (a3$\cdots$b1-b2) the range 130-140° proved most common (61). None of the three torsion angles showed any trend.
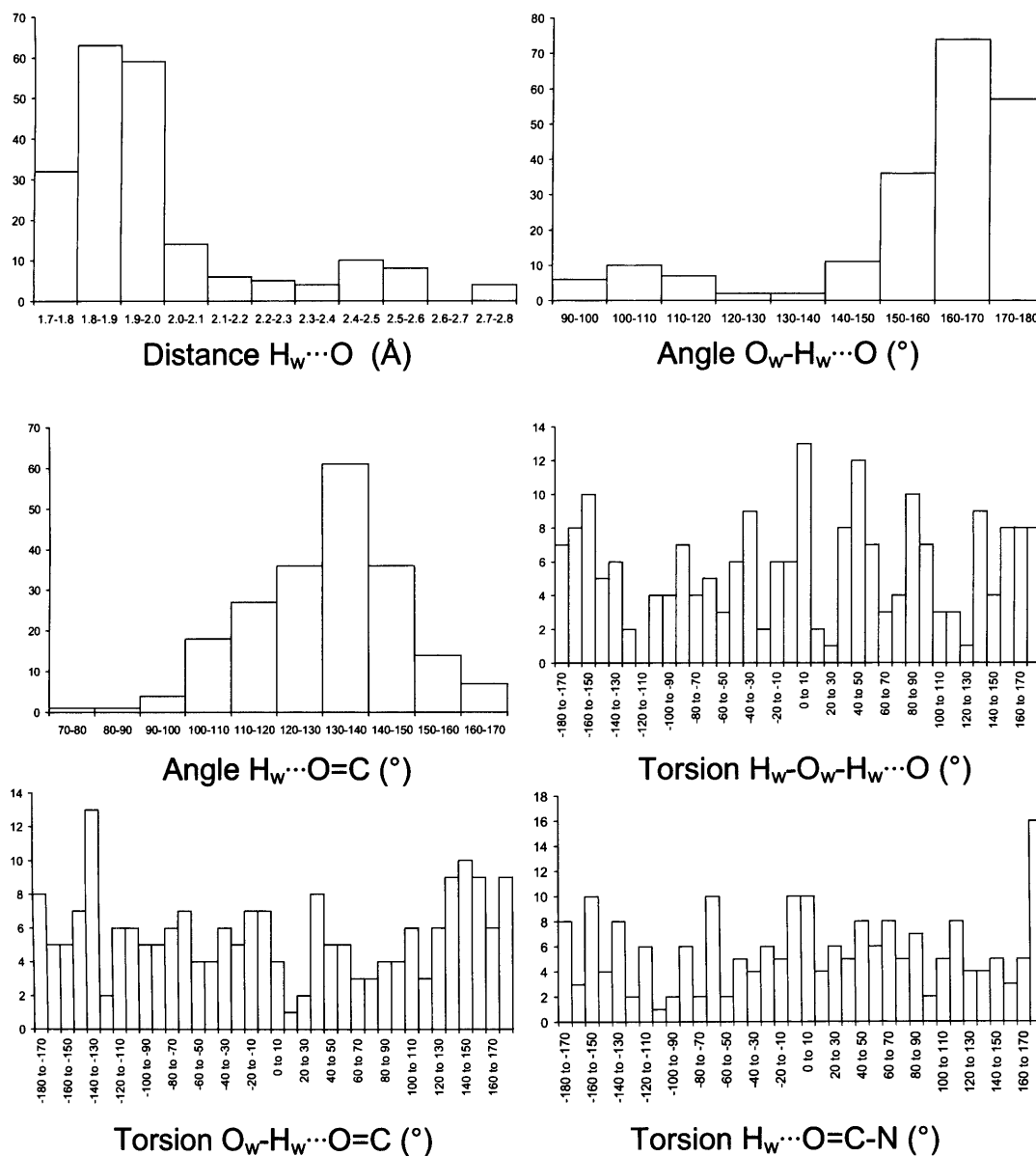
Table 7.11: Histograms of CSD search results for the interaction shown in figure 7.4A. Vertical scale in all cases is frequency. Subscript w refers to water atoms

For B (figure 7.4B) 253 matches were returned (table 7.12). It should be noted that three sequentially bonded atoms are required in each fragment for definition of all six parameters defined according to scheme 7.2. Initially the CSD was searched for water interacting with the fragment N(-C)-C(=O)-[any atom], the aza group with adjacent carbonyl as found in 5-azauracil, but only 5 matches were found. The more generic N(-C)-C fragment as shown in figure 7.4B gave a larger number of matches for five of the six parameters, but without information on the torsion a3···b1-b2-b3. Again single maxima were found for the H_w···N distance (range 1.9-2.0 Å = 95 matches) and the two angles, O_w-H_w···N (range 170-180° = 93 matches) and H_w···N-C (range 110-120° = 85 matches). Neither of the torsions showed any trend.

From the experimental crystal structure[274] of 5-azauracil monohydrate, the parameters for the hydrogen bonds formed between water molecules and the acceptors A1 and A3 were measured. The hydrogen bond distances and angles were found to lie in the common ranges identified by the CSD analyses above. The hydrogen bond to A1 had the following parameters: distance O···H_w = 1.90 Å, angle C=O···H_w = 162°, angle O···H_w-O_w = 124°; and for A3: distance N···H_w = 1.87 Å, angle N···H_w-O_w = 178°, and angle C-N···H_w = 110°.

It is appreciated that a conic correction factor should have been applied to the results for the hydrogen bond angles in both CSD analyses[275] but this was not done. This would have increased the relative population of the 170-180° bins for the hydrogen bond angles O_w-H_w···O and O_w-H_w···N in the CSD searches.

100
90
80
70
60
50
40
30
20
10
0

1.7-1.8 1.8-1.9 1.9-2.0 2.0-2.1 2.1-2.2 2.2-2.3 2.3-2.4 2.4-2.5 2.5-2.6 2.6-2.7 2.7-2.8

Distance $H_w\cdots N$ (Å)

100
90
80
70
60
50
40
30
20
10
0

90-100 100-110 110-120 120-130 130-140 140-150 150-160 160-170 170-180

Angle $O_w-H_w\cdots N$ (°)

90
80
70
60
50
40
30
20
10
0

60-70 70-80 80-90 90-100 100-110 110-120 120-130 130-140 140-150 150-160 160-170

Angle $H_w\cdots N-C$ (°)

16
14
12
10
8
6
4
2
0

-180 to -170 ... 160 to 170

Torsion $H_w-O_w-H_w\cdots N$ (°)

18
16
14
12
10
8
6
4
2
0

-180 to -170 ... 160 to 170

Torsion $O_w-H_w\cdots N-C$ (°)

Table 7.12: Histograms of CSD search results for the interaction shown in figure 7.4B. Vertical scale in all cases is frequency. Subscript w refers to water atoms

7.4.3 Definition of clusters used in the search

For clusters around the acceptors A1 and A2, the a3$\cdots$b1 distance was fixed to 1.9 Å, because from the CSD analysis the bins 1.8-1.9 Å and 1.9-2.0 Å contained the majority of structures and 1.9 Å was the mid-value of this range. Similarly the angle a3$\cdots$b1-b2 was fixed to 135° because the bin with greatest frequency was 130-140°. The angle a2-a3$\cdots$b1 had greatest frequency in the range 160-170° but was fixed at 180° as this

251

produced a linear hydrogen bond and also fixed the torsion a2-a3···b1-b2 to be zero. This left only two torsions that required systematic variation. With four of the six variables now fixed, the torsion a3···b1-b2-b3 now defined the position of the water molecule with respect to the plane of the 5-azauracil molecule and the torsion a1-a2-a3···b1 defined the orientation of the non-hydrogen bonded hydrogen. Due to both 5-azauracil and water being non-chiral, only 0-180° of the a3···b1-b2-b3 torsion was included, with the other hemisphere duplicated by the mirror plane through the plane of the 5-azauracil molecule. Consequently clusters were generated with this torsion set at 0°, 60°, 120° and 180°. At each of these water positions, the torsion a1-a2-a3···b1 was varied in 60° steps between 0-360°. This led to 24 clusters at each of the acceptors A1 and A2. Figures 7.5 shows an overlay of all of the water positions in the clusters at A1 and A2 – for each there are four water sites, and at each site there are six orientations of the free hydrogen of the water molecule.
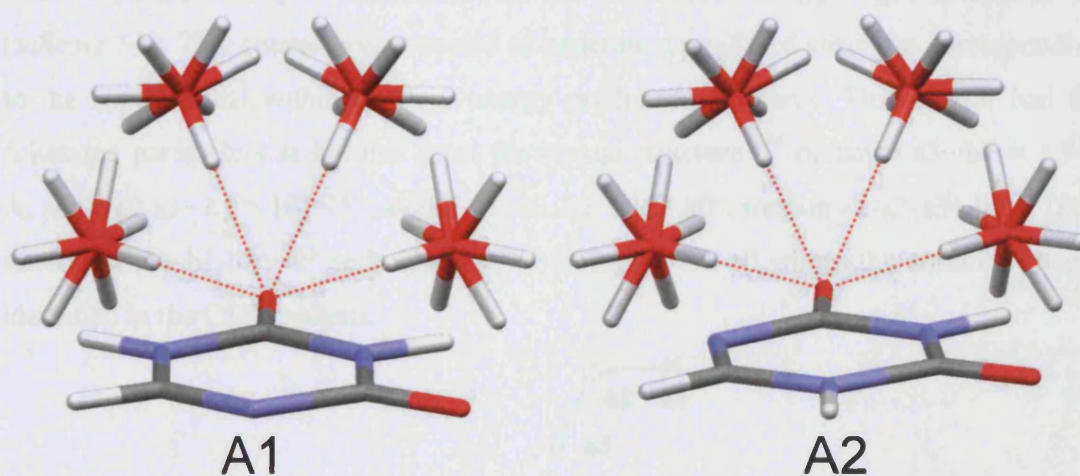
Figure 7.5: Overlay of all clusters generated around acceptors A1 and A2

A similar process was used to define the clusters around A3, but in this case the choice of b2 and b3 determined the placement of the water sites. The atoms C2 and N2 were chosen as b2 and b3 as this led to clusters occupying space closer to C3-H5 rather than C2=O3 around which clusters had already been generated. From the CSD analysis and defining the hydrogen bond as linear led to the fixed variables: a3···b1 = 1.9 Å, angle a2-a3···b1 = 180°, angle a3···b1-b2 = 120° and torsion a1-a2-a3···b1 = 0°. Again the torsion a3···b1-b2-b3 was varied in 60° steps between 0-180° and the torsion a1-a2-a3···b1 was varied in 60° steps between 0-360°. Setting torsion a3···b1-b2-b3 = 0° lead

to overlap of the water molecule with C3-H5 and so this water position was not used. This gave a total of 18 clusters at acceptor site A3 (figure 7.7).
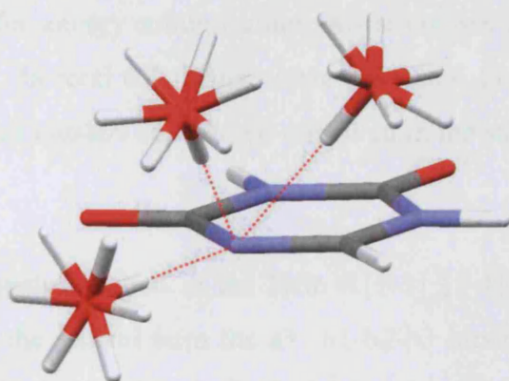
Figure 7.6: Overlay of all clusters generated around acceptor A3

For the purposes of comparison a cluster was considered with the parameters as found in the experimental crystal structure, with the water molecule hydrogen bonded to O2 (scheme 7.3). This cluster was expected to generate a predicted structure corresponding to the experimental within the low energy predicted structures. This cluster had the following parameters as defined from the crystal structure:[274] distance a3···b1 = 1.947 Å, angle a2-a3···b1 = 161.75°, angle a3···b1-b2 = 123.80°, torsion a1-a2-a3···b1 = 180°, torsion a2-a3···b1-b2 = 0° and torsion a3···b1-b2-b3 = 0°, all within the common ranges identified in the CSD analysis.
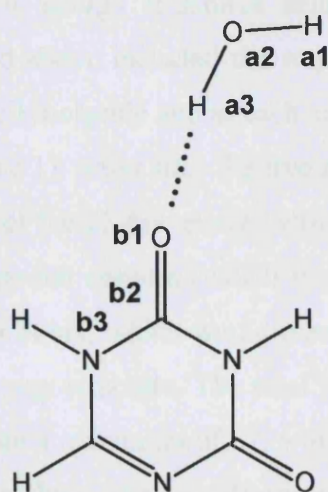
Scheme 7.3: Experimental cluster

The MP2/6-31G(d,p) *ab initio* optimised conformation of 5-azauracil and water were used, with the water O-H lengths 0.96 Å and a H-O-H angle of 104°. For each search

37 MOLPAK packing types covering 18 space groups were used. 5000 structures were generated in each packing type of which the top 125 from each packing type were passed to DMAREL for energy minimisation. All structures that minimised to saddle points were discarded. In total 67 clusters were subjected to an individual MOLPAK search. Definitions of all clusters considered are given in the supporting information.

7.4.4 Results

Each cluster was uniquely labelled in the form A[1-3]_[1-4]_[a-f] with the first term denoting the acceptor, the second term the a3···b1-b2-b3 torsion number, where 0° = 1 and 180° = 4 and the third term the a1-a2-a3···b1 torsion with the six positions denoted from a to f. The low energy structures from all 67 individual searches were collated and the lowest energy example of each unique structure was retained. 24 structures were found lower in energy than the ExptMinOpt structure. A predicted structure was found in 51 of the searches that corresponded exactly to the ExptMinOpt structure and this structure was included in the low energy structure list, leading to a final set of 25 low energy structures, within an energy window of 4.3 kJ mol^{-1} (table 7.13).

The number of individual searches in which each structure is found gives an indication of the completeness of the total search – if all structures were found repeatedly the search procedure could be considered close to complete, with only a small possibility of new low energy structures being found by considering further clusters. The clusters defined above included the water molecule placed at 11 distinct sites relative to the 5-azauracil molecule and at each site six different water orientations were sampled. For each of the 11 water sites the frequency with which the six searches at each site discovered each of the 25 low energy structures have been collated in table 7.14. Except for the experimental column (which includes only one search), all other entries can have a maximum of six, which would mean that at this site all six searches found that particular low energy structure. The total at the end of each row shows the total number of searches, from a maximum of 67, which found each structure. The total at the bottom of each column shows the completeness of the searches at each of the 11 water placement sites, with a maximum of 25 meaning that the six searches at this site found all 25 of the low energy structures.

Structure Rank	Lowest energy example *	Final Energy (kJ mol^{-1})	Space Group	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	Volume (Å^3)	Hydrogen Bond Motif
1	A1_3_c ad/32	-165.01	Pc	5.064	7.084	7.275	90	90.61	90	261	Sheet 2
2	A2_3_b ad/1	-164.207	Pc	5.499	6.972	6.844	90	93.71	90	262	Sheet 2
3	A1_2_d da/92	-164.065	Cc	7.337	7.039	10.04	90	90.31	90	518	Sheet 2
4	A3_1_c ad/9	-162.474	Pc	3.683	5.706	12.646	90	100.22	90	262	Not Sheet
5	A3_3_b aa/93	-162.419	P1	3.701	5.762	6.172	102.43	94.7	91.35	128	Sheet 1
6	A2_2_e da/110	-162.344	Cc	11.934	7.403	10.183	90	144.497	90	523	Sheet 1
7	A3_1_f ad/17	-162.25	Pc	5.04	7.413	7.007	90	93.38	90	261	Sheet 1
8	A3_1_d aa/93	-162.209	P1	4.895	4.995	5.361	92.44	92.01	90.78	131	Sheet 1
9	A1_2_b ah/65	-162.112	P2₁	3.625	12.149	5.802	90	91.11	90	255	Not Sheet
10	A3_1_f ad/31	-162.039	P2₁	4.716	5.261	10.623	90	96.65	90	262	Not Sheet
11	A3_1_c da/66	-161.882	Cc	6.998	7.406	10.113	90	93.09	90	523	Sheet 1
12	A2_4_b da/75	-161.857	Cc	12.858	6.906	6.789	90	119.13	90	527	Sheet 1
13	A2_2_d da/97	-161.371	Cc	7.186	7.012	10.349	90	90.88	90	521	Sheet 2
14	A2_2_b da/84	-161.318	Cc	12.66	6.796	7.194	90	123.67	90	515	Sheet 1
15	A2_2_b ab/123	-161.056	P-1	6.92	6.871	6.911	104.06	109.32	113.49	256	Sheet 1
16	A2_4_d ad/1	-161.056	Pc	3.811	7.475	9.066	90	91.11	90	258	Sheet 1
17	A2_3_e de/8	-160.944	C2/c	11.535	7.57	15.747	90	131.81	90	1024	Sheet 1
18	A1_3_a de/44	-160.939	C2/c	11.535	7.57	11.782	90	95.044	90	1025	Sheet 1
19	A2_4_b ai/7	-160.834	P2₁/c	7.227	6.946	10.125	90	92.25	90	508	Not Sheet
20	A1_3_d dc/13	-160.831	C2/c	11.598	7.55	11.786	90	93.94	90	1030	Sheet 1
21	A2_4_a fa/78	-160.828	P2₁/c	6.705	11.706	9.281	90	130.7	90	552	Not Sheet
22	A3_1_e da/57	-160.774	Cc	11.676	7.482	17.437	90	160.208	90	516	Sheet 1
23	A3_1_a da/28	-160.767	Cc	12.014	7.482	7.566	90	130.68	90	516	Sheet 1
24	A2_2_c am/86	-160.72	P2₁/c	6.935	12.468	6.966	90	120.13	90	521	Sheet 2
25	A1_1_a ab/19 [†]	-160.696	P2₁/m	6.584	5.775	7.065	90	101.8	90	263	Sheet Expt

Table 7.13: Summary of the 25 lowest energy 5-azauracil monohydrate predicted structures. * This structure is the lowest energy example of each unique structure; [†] corresponds to the ExptMinOpt structure

Structure Rank	Exptl	A1_1	A1_2	A1_3	A1_4	A2_1	A2_2	A2_3	A2_4	A3_1	A3_2	A3_3	Total Occurrence
1	1	4	6	6	5	3	0	1	0	2	6	5	39
2	0	1	1	2	0	3	6	6	6	5	1	0	31
3	0	0	4	5	2	1	3	4	0	4	2	4	29
4	0	0	0	0	0	0	0	0	0	1	0	0	1
5	1	6	6	6	6	6	6	5	6	6	6	6	66
6	0	0	4	3	4	4	6	6	6	6	5	4	48
7	0	0	2	5	1	1	1	1	0	6	5	3	25
8	1	6	6	6	6	6	6	6	6	6	6	6	67
9	0	6	6	6	4	6	6	6	6	6	6	6	64
10	0	0	0	0	0	0	0	0	0	1	0	0	1
11	0	0	1	3	2	0	1	2	0	6	0	2	17
12	0	1	4	4	5	6	5	5	6	6	5	4	51
13	0	0	1	4	0	3	5	5	4	2	2	3	29
14	0	5	2	2	4	3	6	4	6	4	5	0	41
15	1	4	6	6	6	6	6	6	6	5	6	6	64
16	0	0	1	0	0	2	0	2	3	2	0	1	11
17	1	3	2	3	3	3	2	1	1	2	1	0	22
18	0	1	4	2	2	0	4	1	1	3	5	4	27
19	0	2	1	0	2	0	0	2	4	2	2	0	15
20	0	2	2	3	5	3	4	4	3	5	5	4	40
21	0	4	6	6	0	1	1	1	2	5	5	3	34
22	0	0	1	0	0	2	3	1	1	1	2	1	12
23	0	0	2	2	1	1	2	0	0	1	1	2	12
24	0	0	1	0	0	0	1	0	0	2	1	0	5
25	1	6	6	6	2	6	6	5	1	3	6	3	51
Completness	6/25	14/25	23/25	19/25	17/25	19/25	20/25	21/25	17/25	25/25	21/25	18/25	

Table 7.14: 5-azauracil monohydrate CSP summary. Last column - frequency with which each structure was found in the total search; last row - completeness for each set of six searches at a single water site

7.4.5 Discussion

The set of six searches at 5/11 water sites found 20 or more of the low energy structures but only the A3_1 set of searches found all 25. It was not be expected that a single water site would be capable of finding all of the low energy structures, even with the use of six different water orientations at each site, and these results confirm the decision to use a range of water sites as well as water orientations. The eighth ranked structure was found by all 67 individual searches, and four predicted structures were found more than 60 times. 11 of the 25 low energy predicted crystal structures were found by more than half of the individual searches, including the structure that corresponds exactly to the ExptMinOpt structure (the 25th ranked structure in table 7.14). Of concern was that two structures were found only once, the fourth and tenth ranked structures, both by the same A3_1 set of searches. Both structures were structurally reasonable and not artefacts of this particular set of searches. The conclusion that must be drawn is that the total search is not close to completeness and other hypothetical structures could be found that are energetically competitive with the low energy structures reported above.

All of the structures were examined to identify recurring hydrogen bond motifs, to assess the qualitative similarities between the crystallographically distinct structures. The ExptMinOpt structure was examined to provide a basis for comparison with the predicted structures. It is comprised of hydrogen bonded sheets that lie on the mirror plane perpendicular to the b axis. Each 5-azauracil molecule hydrogen bonds to two other 5-azauracil molecules and three different water molecules. The water molecule hydrogen bonds to three different 5-azauracil molecules – once as an acceptor and twice as a donor. 5-Azauracil acts as a hydrogen bond donor twice, once to a water molecule and once to another 5-azauracil molecule, and a hydrogen bond acceptor three times, twice to water molecules and once to a 5-azauracil molecule. This balances the acceptor/donor ratio in the structure. There are no strong hydrogen bonds between adjacent sheets. Two hydrogen bonded rings are present in the sheet, a $R_3^3(10)$ ring and a larger $R_5^5(16)$ ring (figure 7.7).
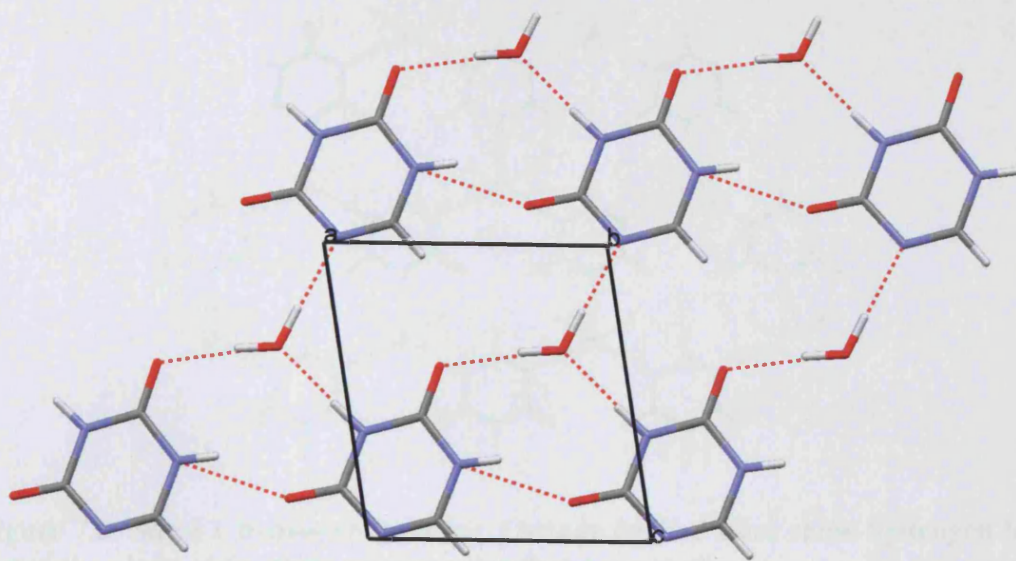
Figure 7.7: The hydrogen bond scheme found in the experimental structure of 5-azauracil monohydrate

Upon examination of the other predicted structures, none showed the same hydrogen bonded motif as ExptMinOpt structure. However, two recurring hydrogen bonded sheets were identified, with sheet 1 (figure 7.8) found in 14 structures and sheet 2 (figure 7.9) in 5 predicted structures, including the three lowest energy structures. The remaining 5 structures all exhibited extended three dimensional hydrogen bonded arrays that do not show strong similarities to one another.

Sheet 1 is comprised of one hydrogen bonded ring, $R_6^5(18)$ (figure 7.8). Within the sheet the 5-azauracil molecules only form three hydrogen bonds, one each from the two amino groups and one to the aza nitrogen. The water molecule forms three hydrogen bonds within the sheet – the oxygen atom acts as acceptor to two hydrogen bonds from 5-azauracil molecules and one of the hydrogen atoms is donated to a third 5-azauracil molecule. The second water hydrogen atom is orientated out of the plane of the sheet and forms a bifurcated interaction with two 5-azauracil carbonyl groups in the adjacent sheet.
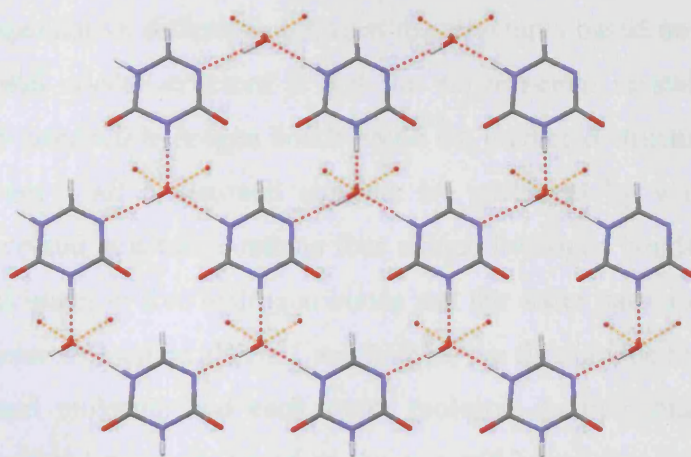
Figure 7.8: Sheet 1 hydrogen bonding. Orange dashed lines show hydrogen bonds out of the plane of the sheet to molecules that are not shown

Sheet 2 is similar to sheet 1, being comprised of the same $R_6^5(18)$ hydrogen bonded rings and in both sheets the water molecule and the 5-azauracil form the same hydrogen bonds. Sheet 1 contains 5-azauracil molecules in only one orientation, leading to all of the constituent rings being orientated in the same direction. In comparison, sheet 2 has 5-azauracil molecules in two different orientations, with every second molecule rotated by 120°, leading to two different orientations of the $R_6^5(18)$ rings, with adjacent rings offset by 120° (figure 7.9). In sheet 2, similarly to sheet 1, one of the water hydrogen atoms points out of the plane of the sheet forming a bifurcated hydrogen bond interaction with two carbonyl oxygen atoms in the adjacent sheet.
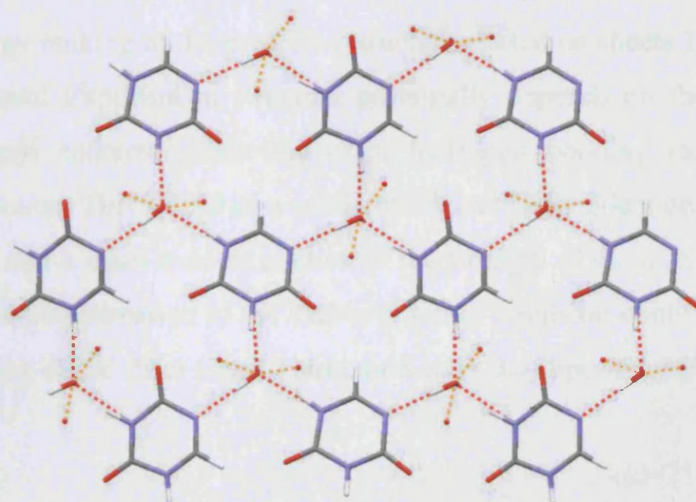
Figure 7.9: Sheet 2 hydrogen bonding. Orange dashed lines show hydrogen bonds out of the plane of the sheet to molecules that are not shown

The primary qualitative difference between the structures based on sheets 1 and 2 and the experimental crystal structure is that the experimental crystal structure contains 5-azauracil···5-azauracil hydrogen bonds while the predicted structures based on sheets 1 and 2 do not – all 5-azauracil contacts are mediated by water molecules. The experimental crystal structure contains four unique hydrogen bonds – each 5-azauracil molecule participates in five hydrogen bonds and the water participates in three. In the predicted structures based on sheets 1 and 2, there are five unique hydrogen bonds, with each 5-azauracil molecule and each water molecule participating in five hydrogen bonds. The predicted structures based on sheets 1 and 2 have two fewer strong hydrogen bonds within the sheet than the experimental structure, but in compensation a water hydrogen atom is orientated out of the plane of the sheet and forms a bifurcated interaction with two carbonyl groups in the adjacent sheet. This bifurcated interaction provides the extra hydrogen bonds and is an advantage compared to the experimental crystal structure where there is no hydrogen bonding between sheets.

In both sheets 1 and 2 this molecular arrangement brings the carbonyl groups into close head-to-head contact with one another. The carbonyl O···O distance is less than 3 Å in all structures, compared to the van der Waal's radius of oxygen which is 1.52 Å. While carbonyl···carbonyl interactions are well established,[276] the carbonyl groups are positioned so the partially positively charged carbon atom interacts with the partially negatively charged oxygen atom, in contrast to the head-to-head interaction observed here.

The energy ranking of the predicted structures based on sheets 1 and 2 compared to the experimental ExptMinOpt structure principally depends on the balance of the repulsive carbonyl contact against the extra hydrogen bonding introduced by the bifurcated interaction. This would be a subtle effect, very dependent on the performance of the potential, and a small over-estimation of the strength of the inter-sheet bifurcated interaction, or under-estimation of the carbonyl group repulsion could lead to the more favourable ranking of the sheet 1 and 2 structures over the experimental structure.

7.5 Conclusion

The successful reproduction of four ordered structures of ice using the FIT dispersion-repulsion potential in conjunction with the distributed multipole electrostatic model, proved it capable of reproducing water⋯water contacts. This potential was also successful in energy minimising 17 out of the 22 hydrate structures tested, returning acceptably small errors in the cell parameters and retaining the hydrogen bonding present in the experimental structures. This proved it capable of adequately modelling contacts between water and organic molecules in hydrate structures. An upper F-value limit for a successful reproduction of a monohydrate of approximately 120 was determined. The energy minimisation of corresponding anhydrous structures for eight of the 22 hydrates showed the inclusion of water in the crystal structure did not significantly degrade the accuracy of the energy minimisation in five of the eight cases.

5-Azauracil was chosen from the set of hydrates used in the potential validation to be the initial test candidate for the crystal structure prediction of a monohydrate system. Modification of the MOLPAK structure generation method was required with a range of clusters, each containing a 5-azauracil molecule and a water molecule, used as separate MOLPAK inputs. Crystallographic knowledge that hydrogen bonds have preferred contact distance and angles was exploited in the definition of the range of clusters. CSD analyses gave common length and angle statistics for hydrogen bonds to each of the acceptor functional groups found in 5-azauracil, and so the hydrogen bond length and angles could be fixed in the clusters to the values found. For each 5-azauracil acceptor the placement of the water molecule on the circle around each acceptor defined by fixing hydrogen bond length and angles to their typical values, and the orientation of the water molecule were varied systematically. This approach limited the number of clusters considered in the total search to 67, upon each of which a MOLPAK search was carried out to produce the final set of predicted monohydrate structures.

The principal requisite of any crystal structure prediction was satisfied by the search results – the search successfully generated a hypothetical structure that corresponded exactly to the energy minimised experimental structure. Many of the individual searches found this structure and it proved energetically competitive with the other predicted structures, only 4.3 kJ mol^{-1} higher in energy than the global energy

minimum. The majority of the low energy predicted structures were based on one of two closely related hydrogen bonded sheet motifs. These sheets were fundamentally different from the experimental structure because they contained no 5-azauracil⋯5-azauracil hydrogen bonds, with all interactions mediated by water molecules. In contrast to the experimental structure these predicted structures had interactions between adjacent sheets, with a bifurcated interaction between water and the two 5-azauracil carbonyl acceptors. These structures also had unconventional repulsive carbonyl head-to-head interactions which could possibly lead to them being classified as unrealistic. Discounting these structures on the basis of these carbonyl contacts would rank the ExptMinOpt structure sixth, 2.1 kJ mol^{-1} above the new energy minimum.

Analysis of the results suggest that the search may not be satisfactorily close to completeness because two of the 25 lowest energy structures were only generated by a single individual search, suggesting that further individual searches starting from clusters with molecular orientations not considered in this search could potentially yield additional low energy predicted structures.

The search strategy employed here was specific to 5-azauracil, an isolated site hydrate, and may not perform as well on other types of hydrates such as channel or sheet hydrates. For example the crystal structures of thymine and cytosine hydrates were poorly reproduced during the potential testing and CSP searches for these structures could be expected to be less successful. It should also be noted that for the truly blind prediction of the hydrates of a candidate molecule, the search would have to address the potential for higher hydrates, such as di- and tri- hydrates, as well as fractional hydrates such as hemihydrates and sesquihydrates. With this would come proportional increases in computational expenditure. However the general methodology is applicable to any system where more than one molecule must be considered, such as other hydrates, solvates, co-crystals and Z' > 1 systems.

Chapter 8 – Conclusion

8.1 Overview

The work presented in chapters 3-7 includes the results of experimental polymorph screening carried out on four small organic molecules: 5-fluorouracil, 5-fluorocytosine, 3-azabicyclo[3.3.1]nonane-2,4-dione and 4-hydroxycoumarin. All four molecules are small, rigid organic molecules containing both hydrogen bond acceptor and donor groups and these highly functionalised molecules proved versatile in crystallising both on their own and with solvent. In the course of this work eight new polymorphs were discovered, with the structures of six of them fully determined by single crystal X-ray diffraction. The crystal structures of 12 solvates and two co-crystals were also fully characterised.

Computational crystal structure prediction was used to generate low energy potential polymorphs for six molecules: 5-fluorouracil, 5-fluorocytosine, 4-hydroxycoumarin, coumarin, 6-methoxycoumarin and 7-hydroxycoumarin. After the experimental polymorph screens described above, nine fully determined crystal structures were available for these six molecules. Using the criteria of the CCDC international blind tests of crystal structure prediction,[121-123] in which a structure ranked within the top three predictions is viewed as a successful prediction, the predictions of 5-fluorouracil form 2, 5-fluorocytosine form 2, coumarin and 7-hydroxycoumarin, were successful, with all four of these structures found at the global energy minimum. The 6-methoxycoumarin computational search fulfilled the principal requirement of crystal structure prediction in that it found the experimental structure within the commonly accepted energy range for polymorphism, but was only ranked 22nd in the search results. 4-Hydroxycoumarin was the only molecule for which the computational search method failed to successfully predict at least one polymorph. This was principally because both of the fully determined polymorphs fell outside of the limitations of the computational search procedure as both contained two molecules in the asymmetric unit. The limitations of the computational search method also led to neither 5-fluorouracil form 1 ($Z' = 4$) nor 5-fluorocytosine form 1 (high symmetry space group) being predicted. Such limitations are an artefact of the method used in this work and more thorough search methods, incorporating more space groups are now available.[253;277]

In all cases crystal structure prediction generated a range of structures within the energy range for potential polymorphism. For each molecule analysis of the hydrogen bonding present in the low energy predicted structures revealed that only two or three hydrogen bonded motifs recurred repeatedly. For 5-fluorouracil three recurring motifs were discovered, with the motif present in the lowest energy structures found in the newly discovered form 2, and with the most common recurring motif present in two solvates, though the form 1 polymorph contained a unique hydrogen bond pattern. 5-Fluorocytosine provided an even more unequivocal example with only two identified motifs from the low energy predicted structures – the vastly more common motif was found in both newly discovered polymorphs and all but one of the solvates, with the remaining solvate containing the alternative motif. The predicted structures of 4-hydroxycoumarin were all found to contain the same hydrogen bonded ribbon, but with two distinct configurations of molecules around the hydrogen bond. Each configuration was observed in one of the polymorphs of 4-hydroxycoumarin found experimentally.

The study on 3-azabicyclo[3.3.1]nonane-2,4-dione provided an exception to the assertion that a crystal structure will be forthcoming with the most common motif found in the crystal structure prediction search results. Extensive crystal structure prediction during the CCDC blind test of crystal structure prediction[122] produced two distinct motifs – hydrogen bonded dimer-based and chain-based structures – but only the less common chain motif was observed experimentally, albeit in all of its discovered crystal structures. The hydrogen bonding in this system was found to be unusually weak, as evidenced by the formation of a high temperature plastic phase, and confirmed by calculation of the electrostatic potential on the van der Waals surface of the molecule which showed the imide N-H to be a weak hydrogen bond donor.

While crystal structure prediction is principally aimed at the prediction of polymorphism, the prevalence of solvates, and especially hydrates, for organic molecules makes the prediction of solvates desirable. The ultimate aim of such an extension of the crystal structure prediction method would be the ability to computationally characterise all of the solid state forms of an organic molecule rather than just its polymorphs.[273] The work presented in chapter 7 describes the application of crystal structure prediction to a monohydrate test subject, 5-azauracil, and shows that

the generation of hypothetical monohydrate structures is possible using existing methods. The intermolecular potential used was found to be capable of reproducing intermolecular interactions between water and organic molecules. The search method proved successful, generating the known 5-azauracil monohydrate structure within 5 kJ mol^{-1} of the global energy minimum, making it energetically feasible. Two principal recurring hydrogen bonded motifs were identified from the hypothetical structures and while neither of these motifs was that oberved in the experimental structure, they provided different realistic hydrogen bond patterns for this system. The most significant factor used to limit the scope of the structure generation was the fore-knowledge that the experimental structure was a monohydrate, and the search was consequently limited to monohydrate structures. Even so the increase in computational expense was commensurate with the increased dimensionality of including a second molecule in the search procedure – the building of a range of hydrogen bonded 5-azauracil-water clusters resulted in 67 individual searches compared to the one commonly required to generate potential polymorphs of a rigid candidate molecule. The search strategy employed was specific to 5-azauracil, but the general method is applicable to other isolated site hydrates and could be used in a similar manner for solvates and co-crystals.

8.2 Further aims

In none of the studies reported here was there an energy gap between any experimental structure and the predicted structures, which would imply that the discovery of further polymorphs could be discounted. For all molecules studied in this work a significantly larger range of predicted polymorphs was generated than were already known, or subsequently found experimentally. This is a general feature of computational crystal structure prediction and a greater understanding of the causes of polymorphism is required to reliably discount many of the predicted structures that would never be found experimentally. Lattice energy can only be used as a structural discriminant to discount structures outside of the commonly accepted range for polymorphism (approximately 8-10 kJ mol^{-1}).[1] The remaining predicted structures have realistic intermolecular interactions, look crystallographically plausible[278] and are calculated to be, at the very least, energetically competitive with the known polymorphs. Some limitations of the

method can account for generation of excess structures, with the lack of treatment of lattice vibrations and the zero-point energy leading to structures being classified as distinct that would in reality have a sufficiently low energy barrier between them that could be overcome by thermal energy, leading to only one of the structures being observed at room temperature.

The most significant criteria for selection of which of the low energy predictions will correspond to experimentally achievable structures must come with the further understanding of nucleation processes. It is at the nucleation stage, between molecular self-assembly in solution and the post-critical-nucleus stage that the vast majority of potential polymorphs are selected against. Different crystallisation conditions can only influence the crystallisation pathway to a certain extent, and many of the potential polymorphs predicted will remain unobserved in the macroscopic crystalline state, even after thorough polymorph screening. Computational methods generate the end result of a 'crystallisation' – the crystal structure – without any incorporation of the selection process that the pathway from solution to crystal structure exerts on the outcome of experimental crystallisations. Only upon understanding the crystallisation pathways available to molecules, can this knowledge be used to discriminate between hypothetical structures, reducing the sub-set of plausible hypothetical structures.

8.3 Conclusion

This thesis contributes to a larger inter-institutional project to control and predict the organic solid state and, as part of this project, a substantial number of compounds will be screened for polymorphism to add to our knowledge about polymorphic and non-polymorphic systems. In the studies reported in this work four molecules were experimentally screened for polymorphism and in all cases the molecule was found to exhibit polymorphism and show a propensity for solvate formation leading to more solid state forms than initially expected at the outset. This clearly shows that crystal structure prediction methods cannot be accurately evaluated without complimentary experimental polymorph screening of the molecules under study. The experimental crystallisation screens documented in this work also include details of the crystallisation conditions that not only yielded new polymorphs, but those that gave known polymorphs and those where the solvent 'interrupted' the common crystallisation

pathway to produce solvate crystal structures. Such knowledge of crystallisation conditions will play an important role in developing an understanding of how kinetic factors influence crystallisation, leading to the incorporation of a kinetic aspect into the computational model for the prediction of polymorphism. Two of the studies presented in this thesis included some investigation of the influence that nucleation processes can have on polymorphic outcome. Molecular dynamics simulations of supersaturated 5-fluorouracil solutions[180] revealed different self-assemblies in solution that led directly to different polymorphs containing these units. In the case of 3-azabicyclo[3.3.1]nonane-2,4-dione the absence of an alternative predicted motif in the crystalline state was attributed to the weakness of the hydrogen bonds, rationalised from the formation of a plastic phase and calculation of the electrostatic surface of the molecule, which along with the globular shape of the molecule would be conducive to facile molecular rearrangement in the fluxional pre-critical nucleation stage. This would enable the molecules to reorientate into the observed, more stable motif, rather than produce the alternative predicted motif.

While such advances will be important in the future, this work shows that current crystal structure prediction methods often generate known polymorphs as the most stable predicted structure. In those cases where known structures were not found in the search, usually because of limitations inherent to the method, the hypothetical structures that were generated could still be used to rationalise the solid state forms found experimentally. The crystal structure prediction results were used to identify the most reliable, robust hydrogen bonded motifs for each molecule with these predicted motifs proving realistic: those predicted for several of the molecules studied were then found experimentally, not only in polymorphs of the molecule in question, but also in many solvates where the role of the solvent was ancillary to the formation of the motif identified from the predicted structures.

Summary of publications

5-Fluorouracil

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

2. Barnett, S.A., Hulme, A.T. & Tocher, D.A. 5-Fluorouracil and thymine form a crystalline solid solution. *Acta Crystallogr.* **C62**, o412-o415 (2006).

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

4. Hulme, A.T. & Tocher, D.A. 5-Fluorouracil-1,4-dioxane (4/1). *Acta Crystallogr.* **E60**, o1781-o1782 (2004).

5. Hulme, A.T. & Tocher, D.A. 5-Fluorouracil-dimethylformamide (2/1). *Acta Crystallogr.* **E60**, o1783-o1785 (2004).

6. Hulme, A.T. & Tocher, D.A. 5-Fluorouracil-dimethyl sulfoxide (1/1). *Acta Crystallogr.* **E60**, o1786-o1787 (2004).

7. Hulme, A.T. & Tocher, D.A. 5-Fluorouracil-2,2,2-trifluoroethanol (1/1). *Acta Crystallogr.* **E61**, o3661-o3663 (2005).

5-Fluorocytosine

8. Hulme, A.T. & Tocher, D.A. The Discovery of New Crystal Forms of 5-Fluorocytosine Consistent with the Results of Computational Crystal Structure Prediction. *Cryst. Growth & Des.* **6**, 481-487 (2005).

9. Hulme, A.T. & Tocher, D.A. 4-Amino-5-fluoropyrimidin-2(1H)-one-2-amino-5-fluoropyrimidin-4(3H)-one-water (1/1/1). *Acta Crystallogr.* **E61**, o2112-o2113 (2005).

3-Azabicyclo[3.3.1]nonane-2,4-dione

10. Hulme, A.T., Johnston, A., Florence, A.J., Fernandes, P., Bedford, C.T., Welch, G.W.A., Sadiq, G., Haynes, D.A., Motherwell, W.D.S., Tocher, D.A., Price, S.L. The Search for a Predicted Hydrogen Bonded Motif – a Multidisciplinary Investigation into the Polymorphism of 3-Azabicyclo[3.3.1]nonane-2,4-dione and Rationalisation of Observed Forms. *J. Am. Chem. Soc.*, in press (2007).

11. Hulme, A.T., Johnston, A., Florence, A.J. & Tocher, D.A. 3-Azabicyclo[3.3.1]nonane-2,4-dione–acetic acid (1/1). *Acta Crystallogr.* **E62**, o545-o547 (2006).

12. Hulme, A.T., Fernandes, P., Florence, A.J., Johnston, A. & Shankland, K. Powder Study of 3-Azabicyclo[3.3.1]nonane-2,4-dione form 2. *Acta Crystallogr.* **E62**, o3046-o3048 (2006).

13. Hulme, A.T., Fernandes, P., Florence, A.J., Johnston, A. & Shankland, K. Powder study of 3-Azabicyclo[3.3.1]nonane-2,4-dione 1-methylnaphthalene hemisolvate. *Acta Crystallogr.* **E62**, o3752-o3754 (2006).

5-Azauracil monohydrate

14. Hulme, A.T. & Price, S.L. Towards the Prediction of Organic Hydrate Crystal Structures. *J. Comput. Theoret. Chem.*, accepted (2007).

Other work

15. Hulme, A.T., Lancaster, R.W. & Cannon, H.F. Clarification of the crystalline forms of androsterone. *Cryst. Eng. Comm.* **8**, 309-312 (2006).

Contribution: Determined the crystal structures of androsterone and its hemihydrate and reviewed the literature on androsterone

16. Lancaster, R.W., Karamertzanis, P.G., Hulme, A.T., Tocher, D.A., Covey, D.F., Price, S.L. Racemic progesterone: predicted in silico and produced in the solid state. *Chem. Comm.*, 4921-4923 (2006).

Contribution: Determined the crystal structure of racemic progesterone

17. Lancaster, R.W., Karamertzanis, P.G., Hulme, A.T., Tocher, D.A., Lewis, T.C., Price, S.L. The Polymorphism of Progesterone: Stabilization of a 'Disappearing' Polymorph by Co-crystallisation. *J. Pharm. Sci.*, submitted (2006).

Contribution: Redetermined the crystal structures of forms 1 and 2 of progesterone, and determined the crystal structure of the related steroid pregnenolone and the structure a co-crystal of pregnenolone and progesterone

18. Barnett, S.A., Hulme, A.T. & Tocher, D.A. A low-temperature redetermination of metaldehyde. *Acta Crystallogr.* **E61**, o857-o859 (2005).

Contribution: Grew and redetermined the crystal structure of metaldehyde

Bibliography

(1)   Bernstein, J. *Polymorphism in Molecular Crystals*; Clarendon Press: Oxford, 2002.

(2)   Byrn, S. R.; Pfeiffer, R. R.; Stowell, J. G. *Solid-state Chemistry of Drugs*; SSCI Inc.: West Lafayette, Indiana, 1999.

(3)   Grant, D. J. W. Theory and Origin of Polymorphism; In *Polymorphism in Pharmaceutical Solids*; Brittain, H. G., ed. Marcel Dekker: New York, 1999; pp 1-34.

(4)   ICH Expert Working Group *ICH Q6A harmonised tripartite guideline. Specifications: test procedures and acceptance criteria for new drug substances and new drug products: chemical substances*; 1999.

(5)   Yu, L.; Reutzel, S. M.; Stephenson, G. A. *Pharm.Res.* **1998**, *1*, 118-127.

(6)   Datta, S.; Grant, D. J. W. *Nature Reviews* **2004**, *3*, 42-57.

(7)   McCrone, W. C. Polymorphism; In *Physics and chemistry of the organic solid state Vol. II*; Fox, D. J., Labes, M. M., Weissberger, A., eds. Wiley Interscience: New York, 1965; pp 726-767.

(8)   Vippagunta, S. R.; Brittain, H. G.; Grant, D. J. W. *Adv.Drug Del.Rev.* **2001**, *48*, 3-26.

(9)   Hancock, B. C.; Zografi, G. *J.Pharm.Sci.* **1997**, *86*, 1-12.

(10) Hancock, D. C.; Shalaev, E. Y.; Shamblin, S. L. *J.Pharm.Pharmacol.* **2002**, *54*, 1151-1152.

(11) Morris, K. R. Structural Aspects of Hydrates and Solvates; In *Polymorphism in Pharmaceutical Solids*; Brittain, H. G., ed. Marcel Dekker Inc.: New York, 1999; pp 125-181.

(12) Perrier, P. R.; Byrn, S. R. *J.Org.Chem* **1982**, *47*, 4671-4676.

(13) Byrn, S. R.; Pfeiffer, R.; Ganey, M.; Hoiberg, C.; Poochikian, G. *Pharm.Res.* **1995**, *12*, 945-954.

(14) Winter, G. Polymorphs and solvates of molecular solids in the pharmaceutical industry; In *Reactivity of Molecular Solids*; Boldyreva, E., Boldyrev, V., eds. John Wiley & Sons Ltd.: 1999; pp 241-270.

(15) Gorbitz, C. H.; Hersleth, H.-P. *Acta Crystallogr.* **2000**, *B56*, 526-534.

(16) Almarsson, O.; Zaworotko, M. J. *Chem.Comm.* **2004**, 1889-1896.

(17) Dunitz, J. D. *Cryst.Eng.Comm.* **2003**, *5*, 506.

(18) Dunitz, J. D. *Pure Appl.Chem.* **1991**, *63*, 177-185.

(19) Threlfall, T. L. *Analyst* **1995**, *120*, 2435-2460.

(20) Brock, C. P.; Schweizer, B.; Dunitz, J. D. *J.Am.Chem.Soc.* **1991**, *113*, 9811-9820.

(21) Seddon, K. R. *Cryst.Growth & Des.* **2004**, *4*, 1087.

(22) Bernstein, J. *Cryst.Growth & Des.* **2005**, *5*, 1661-1662.

(23) Allen, F. H. *Acta Crystallogr.* **2002**, *B58*, 380-388.

(24) Yu, L.; Stephenson, G. A.; Mitchell, C. A.; Bunnell, C. A.; Snorek, S. V.; Bowyer, J. J.; Borchardt, T. B.; Stowell, J. G.; Byrn, S. R. *J.Am.Chem.Soc.* **2000**, *122*, 585-591.

(25) Price, C. P.; Grzesiak, A. L.; Matzger, A. J. *J.Am.Chem.Soc.* **2005**, *127*, 5512-5517.

(26) Griesser, U. Relevance and Analysis of Polymorphism in Drug Development. 5-4-2006. Presentation to British Crystallographic Association.

(27) Grant, D. J. W. Theory and origin of polymorphism; In *Polymorphism in pharmaceutical solids*; Brittain, H. G., ed. Marcel Dekker: New York, 1999; pp 1-34.

(28) Zhang, G. G. Z.; Law, D.; Schmitt, E. A.; Qiu, Y. *Adv.Drug Del.Rev.* **2004**, *56*, 371-390.

(29) DeCamp, W. H. *Am.Pharm.Rev.* **2001**, *4*, 70-77.

(30) Huang, L.-F.; Tong, W.-Q. *Adv.Drug Del.Rev.* **2004**, *56*, 321-334.

(31) Lancaster, R. W. 2006.

(32) Haleblian, J.; McCrone, W. C. *J.Pharm.Sci.* **1969**, *58*, 911-929.

(33) Hilfiker, R.; Berghausen, J.; Blatter, F.; Burkhard, A.; De Paul, S. M.; Freiermuth, B.; Geoffroy, A.; Hofmeier, U.; Marcolli, C.; Siebenhaar, B.; Szelagiewicz, M.; Vit, A.; von Raumer, M. *J.Therm.Anal.Cal.* **2003**, *73*, 429-440.

(34) 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*, 413-417.

(35) Raw, A. S.; Furness, M. S.; Gill, D. S.; Adams, R. C.; Holcombe Jr., F. O.; Yu, L. X. *Adv.Drug Del.Rev.* **2004**, *56*, 397-414.

(36) Burger, A.; Ramberger, R. *Mikrochim.Acta* **1979**, *II*, 273-316.

(37) Rodriguez-Spong, B.; Price, C. P.; Jayasankar, A.; Matzger, A. J.; Rodriguez-Hornedo, N. *Adv.Drug Del.Rev.* **2004**, *56*, 241-274.

(38) Gavezzotti, A. *Modelling Simul.Mater.Sci.Eng.* **2002**, *10*, R1-R29.

(39) Davey, R. J.; Garside, J. *From Molecules to Crystallisers - An Introduction to Crystallisation*; Oxford University Press: Oxford, 2000.

(40) Davey, R. J.; Allen, K.; Blagden, N.; Cross, W. I.; Lieberman, H. F.; Quayle, M. J.; Righini, S.; Seton, L.; Tiddy, G. J. T. *Cryst.Eng.Comm.* **2002**, *4*, 257-264.

(41) Mullin, J. W. *Crystallisation*; Butterworth-Heinemann: Oxford, 1992.

(42) Davey, R. J.; Dent, G.; Mughal, R. K.; Parveen, S. *Cryst.Growth & Des.* **2006**, *ASAP*.

(43) Frankenbach, G. M.; Etter, M. C. *Chem.Mater.* **1992**, *4*, 278.

(44) Threlfall, T. L. Growing Crystals, Growing Polymorphs (and some Thoughts on Crystal Growth). 5-4-2006. Presentation to British Crystallographic Association.

(45) Desiraju, G. R. *Nature Materials* **2002**, *1*, 77-79.

(46) Bernstein, J.; Davey, R. J.; Henck, J.-O. *Angew.Chem.Int.Ed.* **1999**, *38*, 3440-3461.

(47) Nangia, A.; Desiraju, G. R. *Chem.Comm.* **1999**, 605-606.

(48) Kossel, W. *Nachr.Ges.Wiss.Goettingen, Math.Physik.Kl.* **1927**, *IIa*, 135-143.

(49) Verma, A. R.; Krishna, P. *Polymorphism and polytypism in crystals*; John Wiley and Sons: 1966.

(50) Burton, W. K.; Cabrera, N.; Frank, F. C. *Phil.Trans.* **1951**, *A*, 299-358.

(51) Hartman, P.; Perdok, W. G. *Acta Crystallogr.* **1955**, *8*, 49-52.

(52) Lahav, M.; Leiserowitz, L. *Cryst.Growth & Des.* **2006**, *6*, 619-624.

(53) Weissbuch, I.; Addadi, L.; Lahav, M.; Leiserowitz, L. *Science* **1991**, *253*, 637-645.

(54) Dunitz, J. D.; Bernstein, J. *Acc.Chem.Res.* **1995**, *28*, 193-200.

(55) Bernstein, J.; Henck, J.-O. *Mater.Res.Bull.* **1998**, *33*, 119-128.

(56) Jacewicz, V. W.; Nayler, J. H. C. *J.Appl.Cryst.* **1979**, *12*, 396-397.

(57) Morissette, S. L.; Soukasene, S.; Levinson, D.; Cima, M. J.; Almarsson, O. *PNAS* **2003**, *100*, 2180-2184.

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

(59) Florence, A. J.; Johnston, A.; Price, S. L.; Nowell, H.; Kennedy, A. R.; Shankland, N. *J.Pharm.Sci.* **2006**, *95*, 1918-1930.

(60) Thallapally, P. K.; Jetti, R. K. R.; Katz, A. K.; Carrell, H. L.; Singh, K.; Lahiri, K.; Kotha, S.; Boese, R.; Desiraju, G. R. *Angew.Chem.Int.Ed.* **2004**, *43*, 1149-1155.

(61) Blagden, N.; Davey, R. J.; Rowe, R.; Roberts, R. J. *Int.J.Pharm.* **1998**, *172*, 169-177.

(62) Gu, C.-H.; Chatterjee, K.; Young Jr., V.; Grant, D. J. W. *J.Cryst.Growth* **2002**, *235*, 471-481.

(63) Lang, M.; Grzesiak, A. L.; Matzger, A. J. *J.Am.Chem.Soc.* **2002**, *124*, 14834-14835.

(64) Allan, D. R.; Clark, S. J.; Ibberson, R. M.; Parsons, S.; Pulham, C. R.; Sawyer, L. *Chem.Comm.* **1999**, 751-752.

(65) Fabbiani, F. P. A.; Allan, D. R.; Parsons, S.; Pulham, C. R. *Cryst.Eng.Comm.* **2005**, *7*, 179-186.

(66) Fabbiani, F. P. A.; Allan, D. R.; David, W. I. F.; Moggach, S. A.; Parsons, S.; Pulham, C. R. *Cryst.Eng.Comm.* **2004**, *6*, 504-511.

(67) Gavezzotti, A. Crystal Packing. IUCr . 2001. Chester, International Union of Crystallography.

(68) Kitaigorodski, A. I. *Organic Chemical Crystallography*; Consultants Bureau: New York, 1961.

(69) Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry*; John Wiley & Sons: Chichester, UK, 2000.

(70) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, 1996.

(71) Gavezzotti, A. *Synlett* **2002**, *2*, 201-214.

(72) Price, S. L. Towards More Accurate Model Intermolecular Potentials for Organic Molecules; In *Reviews in Computational Chemistry 14*; Lipkowitz, K. B., Boyd, D. B., eds. John Wiley & Sons: New York, 2000; pp 225-289.

(73) Gavezzotti, A. *Cryst.Eng.Comm.* **2002**, *4* , 343-347.

(74) Price, S. L. *Cryst.Eng.Comm.* **2004**, *6*, 344-356.

(75) Mooij, W. T. M.; van Eijck, B. P.; Kroon, J. *J.Am.Chem.Soc.* **2000**, *122*, 3500-3505.

(76) Lennard-Jones, J. E. *Proceedings of the Royal Society* **1924**, *A106*, 463.

(77) Williams, D. E.; Starr, T. L. *Comput.Chem.* **1977**, *1*, 173-177.

(78) Williams, D. E.; Cox, S. R. *Acta Crystallogr.* **1984**, *B40*, 404-417.

(79) Cox, S. R.; Hsu, L.-Y.; Williams, D. E. *Acta Crystallogr.* **1981**, *A37*, 293-301.

(80) Williams, D. E.; Houpt, D. J. *Acta Crystallogr.* **1986**, *B42*, 286-295.

(81) Hsu, L.-Y.; Williams, D. E. *Acta Crystallogr.* **1980**, *A36*, 277-281.

(82) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J.Phys.Chem.* **1996**, *100*, 7352-7360.

(83) Williams, D. E. *J.Mol.Struct.* **1999**, *485-486*, 321-347.

(84) Williams, D. E. *J.Comput.Chem.* **2001**, *22*, 1-20.

(85) Williams, D. E. *J.Comput.Chem.* **2001**, *22*, 1154-1166.

(86) Williams, D. E.; Weller, R. R. *J.Am.Chem.Soc.* **1983**, *105*, 4143-4148.

(87) Karamertzanis, P. G.; Pantelides, C. C. *Mol.Sim.* **2004**, *30*, 413-436.

(88) Stone, A. J.; Alderton, M. *Mol.Phys.* **1985**, *56*, 1047-1064.

(89) Buckingham, A. D.; Fowler, P. W. *Can.J.Chem* **1985**, *63*, 2018.

(90) Price, S. L. *Adv.Drug Del.Rev.* **2004**, *56*, 301-319.

(91) Mooij, W. T. M.; Leusen, F. J. J. *Phys.Chem.Chem.Phys.* **2001**, *3*, 5063-5066.

(92) Day, G. M.; Motherwell, W. D. S.; Jones, W. *Cryst.Growth & Des.* **2005**, *5*, 1023-1033.

(93) Filippini, G.; Gavezzotti, A. *Acta Crystallogr.* **1993**, *B49*, 868-880.

(94) Pauling, L. *The Nature of the Chemical Bond*; Cornell University Press: New York, 1960.

(95) Brammer, L. Hydrogen Bonds in Inorganic Chemistry: Application to Crystal Design; In *Crystal Structure Design and Function*; Desiraju, G. R., ed. John Wiley & Sons: Chichester, 2003; pp 1-76.

(96) Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond*; Oxford University Press: Oxford, 1999.

(97)   Bondi, A. *J.Phys.Chem.* **1964**, *68*, 441-451.

(98)   Desiraju, G. R. *Acc.Chem.Res.* **1991**, *24*, 290-296.

(99)   Steiner, T.; Desiraju, G. R. *Chem.Comm.* **1998**, 891-892.

(100)   Etter, M. C. *Acc.Chem.Res.* **1990**, *23*, 120-126.

(101)   Bolton, W. *Acta Crystallogr.* **1964**, *17*, 147-152.

(102)   Coombes, D. S.; Nagi, G. K.; Price, S. L. *Chem.Phys.Lett.* **1997**, *265*, 532-537.

(103)   Desiraju, G. R. *Chem.Comm.* **2005**, 2995-3001.

(104)   Desiraju, G. R. *Acc.Chem.Res.* **1996**, *29*, 441-449.

(105)   Taylor, R.; Kennard, O. *J.Am.Chem.Soc.* **1982**, *104*, 5063-5070.

(106)   Etter, M. C.; MacDonald, J. C.; Bernstein, J. *Acta Crystallogr.* **1990**, *B46*, 256-262.

(107)   Bernstein, J.; Davis, R. E.; Shimoni, L.; Chang, N.-L. *Angew.Chem.Int.Ed.* **1995**, *34*, 1555-1573.

(108)   Desiraju, G. R. *Angew.Chem.Int.Ed.* **1995**, *34*, 2311-2327.

(109)   Gavezzotti, A. *Curr.Opin.Solid State Mater.Sci.* **1996**, *1*, 501-505.

(110)   Desiraju, G. R. *Curr.Opin.Solid State Mater.Sci.* **1997**, *2*, 451-454.

(111)   Moulton, B.; Zaworotko, M. J. *Chem.Rev.* **2001**, *101*, 1629-1658.

(112)   Maddox, J. *Nature* **1988**, *335*, 201.

(113)   Gavezzotti, A. *Acc.Chem.Res.* **1994**, *27*, 309-314.

(114)   Dunitz, J. D. *Chem.Comm.* **2003**, 545-548.

(115)   Karfunkel, H. R.; Gdanitz, R. J. *J.Comput.Chem.* **1992**, *13*, 1171-1183.

(116)   Nowell, H.; Price, S. L. *Acta Crystallogr.* **2005**, *B61*, 558-568.

(117)   Verwer, P.; Leusen, F. J. J. Computer Simulation to Predict Possible Crystal Polymorphs; Lipkowitz, K. B., Boyd, D. B., eds. John Wiley & Sons: New York, 1998; pp 327-365.

(118)   Gavezzotti, A. Zip-Promet. 1999. Univeristy of Milan, University of Milan.

(119)   van Eijck, B. P.; Kroon, J. *J.Comput.Chem.* **1999**, *20*, 799-812.

(120)   Holden, J. R.; Du, Z.; Ammon, H. L. *J.Comput.Chem.* **1993**, *14*, 422-437.

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

(122)   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. A.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. *Acta Crystallogr.* **2002**, *B58*, 647-661.

(123)   Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwer, P. *Acta Crystallogr.* **2005**, *B61*, 511-527.

(124)   Chisholm, J. A.; Motherwell, W. D. S. *J.Appl.Cryst.* **2005**, *38*, 228-231.

(125)   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*, 3819-3825.

(126)   Cross, W. I.; Blagden, N.; Davey, R. J. *Cryst.Growth & Des.* **2003**, *3*, 151-158.

(127)   Peterson, M. L.; Morissette, S. L.; McNulty, C.; Goldsweig, A.; Shaw, P.; LeQuesne, M.; Monagle, J.; Encina, N.; Marchionna, J.; Johnston, A.; Gonzalez-Zugasti, J.; Lemmo, A. V.; Ellis, S. J.; Cima, M. J.; Almarsson, O. *J.Am.Chem.Soc.* **2002**, *124*, 10958-10959.

(128)   Bruker AXS Inc. SMART (version 5.625). 2001. Madison, Wisconsin, USA.

(129)   Bruker AXS Inc. SAINT+ (version 6.22). 2001. Madison, WI, USA.

(130)   Sheldrick, G. M. SADABS (version 2.03) Program for Bruker Area Detector Absorption Correction. 2001. Gottingen, Germany, University of Gottingen.

(131)   Bruker AXS Inc. SHELXTL (version 6.12). 2000. Madison, Wisconsin, USA.

(132)   Macrae, C. F.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Shields, G. P.; Taylor, R.; Towler, M.; van de Streek, J. *J.Appl.Cryst.* **2006**, *39*, 453-457.

(133)   Watkin, D. J., Prout, C. K., and Pearce, L. J. CAMERON. 1996. Oxford, Chemical Crystallography Laboratory, University of Oxford.

(134)   Holden, J. R.; Du, Z.; Ammon, H. L. *J.Comp.Chem.* **1993**, *14*, 422-437.

(135)   Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. *J.Comput.Chem.* **1995**, *16*, 628-647.

(136)   Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G. R. *J.Chem.Soc.Perkin Trans.* **1987**, *2*, S1-S9.

(137)   Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery Jr., J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa , J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., and Pople, J. A. Gaussian 03. 2003. Wallingford CT, Gaussian Inc.

(138)   Karamertzanis, P. G. OptimalPaste, Cluster and Analysis. 2005.

(139)   Spek, A. L. PLATON - a multipurpose crystallographic tool. 2001. Utrecht, Utrecht University.

(140)   De Gelder, R.; Wehrens, R.; Hageman, J. A. *J.Comput.Chem.* **2001**, *22*, 273-289.

(141)   Motherwell, W. D. S.; Shields, G. P.; Allen, F. H. *Acta Crystallogr.* **1999**, *B55*, 1044-1056.

(142)   Gavezzotti, A. *Cryst.Eng.Comm.* **2002**, *4* , 343-347.

(143)   Price, S. L., Willock, D. J., Leslie, M., and Day, G. M. Online manual for DMAREL 3.02. www.ucl.ac.uk/~ucca17p/dmarelmanual/dmarel.html . 2001.

(144)   Cramer, C. J. *Essentials of Computational Chemistry*; John Wiley & Sons: Chichester, 2002.

(145)   Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons: Chichester, England, 1999.

(146)   Smith, W. *CCP5 Quarterly* **1982**, *4*, 13-25.

(147)   Smith, W. MDMULP, CCP5 program library. 1982. Daresbury Laboratory, UK.

(148)   Price, S. L. *Reviews in Computational Chemistry* **2000**, *14*, 225-289.

(149)   Kitiagorodsky, A. I. *Molecular Crystals and Molecules*; Academic Press: New York, 1973.

(150)   Pertsin, A. J.; Kitaigorodski, A. I. *The Atom-Atom Potential Method*; Springer-Verlag: Berlin Heidelberg, 1987.

(151)   Williams, D. E. PCK83. QCPE Program 548. Quantum Chemistry Program Exchange. 1983. Indiana Univ., Bloomington, Indiana, USA.

(152)   Filippini, G.; Gavezzotti, A. *Acta Crystallogr.* **1993**, *B49*, 868-880.

(153)   Duschinsky, K.; Pleven, E.; Heidelberger, C. *J.Am.Chem.Soc.* **1957**, *79*, 4559-4560.

(154)   Heidelberger, C.; Chaudhuari, N. K.; Daneberg, P.; Morreh, D.; Griesbach, L.; Duschinsky, K.; Schnitzer, R. J.; Pleven, E. *Nature* **1957**, *179*, 663-666.

(155)   Rutman, R. J.; Cantarow, A.; Paschkis, K. E. *Cancer Res.* **1954**, *14*, 119-126.

(156)   Grem, J. L. *Inv.New.Drugs* **2000**, *18*, 299-313.

(157)   *British National Formulary*; British Medical Association/Royal Pharmaceutical Society of Great Britain: UK, 2003.

(158)   Berenbaum, M. C. *Chemotherapy* **1979**, *25*, 54-59.

(159)   Fallon III, L. *Acta Crystallogr.* **1973**, *B29*, 2549-2556.

(160)   Steed, J. W. *Cryst.Eng.Comm.* **2003**, *5*, 169-179.

(161)   Vangala, V. R.; Nangia, A.; Lynch, V. M. *Chem.Comm.* **2002**, 1304-1305.

(162)   Choudhury, A. R.; Guru Row, T. N. *Cryst.Growth & Des.* **2004**, *4*, 47-52.

(163)   Voet, D.; Rich, A. *J.Am.Chem.Soc.* **1969**, *91*, 3069-3075.

(164)   Zaitu, S.; Miwa, Y.; Taga, T. *Acta Crystallogr.* **1995**, *C51*, 1857-1859.

(165)   Kim, S. H.; Rich, A. *J.Mol.Biol.* **1969**, *42*, 87ff.

(166)   Kim, S. H.; Rich, A. *Science* **1967**, *158*, 1046ff.

(167)   Kennedy, A. R.; Okoth, M. O.; Sheen, D. B.; Sherwood, J. N.; Vrcelj, R. M. *Acta Crystallogr.* **1998**, *C54*, 547-550.

(168)   Gopalan, R. S.; Kulkarni, G. U.; Rao, C. N. R. *Chem.Phys.Chem.* **2000**, *1*, 127-135.

(169)   Sternglanz, H.; Bugg, C. E. *Biochim.Biophys.Acta* **1975**, *378*, 1.

(170)   Sternglanz, H.; Bugg, C. E. *Biochim.Biophys.Acta* **1975**, *378*, 1.

(171)   Portalone, G.; Bencivenni, L.; Colapietro, M.; Pieretti, A.; Ramondo, F. *Acta.Chem.Scand.* **1999**, *53*, 57.

(172)   Barnett, S. A. and Copley, R. C. B. Crystal Structure of eniluracil. 2006.

(173)   Hulme, A. T.; Tocher, D. A. *Acta Crystallogr.* **2005**, *E61*, o3661-o3663.

(174)   Fauvet, G.; Massaux, M.; Chevalier, R. *Acta Crystallogr.* **1978**, *B34*, 1376.

(175)   Hulme, A. T.; Tocher, D. A. *Acta Crystallogr.* **2004**, *E60*, o1781-o1782.

(176)   Hulme, A. T.; Tocher, D. A. *Acta Crystallogr.* **2004**, *E60*, o1783-o1785.

(177)   Hulme, A. T.; Tocher, D. A. *Acta Crystallogr.* **2004**, *E60*, o1786-o1787.

(178)   Barnett, S. A.; Hulme, A. T.; Tocher, D. A. *Acta Crystallogr.* **2006**, *C62*, o412-o415.

(179)   Kuhnert-Brandstatter, M. Thermomicroscopy of Organic Compounds; In *Comprehensive Analytical Chemistry Vol. XVI*; Svelha, G., ed. Elsevier Scientific Publishing: Amsterdam, 1982; pp 329-491.

(180)   Hamad, S.; Moon, C.; Catlow, C. R. A.; Hulme, A. T.; Price, S. L. *J.Phys.Chem.B* **2006**, *110*, 3323-3329.

(181)   Tassel, D.; Madoff, M. A. *J.Am.Med.Assoc.* **1968**, *206*, 830-832.

(182)   Vermes, A.; Guchelaar, H.-J.; Dankert, J. *J.Antimicrob.Chemother.* **2000**, *46*, 171-179.

(183)   Akimoto, M.; Miyahara, T.; Arai, J.; Akimoto, A.; Hamada, H.; Yoshida, Y.; Yoshimura, N. *Br.J.Ophthalmol.* **2002**, *86*, 581-586.

(184)   Hulme, A. T.; Tocher, D. A. *Cryst.Growth & Des.* **2005**, *6*, 481-487.

(185)   Louis, T.; Low, J. N.; Tollin, P. *Cryst.Struct.Comm.* **1982**, *11*, 1059.

(186)   Wang, S.; Schaefer III, H. F. *J.Chem.Phys.* **2006**, *124*, 044303.

(187)   Infantes, L. *Cryst.Eng.Comm.* **2002**, *4*, 454-461.

(188)   Hulme, A. T.; Tocher, D. A. *Acta Crystallogr.* **2005**, *E61*, o2112-o2113.

(189)   Barker, D. L.; Marsh, R. E. *Acta Crystallogr.* **1964**, *17*, 1581-1587.

(190) Sternglanz, H.; Bugg, C. E. *Biochim.Biophys.Acta* **1975**, *378*, 1-11.

(191) Okabe, N.; Tamaki, K.; Suga, T. *Acta Crystallogr.* **1995**, *C51*, 2696-2698.

(192) Mailkov, V. M.; Saidkhodzhaev, A. I. *Chem.Nat.Compd.* **1998**, *34*, 202-265.

(193) Booth, N. L.; Nikolic, M. S. D.; van Breemen, R. B.; Geller, S. E.; Banuvar, S.; Shulman, L. P.; Farnsworth, N. R. *Clin.Pharmacol.& Ther.* **2004**, *76*, 511-516.

(194) Morrison, H.; Curtis, H.; McDowell, T. *J.Am.Chem.Soc.* **1966**, *88*, 5415-5419.

(195) Schmidt, G. M. J. *J.Chem.Soc.* **1964**, 2014-2021.

(196) Schmidt, G. M. J. *Pure Appl.Chem.* **1971**, *27*, 647.

(197) Gnanaguru, K.; Ramasubbu, N.; Venkatesan, K.; Ramamurthy, V. *J.Org.Chem.* **1985**, *50*, 2337-2346.

(198) Ramasubbu, N.; Guru Row, T. N.; Venkatesan, K.; Ramamurthy, V.; Ramachandra Rao, C. N. *Chem.Comm.* **1982**, 179.

(199) Thomas, J. M. *Nature* **1981**, *289*, 633-634.

(200) Gao, X.; Friscic, T.; MacGillivray, L. R. *Angew.Chem.Int.Ed.* **2004**, *43*, 232-236.

(201) Cheung, E. Y.; Harris, K. D. M.; Kang, T.; Scheffer, J. R.; Trotter, J. *J.Am.Chem.Soc.* **2006**, *ASAP*.

(202) Bowes, K. F. Solid State Structural Studies on Photoactive Materials. 2005. University of Cambridge.

(203) Gaultier, J.; Hauw, C. *Acta Crystallogr.* **1966**, *20*, 646-651.

(204) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *Chem.Eur.J.* **1996**, *2*, 925-934.

(205) Baures, P. W.; Rush, J. R.; Schroeder, S. D.; Beatty, A. M. *Cryst.Growth & Des.* **2002**, *2*, 107-110.

(206) Ueno, K. *Acta Crystallogr.* **1985**, *C41*, 1786-1789.

(207) Gavuzzo, E.; Mazza, F.; Giglio, E. *Acta Crystallogr.* **1974**, *B30*, 1351-1357.

(208) Munshi, P.; Guru Row, T. N. *J.Phys.Chem.A* **2005**, *109*, 659-672.

(209) Hsieh, T.-J.; Su, C.-C.; Chen, C.-Y.; Liou, C.-H.; Lu, L.-H. *J.Mol.Struct.* **2005**, *741*, 193-199.

(210) Myasnikova, R. M.; Davydova, T. S.; Simonov, V. I. *Kristallografiya (Russ.)* **1973**, *18*, 720.

(211) Csoregh, I. *Chem.Commun.(University of Stockholm)* **1976**, *1*.

(212) De Schryver, F. C.; Put, J.; Leenders, L.; Loos, H. *J.Am.Chem.Soc.* **1974**, *96*, 6994-7000.

(213) Hunter, C. A.; Sanders, J. K. M. *J.Am.Chem.Soc.* **1990**, *112*, 5525-5534.

(214) Anschutz *Ann.* **1909**, *367*, 169.

(215) Stahmann, M. A.; Wolff, I.; Link, K. P. *J.Am.Chem.Soc.* **1943**, *65*, 2285-2287.

(216)   Janiak, C. *J.Chem.Soc.Dalton Trans.* **2000**, 3885-3896.

(217)   Bruker AXS Inc. GEMINI. (1.02). 1998. Madison, Wisconsin, USA.

(218)   Shirley, R. Accuracy in Powder Diffraction; In *NBS Spec. Publ. 567*; Block, S., Hubbard, C. R., eds. 1980; pp 361-382.

(219)   Hunter, B. Rietica - a visual Rietveld program. 1998. International Union of Crystallography. Commission on Powder Diffraction Newsletter No. 20.

(220)   MacGillivray, L. R. *Cryst.Eng.Comm.* **2002**, *4*, 37-41.

(221)   Sarma, J. A. R. P.; Desiraju, G. R. *Crystal Growth and Design* **2002**, *2*, 93-100.

(222)   Allen, F. H. *Acta Crystallogr.* **2002**, *B58*, 380-388.

(223)   Allen, F. H.; Kennard, O. *Chem.Des.Autom.News* **1993**, *8*, 31-37.

(224)   Howie, R. A.; Skakle, J. M. S. *Acta Crystallogr.* **2001**, *E57*, 822-823.

(225)   Threlfall, T. L. *Analyst* **1995**, *120*, 2435-2460.

(226)   Jetti, R. K. R.; Boese, R.; Sarma, J. A. R. P.; Reddy, L. S.; Vishweshwar, P.; Desiraju, G. R. *Angew.Chem.Int.Ed.* **2003**, *42*, 1963-1967.

(227)   Goodwin, W.; Perkin, W. H. *J.Chem.Soc.* **1905**, *87*, 841-855.

(228)   Hall Jr., H. K. *J.Am.Chem.Soc.* **1958**, *80*, 6412-6420.

(229)   Florence, A. J.; Johnston, A.; Fernandes, P.; Shankland, N.; Shankland, K. *J.Appl.Crystallogr.* **2006**, *39*, 922-924.

(230)   Florence, A. J.; Baumgartner, B.; Weston, C.; Shankland, N.; Kennedy, A. R.; Shankland, K.; David, W. I. F. *J.Pharm.Sci.* **2003**, *92*, 389-397.

(231)   Bruker AXS. DiffracPlus Eva. 2003.

(232)   Hulme, A. T.; Johnston, A.; Florence, A. J.; Tocher, D. A. *Acta Crystallogr.* **2006**, *E62*, o545-o547.

(233)   Boese, R.; Blaser, D.; Latz, R.; Baumen, A. *Acta Crystallogr.* **1999**, *C55*, IUC9900001.

(234)   Allan, D. R.; Clark, S. J. *Phys.Rev.B* **1999**, *60*, 6328-6334.

(235)   Hulme, A. T.; Fernandes, P.; Florence, A. J.; Johnston, A.; Shankland, K. *Acta Crystallogr.* **2006**, *E62*, o3752-o3754.

(236)   Timmermans, J. *J.Phys.Chem.Solids* **1961**, *18*, 1-8.

(237)   Hulme, A. T.; Fernandes, P.; Florence, A. J.; Johnston, A.; Shankland, K. *Acta Crystallogr.* **2006**, *E62*, o3046-o3048.

(238)   Ostwald, W. Z. *Z.Phys.Chem* **1897**, *22*, 289-330.

(239)   Lazlo *Rec.J.R.Neth.Chem.Soc.* **1965**, *84*, 251.

(240)   Mora, A. J.; Fitch, A. N. *Z.Kristallogr.* **1999**, *214*, 480-485.

(241) Dunning, W. *J.Phys.Chem.Solids* **1961**, *18*, 21-27.

(242) Fouret, R. Diffuse X-Ray Scattering by Orientationally-Disordered Crystals; In *The plastically crystalline state*; Sherwood, J. N., ed. John Wiley & Sons: Chichester, 1979; pp 85-121.

(243) Pauling, L. *Phys.Rev.* **1930**, *36*, 430-443.

(244) Frenkel, J. *Acta Phys.Chemica U.S.S.R.* **1935**, *3*, 23.

(245) Guthrie, G. B.; McCullough, J. P. *J.Phys.Chem.Solids* **1961**, *18*, 53-61.

(246) Kabo, G. J.; Blokhin, A. V.; Charapennikau, M. B.; Kabo, A. G.; Sevruk, V. M. *Thermochim.Acta* **2000**, *345*, 125-133.

(247) Dunning, W. The Crystal Structure of Some Plastic and Related Crystals; In *The plastically crystalline state*; Sherwood, J. N., ed. John Wiley & Sons: Chichester, 1979; pp 1-37.

(248) Parsons, S. 2006.

(249) Khankari, R. K.; Grant, D. J. W. *Thermochim.Acta* **1995**, *248*, 61-79.

(250) Finney, J. L. *Phil.Trans.R.Soc.Lond.B.* **2004**, *359*, 1145-1165.

(251) Gillon, A. L.; Feeder, N.; Davey, R. J.; Storey, R. *Cryst.Growth & Des.* **2003**, *3*, 663-673.

(252) Desiraju, G. R. *Chem.Comm.* **1991**, 426-428.

(253) van Eijck, B. P.; Kroon, J. *Acta Crystallogr.* **2000**, *B56*, 535-542.

(254) Kamb, B. *Acta Crystallogr.* **1964**, *17*, 1437-1449.

(255) Kamb, B.; Hamilton, W. C.; LaPlaca, S. J.; Prakash, A. *J.Chem.Phys.* **1971**, *55*, 1934-1945.

(256) Kuhs, W. F.; Finney, J. L.; Vettier, C.; Bliss, D. V. *J.Chem.Phys.* **1984**, *81*, 3612-3623.

(257) La Placa, S. J.; Hamilton, W. C.; Kamb, B.; Prakash, A. *J.Chem.Phys.* **1972**, *58*, 567-580.

(258) Jackson, S. M.; Nield, V. M.; Whitworth, R. W.; Oguro, M.; Wilson, C. C. *J.Phys.Chem.B* **1997**, *101*, 6142-6145.

(259) Leadbetter, A. J.; Ward, R. C.; Clark, J. W.; Tucker, P. A.; Matsuo, T.; Suga, H. *J.Chem.Phys.* **1985**, *82*, 424-428.

(260) Salzmann, C. G.; Radaelli, P. G.; Hallbrucker, A.; Mayer, E.; Finney, J. L. *Science* **2006**, *311*, 1758-1761.

(261) Bernal, J. D.; Fowler, R. H. *J.Chem.Phys.* **1933**, *1*, 515.

(262) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., ed. Dordrecht, 1981; p 331.

(263) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J.Chem.Phys.* **1983**, *79*, 926-935.

(264) Guillot, B. *J.Mol.Liq.* **2002**, *101*, 219-260.

(265) Millot, C.; Soetens, J.-C.; Martins Costa, M. T. C.; Hodges, M. P.; Stone, A. J. *J.Phys.Chem.A* **1998**, *102*, 754-770.

(266) Abascal, J. L. F.; Sanz, E.; Fernandez, R. G.; Vega, C. *J.Chem.Phys.* **2005**, *122*, 234511.

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

(268) Boulougouris, G. C.; Economou, I. G.; Theodorou, D. N. *J.Phys.Chem.B* **1998**, *102*, 1029-1035.

(269) Errington, J. R.; Panagiotopoulos, A. Z. *J.Phys.Chem.B* **1998**, *102*, 7470-7475.

(270) Beyer, T.; Price, S. L. *J.Phys.Chem.B* **2000**, *104*, 2647.

(271) Breneman, C. M.; Wiberg, K. H. *J.Comput.Chem.* **1990**, *11*, 361-373.

(272) Whalley, E. *Trans.Faraday Soc.* **1957**, *53*, 1578-1585.

(273) Karamertzanis, P. G.; Price, S. L. *J.Phys.Chem.B* **2005**, *109*, 17134-17150.

(274) Potter, B. S.; Palmer, R. A.; Withnall, R.; Chowdhry, B. Z.; Price, S. L. *J.Mol.Struct.* **1999**, *485-486*, 349-361.

(275) Taylor, R.; Kennard, O. *Acc.Chem.Res.* **1984**, *17*, 320-326.

(276) Allen, F. H.; Baalham, C. A.; Lommerse, J. P. M.; Raithby, P. R. *Acta Crystallogr.* **1998**, *B54*, 320-329.

(277) Karamertzanis, P. G.; Pantelides, C. C. *J.Comput.Chem.* **2005**, *26*, 304-324.

(278) Day, G. M.; Motherwell, W. D. S. *Cryst.Growth & Des.* **2006**, *6*, 1985-1990.