CRYSTAL STRUCTURE SOLUTION OF HYDROGEN BONDED SYSTEMS: A VALIDATION AND AN INVESTIGATION USING HISTORICAL METHODOLOGIES FOLLOWED BY A REVIEW OF CRYSTAL STRUCTURE PREDICTION METHODOLOGIES TO DATE.

Sarah Ann Murphy

Submitted for the degree of Master of Philosophy

Heriot-Watt University

School of Engineering and Physical Sciences

December 2017

ABSTRACT

There are many chemicals that crystallize into more than one form. This phenomenon is called polymorphism. In each form or polymorph, inter and intra-molecular binding differ to varying degrees. As a result of this structural variation, the physical properties of the solid phases may also differ. Even the smallest of changes at the molecular level can result in a significant change in the final adopted crystal structure. Polymorphism in crystal structures allows studies of structure-property relationships since it is only the packing motifs that differ between polymorphs.

In this thesis, a 'computationally assisted' approach to crystal structure solution was taken. X-ray powder diffraction was used to generate unit cell dimensions and space groups while historical in-house molecular modelling methods were used to generate possible trial structures that would be the starting point for refinement.

Finally, a review of the latest methodologies for crystal structure prediction and consideration of polymorphism within the pharmaceutical industry completes this work.

HERIOT WATT UNIVERSITY

ACADEMIC REGISTRY
Research Thesis Submission

Name*:*	Sarah Ann Murphy		
School/PGI:	School of Engineering and Physical Sciences		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought (Award **and** Subject area)	Master of Philosophy Chemical Engineering

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

*   *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate*:*		Date:	

Submission

Submitted By *(name in capitals):*	SARAH ANN MURPHY
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals):*		
Method of Submission (Handed in to SSC; posted through internal/external mail):		
*E-thesis Submitted (***mandatory for final theses)**		
Signature:		Date:	

TABLE OF CONTENTS

CHAPTER 1 – INTRODUCTION

1.1   Polymorphism and Structure Solution

"Polymorphism is the ability of any element or compound to crystallize as more than one distinct crystal species" [1.1]

There are many chemicals that crystallize into more than one form. In each form, or polymorph, inter and intra-molecular binding can differ to varying degrees [1.2]. As a result of this structural variation, the physical properties of the solid phases may also differ [1.3]. Consequently, the phenomenon that is polymorphism in crystal structures allows studies of structure-property relationships [1.4].

There are different categories of polymorphism. Conformational polymorphism is prevalent in "effect chemicals" such as pharmaceuticals since most pharmaceutical compounds contain flexible moieties in their molecular structure. For example, the antiepileptic drug Carbamazepine exists in four reported anhydrous forms alone. The monoclinic and trigonal forms have been examined in this thesis. Packing polymorphism occurs when the same conformational structure can be packed in more than one way in the crystalline state. Forms I and II of Acetaminophen being one such example [1.5, 1.6 and 1.7]. Concomitant polymorphism is when more than one polymorph crystallizes in the same batch. Whilst being an interesting phenomenon to research, it is obviously undesirable within pharmaceutical production [1.8].

Industrial processes including drying, filtration and powder compression need to be optimized for effect chemicals such as pigments and dyes, pharmaceuticals and agrochemicals, where reliable production of a particular polymorph under reproducible conditions is required [1.9]. Consequently, detailed knowledge of the different molecular conformations and subsequent packing motifs manifested by the polymorphs of a given material becomes highly desirable. Logically, it might be predicted that the most likely packing motif for a crystal will be the most thermodynamically stable. However, in some circumstances, this is not the case [1.10, 1.11].

Such structural information can be used as a basis for carrying out molecular modelling simulations when attempting to predict differences in the performance of polymorphs. Such differences include variations in crystal morphology, crystal purity and mechanical properties of the solids. Research has shown that there is considerable occurrence of multiple packing

motifs in organic molecular crystals [1.2]. Taking the example of pharmaceutical compounds, the bio-availability or toxicity of one polymorph will most likely differ from another, resulting in a mixture of polymorphs being potentially lethal, or at the other end of the scale, totally ineffective.

Since polymorphism is so unpredictable [1.12, 1.13, 1.14], every system will provide a novel problem. Without necessary experimentation, there is no way of knowing how many different forms exist or what the processing effects on each form might be [1.15, 1.16, 1.17, 1.9 and 1.18]. This could lead to expensive delays in production, even starting the process all over again.

Solid-form screening during the initial stages of pharmaceutical design has become increasingly important. However unless sufficient time is given to this stage of the process, it is still likely that despite the numerous techniques available, the solid-form screen will not find some of the most useful polymorphic forms until the later stages of development [1.19, 1.20, 1.21].

High through-put screening methods have been widely used in order to expose the compound under investigation to hundreds of different experimental conditions. Whilst this 'trial-and-error' approach considered the variations of crystallization processing conditions, it could not take into account the complexities of the individual molecule. Particularly within pharmaceutical compounds, hydrogen bonding can have a large effect on the final structure [1.22] since even weak hydrogen bonds maintain a directional preference [1.23]. Consequently, 'bespoke' manual methods of screening should be used to complement the automated techniques.

A great deal of interest has been shown in the study of crystals comprising more than one chemically distinct molecule in the molecular unit. Known as cocrystals, they have been seen to display enhanced properties over the individually formulated active pharmaceutical ingredients (API's) [1.24]. It is common to use an API teamed with a 'biologically passive structure-directing component'. This is not the only advantage of utilizing co-crystals. Another attractive feature is that enhancing chemical properties can be brought about without the inconvenience of making or breaking any covalent bonds [1.25, 1.26 and 1.27].

One long-term goal of the scientific community is to be able to predict crystal structures *ab initio* from knowledge of the molecular structure [1.11, 1.28, 1.29 and 1.30]. A lesser but still considerable problem is to develop X-ray powder data into a routine method in the

laboratory. The main problem with this method of structure solution is the generation of reasonable starting structures to refine [1.31].

There have been many methods for structure determination proposed, especially in the latter half of the 20th century aided by all the advances and improvements made in electronics and mechanics.

Since single crystal analysis is now a routine method along with the advances in computer power and experimental techniques, the number of solved crystal structures has risen dramatically. In the mid-1960s, in an attempt to provide a bibliographic collection of chemical and structural data obtained from X-ray or neutron diffraction, the Cambridge Structural Database (CSD) was developed [1.32]. It contains organic and organometallic compounds, each one having a specific reference code. There is another database called the Protein Data Bank (PDB) [1.33], which as the name suggests contains biological macromolecular structures. This is managed by the Research Collaboratory for Structural Bioinformatics (RCSB). A third database called Isostar [1.34(a)] contains information on the intermolecular non-bonded interactions of structures. It has used data from the CSD and the PDB combined with theoretical energy calculations using the InterMolecular Perturbation Theory (IMPT) [1.34(b)].

A popular and reliable technique for unit cell and space group determination is X-ray diffraction. As previously mentioned, it is now considered routine to analyze structures using single crystal diffraction [1.35]. However this method is limited to compounds where crystals of sufficient size and quality may be grown. X-ray powder diffraction can be used as an alternative, although its application is not straightforward [1.36-1.39].
Synchrotron radiation sources have made vast improvements in the quality of collectable powder data [1.40, 1.41]. The advantages of Synchrotron radiation over conventional sources is that of high intensity, tuneable wavelength and a high degree of polarization. Accurate peak positions are produced, which make it ideal for indexing and refinements.
X-ray powder diffraction coupled with molecular modelling techniques has been applied to determine the structure of many small inorganic compounds [1.42, 1.43] and, as time passes, the structures of more and more organic molecules and metallic clusters of increasing size and complexity are being solved [1.44, 1.45].

Since 1999, the CCDC (Cambridge Crystallographic Data Centre) has arranged a series of problems designed to test the strengths of the variety of techniques available to the field of CSP. Research groups use their preferred methods to predict the crystal structures of a

number of 'unknown' structures from the CSD. The structures range from small rigid structures to large flexible molecules and structures with more than one molecule in the asymmetric unit. The methods and results are discussed and progress of methodologies within the community is assessed.

In this research, a 'computationally assisted' approach to crystal structure solution was taken. X-ray powder diffraction was used to generate unit cell dimensions and space groups while molecular modelling methods were used to generate possible trial structures that would be the starting point for refinement.

Using this combination of experimental and theoretical techniques, the following work was carried out:

- Validation of an in-house computational search method involving a systematic search of the unit cell. Results are ranked based on final lattice energy and comparison of the generated and experimental X-ray powder diffraction patterns. A comparison of this method with a simulated annealing 'random search' method was also carried out. Two examples of polymorphic pharmaceuticals were used for this purpose, Carbamazepine (monoclinic and trigonal) and the two published structures of Primidone.

- The simulated annealing method was tested further to solve the crystal structure of the co-crystal of the dihydroxybenzene catechol and urea.

1.2   2016 Review

Since this research was completed, the stages in Crystal Structure Solution (CSS) from Powder Diffraction Data (PDD) haven't changed significantly. Ultimately the stages [1.3] still consist of;

1) Conformational analysis and selection of Degrees of Freedom (DOF). A sensible starting molecular conformation is advantageous. Decisions about degrees of freedom can be made by looking to the CSD for similar structures/structural fragments. Alternatively, the changes in energy associated with changes in conformation can be calculated. It makes sense to allow a small number of degrees of freedom to be treated in the search and to have an idea as to the range of values they are likely to take. Whilst keeping computing time to a minimum is desirable, it is also essential not to discard potential conformations at the outset.

2) Selection of the theoretical set which will describe the intramolecular energy, the electrostatic intermolecular energy and selection of the potential function to describe intermolecular repulsion/dispersion forces. The development of potentials for dispersion forces has been considerable and have had significant effects in the quality of structures generated to take into the optimisation stage.

3) The global search of the unit cell which will generally find numerous pockets of low-energy regions with similar lattice energy values. In modern stochastic searches, functions have been implemented to avoid losing the molecule in a local minima or basin. The practitioner does well to optimise the balance between the amount of search space investigated and the length of computational time spent searching. Current methods can be divided into four categories;

   a. Deterministic Methods – given the same input, the same output will be generated.

   b. Stochastic Methods – given the same input, the output will be unpredictable.

   c. Heuristic Methods – a method which might not give a perfect result, but one that will provide suitable results for the current purposes.

   d. Smoothing Methods – a statistical method to remove irregularities to improve accuracy.

4) Local Lattice-Energy Minimization of the lowest energy/highest ranking structures generated within the previous stage. The molecular conformation may be treated as rigid in this step, however greater accuracy has been seen when further degrees of freedom, such as torsions and bond angles, have been considered.

Therefore, in terms of organic compounds, the advances in CSS have been changes in both computational structure prediction methods and experimental diffraction methods.

Complementary analytical techniques used in combination with crystal structure prediction methods, such as polymorph screening and NMR [1.46], have been shown to obtain repeatable and reliable results. Routine crystal structure solution of molecular crystals remains a formidable problem however [1.47]. The complexity of the molecular structure, the quality of the powder data and the number of possible polymorphs are just a few key factors which can hinder an analyst's progress to finding the global minimum of the energy landscape [1.48]. It is still hoped that when crystal structure prediction methods are reliable enough, only the chemical diagram should be required as input data. In this way, the design of new molecules crystallizing with favourable structural and physical properties may be computationally led prior to any laboratory work taking place. If the crystal structure of a molecule is already known, crystal structure prediction methods may also be used to predict the likelihood of as-yet unknown polymorphs being formed which have undesirable physical properties.

1.3   References

[1.1]   J. Haleblian, *Pharmaceutical Applications of Polymorphism*, Journal of Pharmaceutical Sciences, **58**(8), 911-929 (1969)

[1.2]   T.L. Threlfall, *Analysis of Organic Polymorphs,* Analyst, **120**, 2435-2460 (1995)

[1.3]   M. Vasileiadis, *The Polymorphs of ROY: Application of a Systematic Crystal Structure Prediction Technique,* Acta Cryst., B68, 677-685 (2012)

[1.4]   J. Bernstein, *Crystal Growth, Polymorphism and Structure-Property Relationships in Organic Crystals*, J. Phys. D: Appl. Phys., **26**, B66-B76 (1993)

[1.5] (a) Chem Image Application Note, *Discrimination of Acetaminophen Polymorphs Using Raman Chemical Imaging,* http://www.chemimage.com/docs/application-notes/Pharmaceutical/CI-appnote-Polymorphs.pdf; (b) Application Brief *Polymorphism in Acetaminophen Studied by Simultaneous DSC and Raman Spectroscopy,* http://www.perkinelmer.co.uk/lab-solutions/resources/docs/APP_PolymorphisminAcetaminophenbyDSC.pdf

[1.6] J. Espeau, *Polymorphism of Paracetamol: Relative Stabilities of the Monoclinic and Orthorhombic Phases Inferred from Topological Pressure-Temperature and Temperature-Volume Phase Diagrams*, J. Pharm. Sci., **94**(3), 524-539 (2005)

[1.7] G. Di Profio, *Controlling Polymorphism with Membrane-Based Crystallizers: Application to Form I and II of Paracetamol,* Chem. Mater., **19**, 2386-2388 (2007)

[1.8] J. Bernstein, *Concomitant Polymorphs,* Angew. Chem. Int. Ed., **38**, 3440-3461 (1999)

[1.9]   (a) J.D. Dunitz, *Disappearing Polymorphs*, Acc. Chem. Res., **28**, 193-200 (1995); (b) M. Perrin, *Crystal Structure Determination of the Elusive Paracetamol Form III,* Chem. Comm., 3181-3183 (2009)

[1.10]   T. Beyer, *Which Organic Crystal Structures are Predictable by Lattice Energy Minimisation?*, CrystEngComm, **44**, 1-35 (2001)

[1.11]   J.D. Dunitz, *Are Crystal Structures Predictable?*, Chem. Commun., 545-548 (2003)

[1.12]   (a) A.J. Cruz-Cabeza, *Facts and Fictions about Polymorphism,* Chem. Soc. Rev., **44** 8619-8635 (2015); (b) A. D. Bond, *Polymorphism in Molecular Crystals,* Curr. Op. Solid State Mat. Sci., **13** 91-97 (2009)

[1.13]   E. H. Lee, *A Practical Guide to Pharmaceutical Polymorph Screening and Selection,* Asian Journal of Pharmaceutical Sciences, **9**, 163-175 (2014)

[1.14]   A. Nangia, *Conformational Polymorphism in Organic Crystals,* Acc. Chem. Res., **41**(5), 595-604 (2008)

[1.15]   (a) L. Yu, *Survival of the Fittest Polymorph: How Fast Nucleater can Lose to Fast Grower,* Cryst. Eng. Comm., **9**, 847-851 (2007); (b) S. Chen, *Cross Nucleation Between ROY Polymorphs,* J. Am. Chem. Soc., **127**, 17439-17444 (2005)

[1.16]   (a) A. D. Bond, *On the Polymorphism of Aspirin,* Angew. Chem. Int. Ed., **46**, 615-617 (2007); (b) John Wiley & Sons, Inc. *How Does Aspirin Crystallize?*, Science Daily, 23 December 2006; (c) C. Ouvrard, *Toward Crystal Structure Prediction for Conformationally Flexible Molecules: The Headaches Illustrated by Aspirin,* Cryst. Growth and Des., **4**(6) 1119-1127 (2004)

[1.17]   Z. Nėmet, *Polymorph Transitions of Bicalutamide: A Remarkable Example of Mechanical Activation,* J. Pharm. Sci., **97**(8), 3222-3232 (2008)

[1.18]   J. Bauer, *Ritonavir: An Extraordinary Example of Conformation Polymorphism,* Pharm. Res., **18**(6), 859-866 (2001)

[1.19]   A. M. Campeta, *Development of a Targeted Polymorph Screening Approach for a Complex Polymorphic and Highly Solvating API,* J. Pharm. Sci., **99**(9) 3874-3886 (2010)

[1.20]   S. Price, *The Potential of Computed Crystal Energy Landscapes to Aid Solid-Form Development,* Drug Discovery Today, **21**(6), 912-923 (2016)

[1.21]   S. L. Morissette, *High-throughput Crystallization: Polymorphs, Salts, Co-crystals and Solvates of Pharmaceutical Solids,* Adv. Drug Del. Rev., **56**, 275-300 (2004)

[1.22]   P. T. A. Galek, *Persistent Hydrogen Bonding in Polymorphic Crystal Structures,* Acta Cryst., **B65**, 68-85 (2009)

[1.23]   I. A. Finneran, *Hydrogen Bonding in the Ethanol-Water Dimer,* Phys. Chem. Chem. Phys., **17**, 24210-24214 (2015)

[1.24]   Ö. Almarsson, *Crystal Engineering of the Composition of Pharmaceutical Phases. Do Pharmaceutical Co-crystals Represent a New Path to Improved Medicines?*, Chem. Commun., 1889-1896 (2004)

[1.25]   A.D. Bond, *What is a Co-crystal?*, CrystEngComm, **9**, 833-834 (2007)

[1.26] P. Vishweshwar, *Pharmaceutical Co-crystals*, Journal of Pharmaceutical Sciences, **95**(3), 499-516 (2006)

[1.27] G.P. Stahly, *Diversity in Single- and Multiple-Component Crystals. The Search for and Prevalence of Polymorphs and Cocrystals*, Crystal Growth and Design, **7**(6), 1007-1026 (2007)

[1.28] J. Maddox, *Crystals From First Principles*, Nature, **335**, 201 (1988)

[1.29] A. Gavezzotti, *Are Crystal Structures Predictable?*, Acc. Chem. Res., **27**, 309-314 (1994)

[1.30] H.R. Karfunkel, *Ab Initio Prediction of Possible Crystal Structures for General Organic Molecules*, Journal of Computational Chemistry, **13**(10), 1171-1183 (1992)

[1.31] R.B.Hammond, *Computationally Assisted Structure Determination for Molecular Materials from X-ray Powder Diffraction Data*, J. Phys. Chem. B., **101**, 6532-6536 (1997)

[1.32] F.H. Allen, *The Cambridge Crystallographic Data Centre: Computer-Based Search, Retrieval, Analysis and Display of Information*, Acta Cryst., **B35**(10), 2331-2339 (1979)

[1.33] H. M. Berman, *The Protein Data Bank,* Nucl. Acids. Res., **28**(1), 235-242 (2000); www.rcsb.org

[1.34] (a) I.J. Bruno, *IsoStar: a Library of Information About Nonbonded Interactions,* J. Comput. Aided Mol. Des., **11**, 525-537 (1997) (b) A.J. Stone, *Intermolecular Perturbation Theory for Van der Waals Molecules,* Faraday Discuss. Chem. Soc., **73**, 19-31 (1982)

[1.35] P.R. Rudolf, *Techniques for Ab Initio Structure Determination from X-ray Powder Diffraction data*, Materials Chemistry and Physics, **35**, 267-272 (1993)

[1.36] D. Louër, *Some recent Developments and Applications of Powder Diffraction*, Chemica Scripta, **28**, 89-95 (1988)

[1.37] A. Clearfield, *Crystal Structure Determination from X-ray Powder Diffraction Data*, The Rigaku Journal, **11**(1), 1-3 (1994)

[1.38] K.M. Harris, *Crystal Structure Determination from Powder Diffraction Data*, Chem. Mater., **8**(11), 2554-2570 (1996)

[1.39] P. Lightfoot, *Determination of a Molecular Crystal Structure by X-ray Powder Diffraction on a Conventional Laboratory Instrument*, J. Chem. Soc., Chem. Commun., 1012-1013 (1992)

[1.40] A.N. Fitch, *High Resolution Powder Diffraction Studies of Polycrystalline Materials*, Nuclear Instruments and Methods in Physics Research B, **97**, 63-69 (1995)

[1.41] P. Barnes, *Synchrotron Radiation for Materials Science Research*, Analytical Techniques, 708-715 (1990)

[1.42] A. Le Bail, *Structure Determination of $NaPbFe_2F_9$ by X-ray Powder Diffraction*, Journal of Solid State Chemistry, **83**, 267-271 (1989)

[1.43] A. Clearfield, *Crystal Structures from Powder Data. 1. Crystal Structure of ZrKH(PO₄)₂*, Inorg. Chem., **23**, 4679-4682 (1984)

[1.44] E.D.L. Smith, *The Determination of the Crystal Structure of Anhydrous Theophylline by X-ray Powder Diffraction with a Systematic Search Algorithm, Lattice Energy Calculations, and ^{13}C and ^{15}N Solid-State NMR: A Question of Polymorphism in a Given Unit Cell*, J. Phys. Chem. B, **105**, 5818-5826 (2001)

[1.45] S.J. Hibble, *The Synthesis and Structure Determination from Powder Diffraction data of LaMo₅O₈, a New Oxymolybdate Containing Mo₁₀ Clusters*, J. Am. Chem. Soc., **110**, 3295-3296 (1988)

[1.46] F. G. Vogt, *Solid-State NMR Analysis of Organic Cocrystals and Complexes,* Cryst. Growth Des., **9**(2), 921-937 (2009)

[1.47] A. R. Oganov, *Introduction: Crystal Structure Prediction, a Formidable Problem,* Modern Methods of Crystal Structure Prediction, A. R. Oganov (ed), xi-xxi, (2011)

[1.48] W. I. F. David, *Structure Determination from Powder Diffraction Data,* Acta Cryst., **A64** 52-64 (2007)

A.   APPENDIX

A.1   Principles of X-ray Diffraction

X-rays were first discovered by Röntgen in 1895 and it wasn't until 1912 in Münich that Laue, Friedrich and Knipping performed diffraction experiments. However, it was Bragg in 1913 who developed the method of X-ray diffraction and realised its potential as a technique for crystal structure solution. Laue had found that it was possible to treat crystals as a three-dimensional diffraction grating where the atoms were arranged in large clusters. However, the real breakthrough came from W.L. Bragg when the father and son team considered crystals as layers or planes of atoms. The Braggs discovered that diffracted X-ray beams were produced when the path differences between reflections from successive planes were equivalent to an integral number of wavelengths. This observation gave rise to the well-known Bragg equation, equation (1), which is illustrated in figure 1.

$$n\lambda = 2d_{hkl}\sin\theta \qquad\qquad (A.1)$$
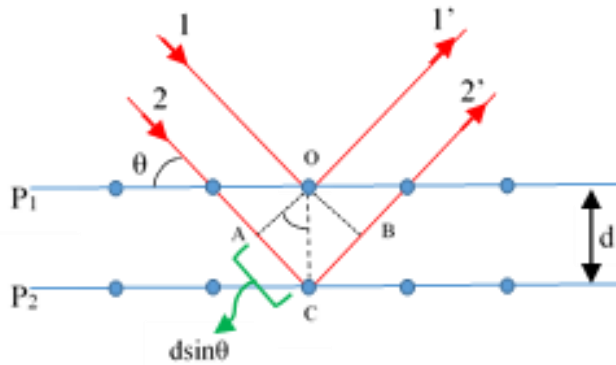
Figure A.1; X-ray beam incident on a pair of parallel planes P_1 and P_2 with interplanar spacing d. The parallel incident rays 1 and 2 make an angle of θ with these planes. Electrons positioned at O and C are forced to vibrate and radiate in all different directions. Where parallel secondary rays 1' and 2' radiate out from the planes at angle θ, a diffracted beam of maximum intensity is produced assuming the rays are in phase.

The perpendiculars dropped from O to A and B reveal that the angles created AOC = BOC = θ and that AC = BC. Waves in ray 1' will be in phase with waves in ray 2' if AC + CB (or 2AC) is an integral number of wavelengths λ, giving rise to the expression 2AC = nλ where 'n' is an integer. Since AC/d = sinθ, we arrive at Bragg's Law.

X-rays lie in the region of the electromagnetic spectrum between UV and γ radiation 0.1 (lower γ) to 100Å (upper far UV) (see figure 2). X-rays for crystal structure analysis lie in the region of 0.5 to 2.5Å.
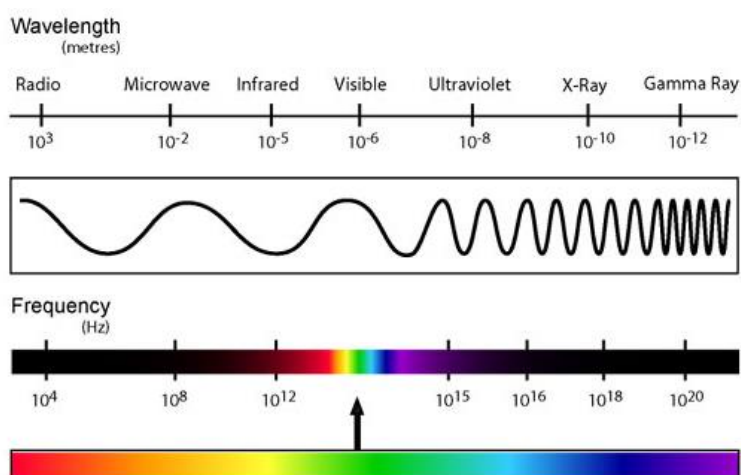
Figure A.2; the electromagnetic spectrum.

X-rays are produced when electrons of a source material are accelerated towards a target material as shown in figure 3. Multiple collisions of the accelerated electrons with the target material causes reductions in the electron velocity but does not bring them to a standstill. The loss of energy produced from these multiple collisions is emitted as a continuum of X-ray radiation.
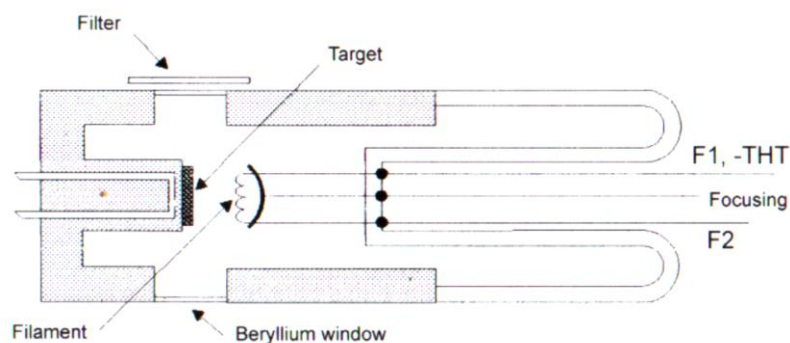
Figure A.3; accelerated electrons bombarding a target.

The λ_{min}(Å), or the maximum energy, of the radiation is determined using a combination of physical constants including Plank's constant and the speed of light. The relationship

between these constants, the charge on the electrons and the accelerating voltage is shown in equation (2).

$$E_{max} = eV_{acc} = \frac{hc}{\lambda_{min}} \tag{A.2i}$$

$$\lambda_{min} = \frac{12398}{V_{acc}} \tag{A.2ii}$$

Where h (Plank's constant) = $6.626\,070\,040 \times 10^{-34}\ m^2\ kg/s$

c (the speed of light) = $299\,792\,458\ m/s$

e is the charge of an electron = $1.602 \times 10^{-19}\ Coulomb$

and V is the accelerating voltage, measured in electron-volts.

The efficiency (η) of X-ray production can be calculated using equation (3) where Z is the atomic number and V is the voltage.

$$\eta = \frac{E_{photon}}{E_{electron}} = 1.1ZV10^{-9} \tag{A.3}$$

Since absorption of X-rays will occur the actual intensity of the radiation used can be calculated using;

$$I(\lambda) = I_0(\lambda)\exp^{(-\mu\rho x)} \tag{A.4}$$

$I(\lambda)$, $I_0(\lambda)$ are the resultant and incident radiation respectively, x is the width (depth) of sample in cm, μ is the mass absorption coefficient in cm^2g^{-1} and ρ is the density of material in gcm^{-3}

The absorption of X-rays (μ) is dependent on two factors, photoelectric absorption (τ) and scattering effects (σ) and can be considered as the stopping power of the material.

A.2   Crystal Lattices and Crystal Systems

Crystals are solid chemical substances with a three-dimensional periodic array of atoms, ions or molecules. This array is called a crystal structure. When the substances are defined as single points, a 'point lattice', figure 4, is formed. This can then be built up from a two dimensional form to a three dimensional structure. The vectors describing the sides of this basic unit are chosen to be right handed (the right thumb runs along a, first finger along b and middle along c). The crystallographic axes (also known as the base vectors) a, b and c are assigned in the same manner. The angles between these base vectors are called α, β and γ where $\alpha=\{b,c\}$, $\beta=\{a,c\}$ and $\gamma=\{a,b\}$. The basic three-dimensional lattice known as the 'unit cell' therefore, is the three dimensional array of the smallest repeating unit of the substance. Figure 5 shows the three-dimensional lattice with a unit cell showing how the cell edges and angles are related.

(a)   (b)   (c)
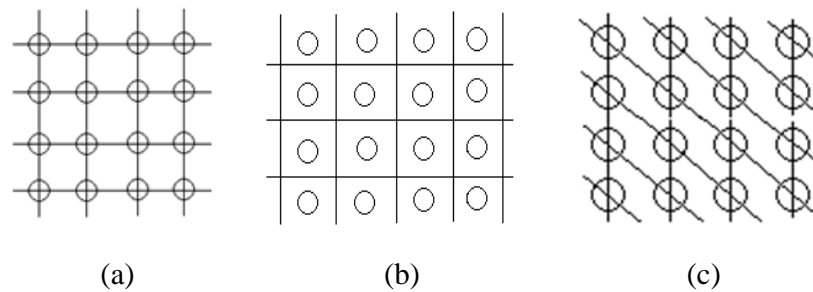
Figure A.4; regular 2 dimensional array - the point lattice. (a), (b) and (c) show how the points can be fixed within the two-dimensional lattice structure.
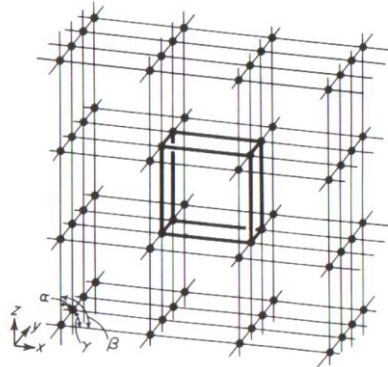
Figure A.5; the three-dimensional array – the unit cell is highlighted in bold.

There are seven crystal systems that are defined using the relationships of the unit cell parameters, these are listed in table 1.

Crystal System	Lattice Parameters	Degrees of Freedom
Triclinic	$a \neq b \neq c$; $\alpha \neq \beta \neq \gamma$	6
Monoclinic	$a \neq b \neq c$; $\alpha = \beta = 90°$, $\gamma \neq 90°$	4
Orthorhombic	$a \neq b \neq c$; $\alpha = \beta = \gamma = 90°$	3
Tetragonal	$a = b \neq c$; $\alpha = \beta = \gamma = 90°$	2
Trigonal (rhombohedral)	$a = b = c$; $\alpha = \beta = 90°$, $\gamma = 120°$	2
Hexagonal	$a = b \neq c$; $\alpha = \beta = 90°$, $\gamma = 120°$	2
Cubic	$a = b = c$; $\alpha = \beta = \gamma = 90°$	1

Table A.1; the seven crystal systems, the lattice parameter conditions and the resulting degrees of freedom.

There are also different ways atoms or molecules may be arranged within the unit cell. These factors have been classified too and are shown in table 2.

Cell	Symbol	Arrangement
Primitive	P	Atoms only at corners of cell
Body-centred	I	One atom in the centre of the cell
Face-centred	F	An atom at the centre of each face
Base-centred	C	Atom at centre of each face on a pair of opposite faces
Rhombohedral	R	Special case for primitive trigonal cell

Table A.2; the various atomic arrangements occurring in the unit cell, the cell nomenclature and identifying symbols.

In 1866, Bravais devised a classification system for every possible type of lattice. The Bravais classifications are shown in table 3 and figure 6.

Crystal System	Type of Centring Possible
Triclinic	Primitive
Monoclinic	Primitive, base centred
Orthorhombic	Primitive, base centred, face centred, body centred
Tetragonal	Primitive, body centred
Trigonal (rhombohedral)	Primitive
Hexagonal	Primitive
Cubic	Primitive, face centred, body centred

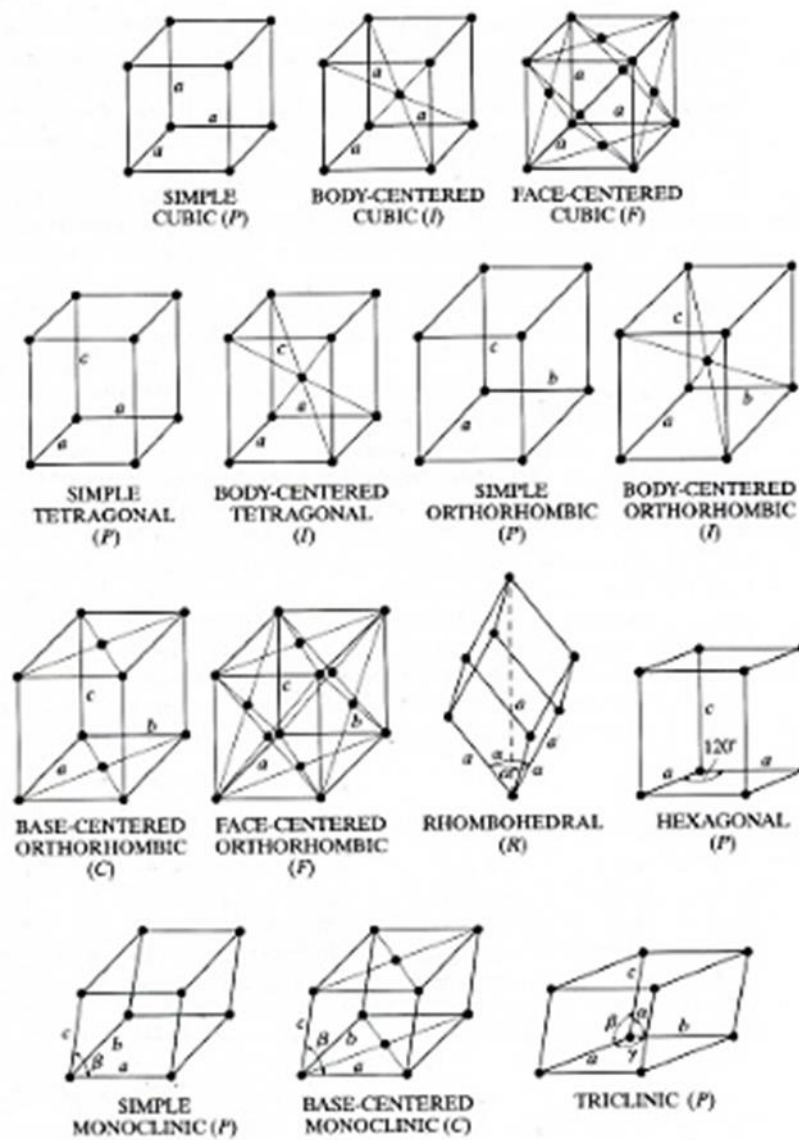**Table A.3;** the 14 Bravais lattice classifications.

Figure A.6; the 14 Bravais lattices: Image clipped from;

https://www.slideshare.net/praying1/972-b3102005-cullity-chapter-2/17

A.3　Symmetry Operators and Space Groups

Molecules in the unit cell are related to each other by functions called symmetry operators. In other words, the molecules can be superimposed onto each other using a number of different operations such as rotations or inversions. The point about which this operation occurs is called the symmetry element. The group of functions, excluding any translation operations, is called the point group. In three dimensions, there are 32 different point groups, known as crystal classes, which are divided up between the seven crystal systems introduced above. Each point group is described using symbols such as a rotation axis of order x, an inverse rotation axis of order $^-$, or a mirror plane m. In summary, the crystallographic point group is the operation or operations that will leave at least one point of the molecule unmoved and does not change the appearance of the crystal structure. Space groups also define symmetry in crystal structures. This group of operations not only includes the symmetry operations (reflection and rotation) of the point group, but also translational symmetry, which allow the repetition of the unit cell ultimately creating the infinite periodic system. There are three types of translational symmetry; lattice translations, screw axes and glide planes. Lattice translations allow movement of the repeat unit along a cell direction, screw axes involve a rotation of order 360/x immediately followed by a translation of n/x. Finally, a reflection in a plane followed by a translation parallel to the plane of half the repeat distance defines a glide plane. There are 230 different space groups defining all possible crystal symmetries.

The unit cell is the smallest repeating unit of the crystal structure and may contain one or more molecules. The total number of molecules in the unit cell is defined as Z. The asymmetric unit (Z') is the smallest subset Z which replicates into a complete unit cell following the application of symmetry operators. The space group defines the position and relationship of the molecule or molecules in the unit cell.

A.4　Miller Indices and the Reciprocal Lattice

It can also useful to describe sets of lattice planes as well as lattice points in unit cells. Series of parallel lattice planes are constructed such that every lattice point lies on one plane in the series. The edge is always cut up into an integral number of common fractions. The reciprocals of these factions become a triplet of numbers expressing the sets of possible planes. For example if a plane cuts the x, y and z-axes ½, ½, 1, the triple of reciprocal numbers becomes (221). These representations of a lattice planes are called Miller Indices, some examples of which are shown in figure 7.
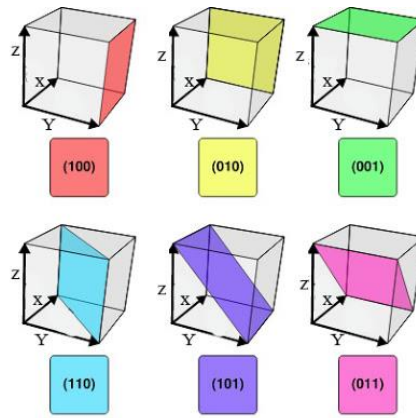
Figure A.7; Miller Indices, so named after Miller who devised the nomenclature in 1839. Image taken from;

http://www.smartway2study.com/2016/03/salient-features-of-miller-indices-of.html

So far, the lattice has been considered in direct space. In the Bragg equation sinθ is a measure of the deviation of 1' and is inversely proportional to the lattice interplanar spacing d.

$$\sin\theta = n\lambda/2(1/d) \tag{A.5}$$

Therefore, structures with large d values will have compressed diffraction patterns and vice versa.

By constructing the reciprocal lattice based on 1/d, the inverse relationship between sinθ and d becomes a direct one, thus facilitating the interpretation of X-ray diffraction patterns.

Generally, the reciprocal lattice parameters; reciprocal base vectors A*, B* and C* and reciprocal interaxial angles α*, β* and γ* are sometimes marked with an asterisk and are calculated using the direct lattice parameters and the unit cell volume as shown in equations (6), (7a), (7b) and (7c).

$$A* = \frac{b \wedge c}{V} \qquad B* = \frac{c \wedge a}{V} \qquad C* = \frac{a \wedge b}{V} \tag{A.6}$$

$$\alpha* = \vec{b}* \wedge \vec{c}*; \qquad \cos\alpha* = \frac{\cos\beta\cos\gamma - \cos\alpha}{\sin\beta\sin\gamma} \tag{A.7a}$$

$$\beta* = \vec{a}* \wedge \vec{c}*; \qquad \cos\beta* = \frac{\cos\alpha\cos\gamma - \cos\beta}{\sin\alpha\sin\gamma} \tag{A.7b}$$

$$\gamma* = \vec{a}* \wedge \vec{b}*; \qquad \cos\gamma* = \frac{\cos\alpha\cos\beta - \cos\gamma}{\sin\alpha\sin\beta} \tag{A.7c}$$

The reciprocal lattice can be considered as normals to all direct lattice planes (h,k,l) radiating from one lattice point which is designated as the origin. These normals all end at a distance $1/d_{hkl}$ from the appointed origin and d_{hkl} is the perpendicular interplanar distance between the set (hkl). These normal points make the reciprocal lattice. Figure 8a shows a representation of a reciprocal lattice where the asterisks are the reciprocal points and the integers are the direct indices of the relevant planes. Like the direct lattice, the reciprocal lattice can also be built up from two dimensions to three and this is shown in figure 8b. Since the reciprocal points are derived from the planes in the direct lattice, these values correspond to the Miller indices.
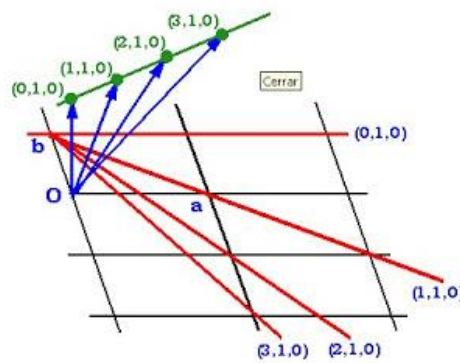
Figure A.8(a); 2D representation of points on a reciprocal lattice (green points). Miller indices are shown in blue. Image taken from;

http://orlaningsolidstate.blogspot.co.uk/2010/03/reciprocal-lattice.html

A.5 Atomic Scattering

Incoming X-rays are scattered by electrons of atoms within the unit cell. An atomic scattering factor '*f*', can be calculated using equation (8a) since only electrons are involved in the scattering of X-rays.

$$atomic\ scattering\ factor\ f = \frac{amplitude\ scattered\ by\ an\ atom}{amplitude\ scattered\ by\ a\ single\ electron} \qquad (A.8a)$$

A structure factor is represented by F_{hkl} and it expresses both the amplitude and the phase of a diffracted beam in other words, how the material scatters incoming radiation.
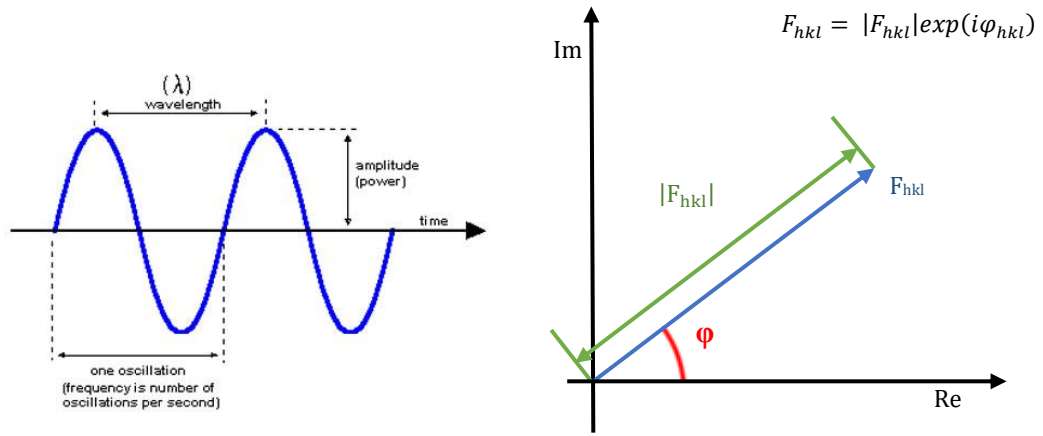
Figure A.9; (left) the characteristics of a wave. Image taken from;
http://www.sunapeeschools.org/smhs/staff-directory/mr-baker/physical-science/waves
and (right) the magnitude (green) and phase (red) of a structure factor (blue).

The amplitude of the scattered wave is given by the magnitude of the structure factor expressed as a ratio;

$$|F_{hkl}| = \frac{amplitude\ of\ the\ wave\ scattered\ by\ all\ atoms}{amplitude\ of\ the\ wave\ scattered\ by\ one\ atom} \tag{A.9a}$$

The structure factor is also directly related to the intensity of a reflection.

$$|F_{hkl}|^2 = the\ intensity\ of\ a\ reflected\ wave \tag{A.9b}$$

For any given structure with atomic positions at uvw, the structure factor is also used to determine which reflections to expect in a diffraction pattern.

$$F_{hkl} = \sum_{i=1}^{N} f_i\ e^{2\pi i(hu_i+kv_i+lw_i)} \tag{A.9c}$$

The structure factor is also required to estimate electron density (ρ) at a point in a unit cell (xyz) which has a volume of V and a phase α_{hkl}.

$$\rho_{(xyz)} = \frac{1}{V}\sum |F_{hkl}|exp^{(i\alpha_{hkl})}exp^{(-2\pi ihx+ky+lz)} \tag{A9.d}$$

CHAPTER 2 – METHODOLOGIES

2.1   Introduction

The procedure for determining crystal structures from materials presented in a powdered form consists of numerous steps [2.1, 2.2]:

- Collection of good quality X-ray diffraction data. The theory of X-ray diffraction and basic crystallography is not discussed within this chapter but is included in this thesis in Appendix 1. Suffice to say that depending on the types and the positions of the atoms in the crystal structure, X-rays will be diffracted in specific directions.

- Indexing the peak positions of the X-ray powder pattern to generate unit cell parameters and to identify the space group. This reveals the size and shape of the smallest repeating unit of the lattice and how the lattice can be displaced without being changed.

- The three-dimensional picture of the electron density can be generated by measuring the angles and intensities of the diffracted beams. This electron density 'picture' may be used to calculate the positions of the atoms and the bonds between them. Phase definition is required for this calculation however, which is information that is lost in the powder diffraction experiment. Values for phases are usually calculated by empirical methods where prior knowledge of electron density or structure is assumed. Methods which have been used to produce initial estimates of phases are;

  - *Ab Initio* Phasing or Direct Methods
  - Molecular Replacement
  - Heavy Atom Methods

- Trial structure generation in the unit cell using the estimated phases and the space group information. This chapter reviews the modelling methods employed during the period of this research. A modern perspective looking at modifications of the original methodologies is also included to bring this overview up to date.

- Refinement of best trial structures generated.

2.2 X-ray Diffraction in Crystal Structure Prediction

2.2.1 *Sample Preparation*

Samples used in X-ray powder diffraction should be ground until they are a very fine powder in order to attain a reasonably constant particle size. The ground sample is then prepared for analysis by tightly packing and sealing it into a glass capillary tube. The smaller the particle size, the more of the sample will be exposed to the X-rays and the more representative the result. Powder diffraction is based on a truly random distribution of crystallites of equal size. Deviating from this random distribution will have an effect on the measured intensities of the peaks in the diffraction pattern. This effect of non-randomness is called 'preferred orientation' [2.3] the effects of which are shown in figure 1.
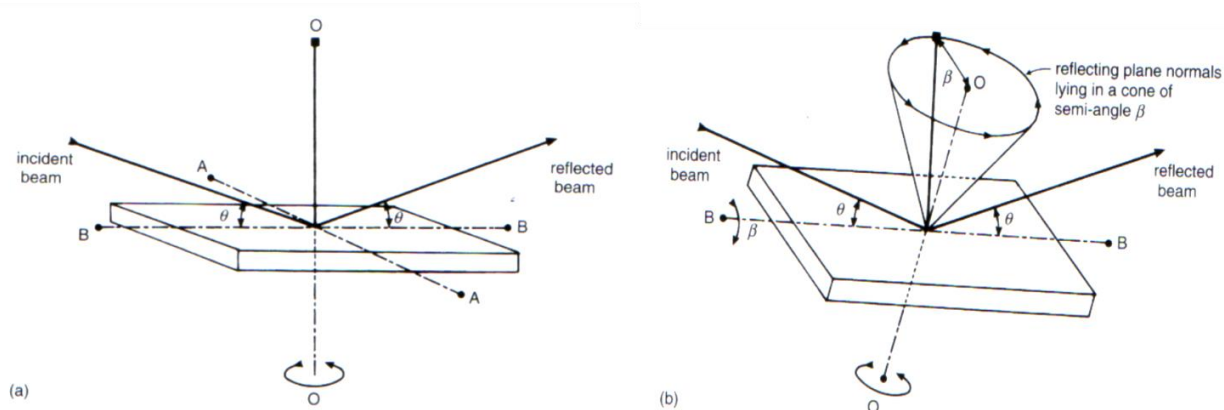
Figure 2.1; preferred orientation effects in powder samples. Reference from [2.3](c).

The finely ground sample in the sealed capillary tube is mounted onto a goniometer and rotated in the beam of X-rays in order to try and overcome any orientation effects (figure 3). The fewer air pockets and gaps in the sample packing, the higher the signal to noise ratio will be.
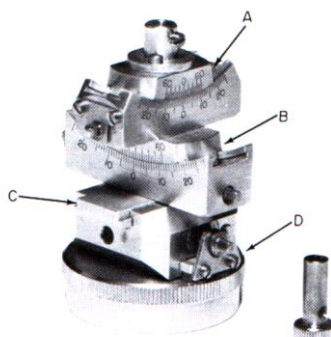
Figure 2.3; a laboratory goniometer head used to mount a capillary packed with finely ground sample. A and B are moveable arcs, C and D are moveable sledges. Reference from [2.3](c).

Finally, the thinner the sample, the better the peak definition will be. This is due to a scattering coefficient determining interference effects which is described in Appendix 1.

2.2.2 Calculating Unit Cell Dimensions and Space Group Assignment

Once a powder diffraction pattern of sufficient quality has been generated, the unit cell dimensions can be determined and its accuracy proven. De Wolff developed a figure of merit which calculated the reliability of powder pattern indexing using the geometry of the diffraction pattern [2.4]. A decade later, Smith and Snyder published a criterion for rating powder diffraction data [2.5], giving each pattern a 'quantitative quality factor'. As long as the pattern is of sufficient quality the indexing step is relatively straightforward. It is necessary to index the X-ray diffraction pattern in order to determine which planes give rise to which peak, sometimes even when the pattern is of poor quality. Early indexing methods required the user to define a series of peaks from experimental data for the program to calculate the unit cell parameters [2.6 – 2.11]. These methods break down into two types of methods, the search methods and the deductive methods. Search methods, also known as optimization or dichotomy methods, such as DICVOL [2.11], generate unit cell parameters and generate peak positions for this unit cell. The suitability of the solution is given a 'Figure of Merit' by comparing the calculated peak positions with the experimental peak positions. The parameters are successively reduced until an exhaustive search of all parameters has been carried out. Deductive methods, such as TREOR [2.10a] and ITO [2.7] recognize well-defined relationships between sets of peak positions and propose unit cell parameters based on these comparisons. A new approach proposed by Kariuki [2.12] suggested that the previous methods were susceptible to user error and put forward a 'whole-profile' fitting and global optimization process for powder pattern indexing called GAIN.

Since the original three 'classic' programs for indexing (DICVOL, TREOR and ITO), [2.23, 2.24] revisions of these methodologies and new processes have been developed to generate a host of options available to the practitioner. Monte Carlo and Artificial Neural Networks are just a couple of the approaches being used in the more recent programs [2.13 – 2.19]. In order to be as 'user-friendly' as possible, 'user interfaces' and 'wizards' guiding the practitioner through the process have been developed considerably too [2.20 – 2.22, 2.25].
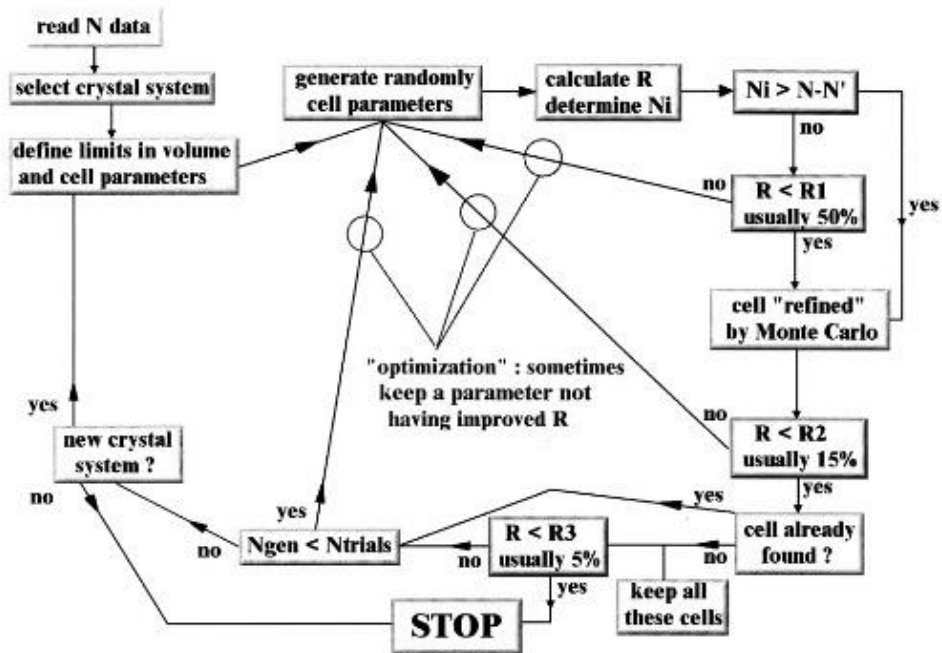
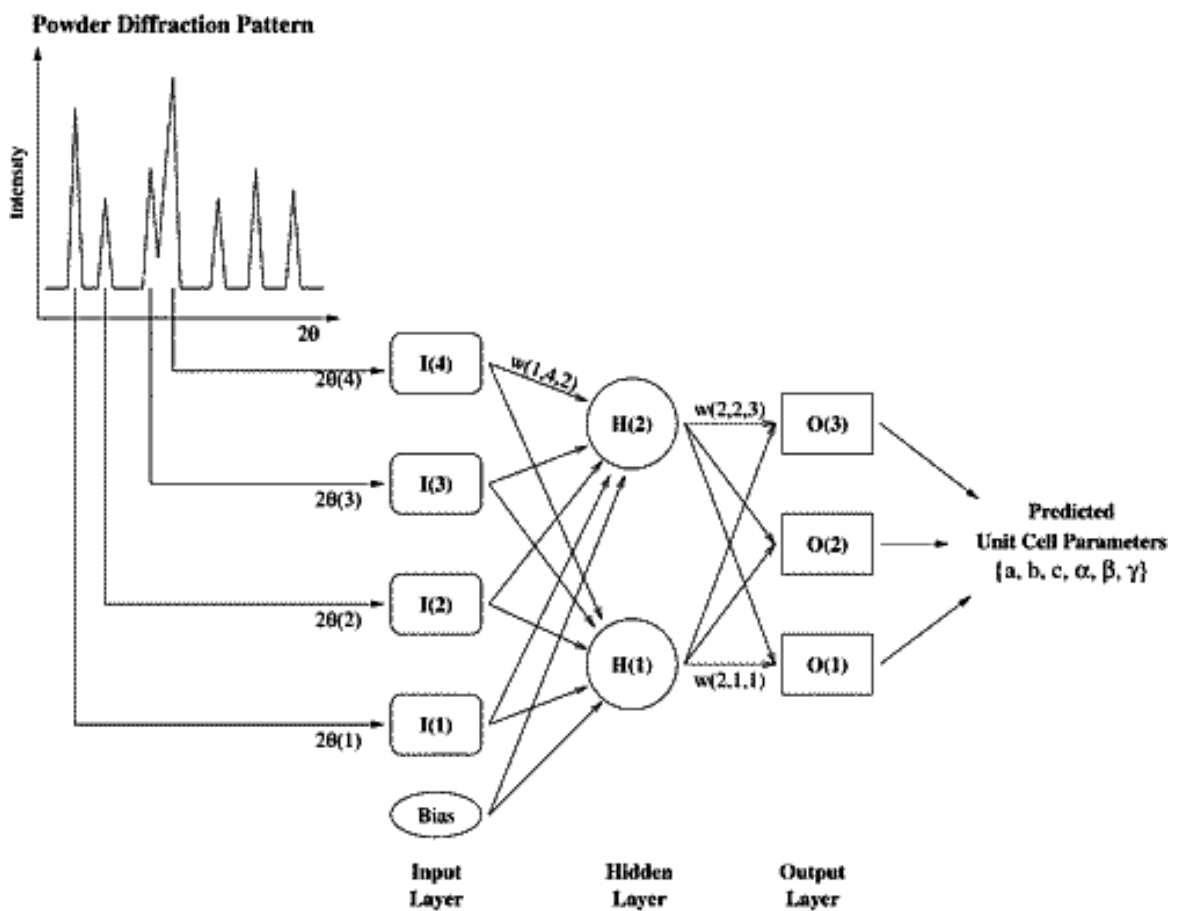**Figure 2.4;** Monte Carlo indexing protocol, taken from [2.16]

Figure 2.5; indexing using an artificial neural network approach, taken from [2.18]

Next a space group is assigned using the systematic extinctions in the powder pattern. Due to the extent of peak overlap in powder diffraction experiments, absent peaks may be difficult to spot therefore during this stage of the process, consideration is also often given to the likelihood of a space group occurring based on information relating to similar organic compounds [2.26].

Occasionally, the intensities of certain reflections from a plane will be equal to zero i.e. the reflection is forbidden or systematically absent [2.3b]. Such absences can arise from the centering of the unit cell and/or the presence of translational symmetry elements like glide planes or screw axes. Therefore, systematic absences of groups of specific reflections can imply the presence or absence of certain symmetry elements.

2.2.3 The Phase Problem

The intensities of waves scattered from planes within a crystal are measured in diffraction experiments [2.27]. Since the contribution of the nucleus to scattering is negligible, the electron density, which is centered on the nucleus, is effectively measured by X-ray diffraction. In order to calculate the position of the electron density in a unit cell however, the phase related to the particular structure factor amplitude is required and it is this information which cannot be calculated. This is the phase problem and has led to a number of phasing methods being developed to try and determine as many unique intensities as accurately as possible [2.28- 2.34].

Program	Description	www
Superflip	Program for solving structures by charge flipping. **Distributed with Jana** and immediately available.	**superflip.fzu.cz**
MCE	Program for electron densities visualization. **Distributed with Jana**, available after defining the path in Tools->Preferences->3d map visualization.	**www.vscht.cz**
VESTA	Program for electron densities visualization. Available after download, installation and defining the path in Tools->Preferences->3d map visualization.	**www.geocities.jp**
SIR97, SIR2002, SIR2004	Program for solving structures by direct methods from single crystal diffraction data. Available after download, installation and defining the path in Tools->Preferences.	**www.ic.cnr.it**
EXPO, EXPO2004	Program for solving structures by direct methods from powder diffraction data. Available after download, installation and defining the path in Tools->Preferences.	**www.ic.cnr.it**

Table 2.1; summary of a sample of methods for deconvolution and structure solution in reciprocal space. Sourced from http://jana.fzu.cz/

2.3   Crystal Structure Solution Methods from X-ray Powder Diffraction Data

2.3.1   Solving the Phase Problem in Reciprocal Space; Direct methods, Patterson method, Maximum Entropy and Likelihood Ranking.

Direct methods were developed to be an effective way of solving the phase problem, for small structures. The methods don't rely on the presence of an atom which acts as a dominant scatterer or prior knowledge of the geometry of a well-defined fragment. Direct methods assume a crystal is made up of atoms which are similar in shape and all have a positive electron density. As a result, statistical relationships exist between sets of structure factors and these relationships are used in order to assign possible values for the phases. However, problems with this method arise as the number of atoms increases and the statistical relationships become weaker. Problems also occur with this method when the quality of the diffraction data is low. Even weak reflections are required in these calculations and some are lost in the background noise, and in the case of powder diffraction, from overlapping peaks [2.35 – 2.37].

Direct methods are still useful for processes for example in isomorphous replacement where phase values are deduced from the way the structure factors are perturbed. Changes in the crystal will give rise to this effect and in isomorphous replacement, this is achieved by the addition of a heavy atom.

In the Patterson method, calculations require quantities directly available from experimental data such as the angular positions and intensities of the diffracted beams. The Patterson function is then used to construct a "map" of the relative positions of the atoms in the unit cell. The peaks actually relate to the locations of the interatomic vectors rather than the atoms themselves. The Patterson function is most useful to use when attempting to define the location of a heavy atom (a dominant scatterer) for a cell also containing many light atoms. Here the dominant scatterer produces a large peak, which is easy to define. The drawbacks to this method arise when only light atoms occur in the unit cell; the peaks overlap and became difficult to resolve [2.38].

The maximum entropy coupled with the likelihood ranking method supports a single crystal like approach to structure solution and has been described as an improved direct method approach. Data from powder diffraction experiments is compressed into one dimension, unlike single crystal data, which causes peaks to overlap. When attempting to extract the intensities, ambiguities arise potentially causing errors when trying to

generate a trial structure. Prior to the development of this method the overlapping peaks were either assigned arbitrary values or were simply ignored, neither approach being acceptable for accurate structure generation. With maximum entropy and likelihood ranking, the starting structure was believed to consist of atoms whose chemical identity was known but whose positions were random and unknown. Structure solution effectively 'reduced the randomness', of the atomic positions. Since the method relied on dominant scatterers, it was only really suitable for compounds containing a "heavy atom" [2.39 – 2.45].

Dual space methods have since been developed which use information from both the reciprocal space data and the real space information [2.46 – 2.48]

2.3.2 Crystal Structure Solution in Direct Space

Alternative methods allow structure solution methods to search direct space for possible trial structures [2.49 – 2.51]. As well as comparing diffraction patterns, calculated structures can be ranked based on their lattice energies and by monitoring the separation distances between non-bonded interactions.

At the time of this research, the methods using the asymmetric unit to search direct space included Monte Carlo (and Metropolis or 'modified' Monte Carlo), Simulated Annealing, Systematic Searching, Clustering and Genetic Algorithms.

The Monte Carlo and the modified Monte Carlo (Metropolis) methods involve a random search of the unit cell. The molecular unit, treated as a rigid body, is given three degrees of translational and three degrees of rotational freedom. The differences in a function reflecting the new structure quality determine if the move is permissible. If the move is accepted, the new position is then redefined as the new starting position.

Typically the energy change in the system or the fit of the calculated and experimental diffraction profiles are the functions employed to accept or reject a move [2.52 – 2.53].

The equation to calculate the energy of the system is shown in equation (1).

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}V(d_{ij}) \qquad (2.1)$$

where E is the potential energy of a system, N is the number of particles in the system, i is one particle, j is a second particle, V is the potential energy between particles i and j and đ$_{ij}$ is the minimum distance between particles i and j

The crystallographic agreement factor is calculated using equation (2)

$$R_{wp} = 100 \times \sqrt{\frac{\sum w_i (y_i - y_{ci})^2}{\sum w_i y_i^2}} \tag{2.2}$$

R_{wp} = the crystallographic agreement factor

w_i = a weighting factor for the ith point in the pattern

y_i = the intensity of the ith observed point in the experimental pattern

y_{ci} = the intensity of the corresponding ith point in the calculated powder pattern

New structural models are generated independently of any diffraction data to begin with. The diffraction pattern of the new model is then calculated and compared against the experimental data. This approach avoids the need to extract information from overlapping or weak reflections [2.54 – 2.56].

Simulated annealing, similarly a random search of direct space, involves 'melting' the material until a predefined number of consecutive moves are accepted. Here the 'temperature' is the normalising quantity in the calculation of the Boltzman probability factor. The temperature is increased until the requisite number of consecutively accepted moves is achieved. The system is then described as being in 'a molten state'. The molten condition implies that all rotations (including sub-rotations) are acceptable, after which the system is 'cooled' to a user-defined temperature. At the lower temperature the system is described as frozen (the heating and cooling rates are also user-defined). More moves allowed per unit temperature change give greater coverage of the potential energy surface, however, this also means many similar structures are generated [2.57 – 2.63]. It becomes difficult to determine significant differences between structures having a similar rating. This has been an issue which the blind tests organized by the CCSD have addressed. More details of these tests are in section 2.6.

Systematic searching, as the name implies, employs a grid-based search of a unit cell [2.1] using molecular packing considerations and lattice energy calculations [2.64 – 2.65]. The

27

asymmetric unit, treated as a rigid body, is given three degrees of rotational freedom and three degrees of translational freedom. In further studies, intramolecular degrees of freedom are also allowed. User-defined limits are set to constrain various parameters of the search. If a trial structure is accepted it is stored for assessment and refinement. To find the global minimum, the molecule is set in a defined starting position and all local minima nearby are visited. This is carried out for a selection of starting positions. The local minima are then ranked according to energy values, the lowest of which is expected to be the global minimum. The generated structures are assessed using lattice energy or powder pattern fits in a similar way to the random search methods described above. The limitations to this method are seen when predictions of flexible structures are attempted. The method becomes time consuming, especially if large unit cells are being searched, although consideration of symmetry aspects does sometimes allow for only a fraction of a cell direction to be considered. Also at this time, the method could not be used if the asymmetric unit contains more than one molecule.

In the clustering approach, small units are constructed (usually dimers) using the most commonly encountered symmetry elements. The assessment of the resulting structures is carried out using statistical analysis of known assemblies with a similar structure. A full crystal structure is then built in the most frequently encountered space groups by translation of the initial cluster, and a final lattice energy is calculated [2.66 – 2.69].

Genetic algorithms employ a genetic approach to trial structure generation by mating, mutating and using natural selection to generate a population of "fittest" structures [2.50, 2.70 and 2.71]. These are the structures with the lowest energy. Like the previous random search methods, the diffraction patterns are usually compared as the acceptance criteria, however lattice energy may also considered as acceptance criteria to allow two different ranking methods.
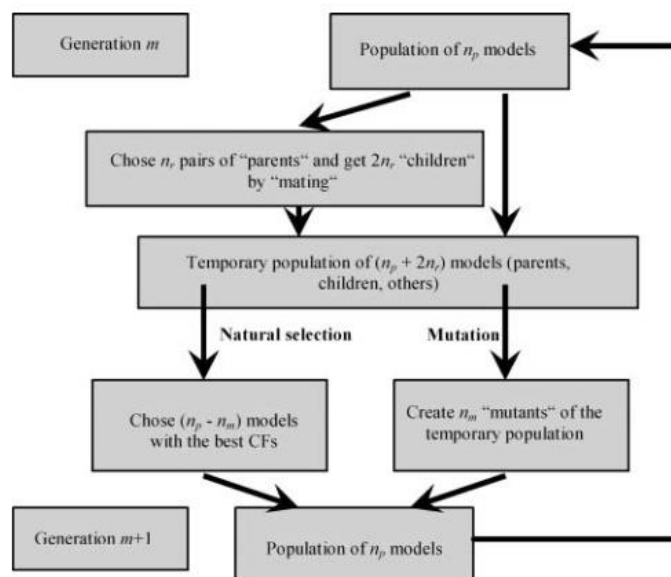
Figure 2.6; steps in a Genetic Algorithm Sequence. From [2.49]

2.4 Review of Crystal Structure Solution Methods

Since the conclusion of the practical side of this research, the dominant method in CSS are Global Optimization methods [2.49, 2.72], also known as direct space methods, model building methods and sometimes real space methods. In structure determination from powder diffraction data, these methods involve moving a molecular model of the study molecule around a known unit cell. The conformation, position and orientation are constantly being adjusted until the best agreement with the observed diffraction data is obtained or the global minimum on the energy surface is reached.
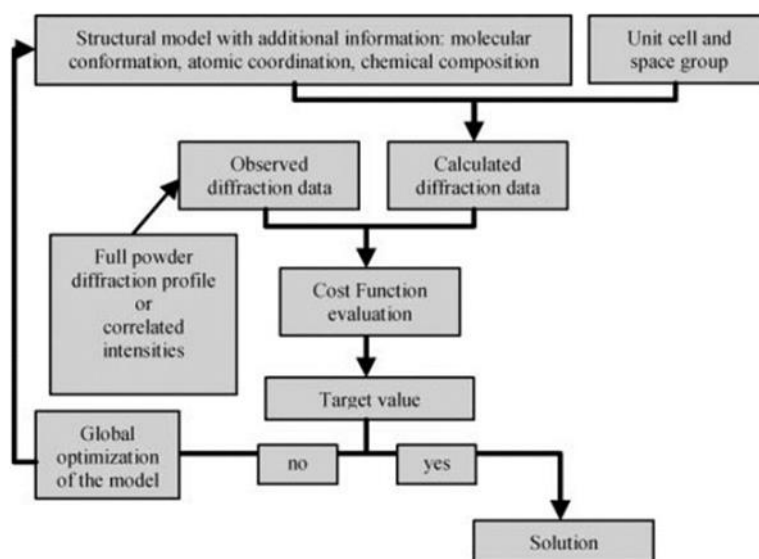

Figure 2.7; steps in a Global Optimization process. From [2.49].

Finding the global minimum of the energy surface is not guaranteed and there isn't a way to test if the generated result lies within it or whether only one of the local minima has actually been found. One simple approach to testing for the global minimum has been to generate many random starting conformations and to perform optimizations on each. Clearly, this can only work for very small systems. Another simple approach has been to apply a weighting using a Boltzmann distribution by the Monte Carlo method. The generated structures are weighted based on their energy; the lower the energy the higher the weighting. However, when the thermal energy (k_BT) is low, the high energy regions cannot be crossed and simulations become trapped in a catchment basin. Even when a transformed potential energy surface is used whereby the energy restrictions from funnels are effectively removed by assigning the only the local minima energy value in each catchment basin, there are still energy barriers between the different levels of the transformed surface which cannot be crossed.

Since the completion of the practical work in this research, different types of Global Optimization methods have been developed;

2.4.1 The Grid search

As previously described in 2.3.2, the unit cell is broken down into a grid and the asymmetric unit is systematically moved within it. With this method, the global minimum is guaranteed to be found if the grid search is fine enough. The balance must be struck between computational time and fineness of the grid. Since fine grid searches can be too time consuming, it is often preferable to run numerous searches in parallel [2.73]. However, for complicated examples such as larger, flexible molecules, even with Parallel Computing, this might not be the best overall approach.

Rigid molecular conformations have been used in stepwise searches. The rigid conformer is used to probe the regions of conformational space in a grid-based search to determine which areas may give rise to low-energy structures [2.74, 2.75].

The benefits of a Monte Carlo approach has been combined with a grid search in the program 'Organa' [2.76]. Here an energy-guiding Monte Carlo search of direct space is used with the option of a grid-based search.

2.4.2 Random searches;

Monte Carlo (MC) sampling was initially used in Crystal Structure Solution (CSS) from Powder Data, with Simulated Annealing (SA) following soon afterwards. SA quickly became the more popular search method thanks to its ease of use and quick computational time compared to grid searching. Parallel Tempering (PT) offered further advantages to SA since fewer starting parameters were required (e.g. annealing rates, starting temperatures). The PT method is also generally able to escape from local minima within the search space [2.77].

Evolutionary Theory (ET) – which includes Genetic Algorithms (GA) – is a population based search using evolutionary algorithms [2.78-2.88]. USPEX [2.89-2.90], EVO [2.91], XTALOPT [2.92] are notable examples. Evolutionary algorithms mimic evolutionary theory by forming new generations using selection, cross-over and mutation until a 'stop' condition is reached. GA [2.93] avoid the restrictions imposed using a single starting point by beginning with an initial population of structures and EAs tend to search different areas of the energy landscape simultaneously.

Both SA and GA have been modified to improve the rates of convergence and improve minimization strategies. This is achieved by adding Variable or Dynamic Penalty Functions [2.94].

Other modifications or complementary methods to GO methods include;
- Hybrid Monte Carlo (HMC) combines a Monte Carlo (MC) approach and Molecular Dynamic (MD) calculations [2.95, 2.96].
- Hybrid GA [2.97]. The efficiency of the GA is improved in this method with the implementation of neural network within the procedure. A neural network is an information processing technique which mimics the learning capability of the brain.
- Semi-global and local searches – optimizing the rates of convergence during the optimization step at the end of the SA cycle [2.98].
- The Monte Carlo Minima hopping (MCMH) algorithm has been developed to try to alleviate the catchment trapping problem. This method effectively recognizes when a low energy region has been visited already, does not try to search it again, but can then hop across high energy regions if necessary in order to reach other local minima ("tabu" regions [2.99]). Minima hopping methods also look to cross

the lowest energy barriers when possible since local minima are more likely to be found behind a low energy barrier than a high energy barrier. The Minima Hopping Method has been modified to allow for a systematic search approach [2.100 – 2.102].

- Multi-Canonical Basin hopping (MCBH) method. This global optimization method works by proposing a change to a current system and accepting this change if it reduces the energy. It combines the advantages of using the basin hopping method with the Monte Carlo method. Therefore, it can efficiently find minima on the energy surface and can avoid being trapped in them by making a move ('hop') out of local minima in order to continue searching the energy surface [2.103 – 2.105].

- Data Mining requires a large database of experimentally observed structures from which to gather information and predict properties [2.106 - 2.108]. "Cluster analysis" is an important tool within Data Mining. As the name suggests, cluster analysis compares and groups crystal structures with similar traits. There are different ways in which data is assessed [2.109] and different ways of linking data within these assessments;
    - Hierarchical Clustering undertakes a series of steps to partition structures into particular clusters. Clustering does not occur within one step.
    - Agglomerative Clustering works the opposite way from hierarchical clustering. Each structure starts from a single point and with each step, the structures are grouped into clusters.

- Parallel GA; these can be broken down into three types;
    - Single population master-slave PGA
    - Single population fine-grained PGA
    - Multiple population coarse-grained PGA

  The third type offering the biggest difference in the original GA methodology by not only generating more than one population in parallel, but by adding in a new evolutionary operation "migration". By allowing communication between multiple populations, the operator adds further diversity to the population and helps prevent stagnation [2.110, 2.111].

- Differential Evolution (DE); is an evolutionary algorithm which generates populations of structures over many generations. The Cultural Differential Evolution Hybrid (CDEH) methodology uses the experiences of previous generations to influence the outcome of future ones. Dynamic boundaries

'shadow' the evolution and clustering of a population over generations [2.112]. DE has also been combined with MC methods to predict the structure of disordered systems [2.113].

- Particle-swarm Optimization (PSO) [2.114]. This method has been described as an EA like GA but without the crossover and mutation processes. PSO was "inspired by the choreography of a bird flock" which searches areas of multidimensional space. Within PSO, an individual learns from previous experiences and can change its direction and "flying speed".

- Periodic Graph Approach [2.115] which has been applied mostly to supramolecular architectures containing intermolecular bonds. Instead of describing a crystal structure as a set of points in space, topological representation focuses on describing the system of chemical bonds within the structure. Topology-based ab initio computational approaches for CSS and more recently Crystal Structure Prediction (CSP) excludes random walk approximations in the search for new crystal structures [2.116]. New structures are generated using 'topological nets' collected from databases. Relevant structures are selected at this stage by examining bond lengths and angles [2.117]. A quick geometrical relaxation is performed using classical force-fields or approximate DFT and the structures with the lowest energy are retained. A full DFT calculation is finally performed on these structures.

- GA with Dynamic Diversity Control (DDC). At the outset of the search, a threshold value is set for population diversity. Following an iteration, the population is assessed. If the diversity is calculated to be below the threshold value, the structures with the highest 'fitness' factor will be extracted while remaining structures will be rejected. As this is happening, artificial structures with greater diversity will be introduced into the population. The iteration starts again and the diversity calculation should be higher than the threshold. Once it is, the search continues along the normal GA routes [2.118]. This process is also carried out with multiple starting populations (MDDCGA).

- MUSE [2.119]: Multi-Algorithm-Collaborative Universal Structure-Prediction Environment. Crystal Structure Prediction is achieved by combining Evolutionary Algorithms, Simulated Annealing and Basin Hopping techniques. The MAC (Multi-Algorithm Collaborative) also includes two new functions to increase diversity; slip and twist.

Slip; moves a molecule a random distance along a random direction which is parallel to a specific plane. This operation was inspired by real crystal phase transitions.

Twist; the atoms are rotated in the cell around an axis which is chosen at random. Only the atomic fractional co-ordinates are moved in this operation, the cell remains fixed.

MAC also incorporates a 'ripple' function. This function shifts the co-ordinates of each atom along a randomly chosen axis. The ripple has been combined with a mutation operator (the 'mutation-ripple'). With this hybrid operator, the ripple function is applied directly after a mutation.

2.5 Molecular Simulations from X-ray Diffraction Patterns

When constructing a chemical model, the aim is to make simulations reflecting reality as accurately as possible and in order to use Molecular Dynamics or Monte Carlo methods, the user must first define the interactions between the atoms in the system. To be effective, it is important that a potential model describing these interactions is transferable between related systems. Various approaches have been developed in the past which include;


2.5.1 Empirical Potential Functions (Pair Potentials)

Molecular mechanics (MM) calculations are based on a simple classical-mechanical model of molecular structure, resulting in highly simplified versions of an actual physical model. Early force fields contained very few parameters and as such only really worked for very small molecules. They included terms to describe each chemical bond, disregarding any non-bonded interactions. Included in these early data sets were terms for bond stretching, angle bending and torsional parameters. As such they were called the "Valence Force Fields". The need to describe interatomic interactions was considered though and these were included in later MM force fields such as MM2 (and subsequently MM3 and MM4), CHARMM, AMBER, All-Atom and UFF [2.120 – 2.136]. These proved useful for small organic molecules (mainly hydrocarbons). The user assumes a functional form for the potential function and selects parameters to fit experimental data. Pair potentials, as the name suggests consist of only pair-wise terms.

In MM calculations, molecules are described as hard spheres, lying in equilibrium with respect to each other where the potential energy is at a minimum and bond terms are in a

'natural' state. The molecules are treated as an array of atoms governed by a set of classical-mechanical potential functions derived from experimental data. The overall potential energy (E_{pot}) of a system is calculated by a summation of the energy functions;

$$E_{pot} = E_b + E_\theta + E_\phi + E_w + E_{vdw} + E_{el} + E_{b-\theta} + E_{hb} \tag{2.3}$$
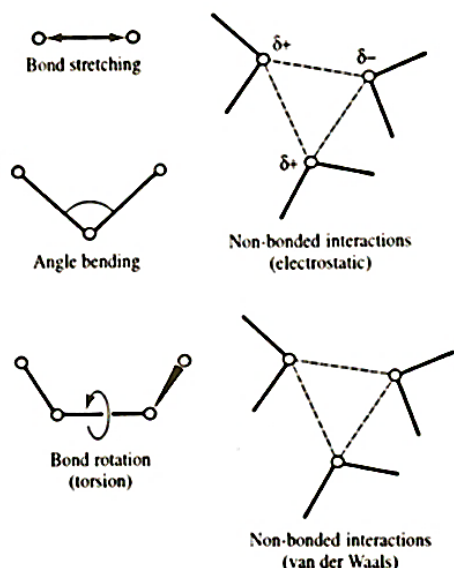
where the energy 'E' terms are E_b = bond stretching, E_θ = angle bending, E_ϕ = torsions, E_w = out of plane deformations, E_{vdw} = non-bonded interactions, E_{el} = electrostatic, $E_{b-\theta}$ = bond stretching, angle bending and E_{hb} = hydrogen bonds.

Figure 2.8; (left) bonded and non-bonded interactions parameterized in force fields.

In a simple structure, the bond stretching term for example is well defined by the Morse curve, figure 9 (right), where D_0 is the bond energy in kcal/mol, R_0 is the equilibrium bond distance in Å. For the majority of examples, this is more fitting than the idea of the hard-sphere potential where atoms are considered as spheres which cannot overlap in space. This model assumes that when the interatomic distance is too short, the potential energy very quickly tends towards infinity.
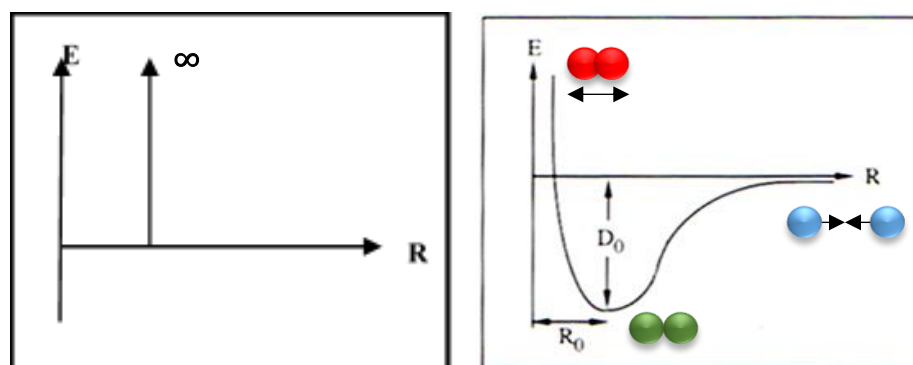

Figure 2.9; (left) graphical representation of the Hard Sphere Potential and (right) the Morse Potential. ● = strong repulsive forces when the atoms are too close together. ● = atomic separation is at an energy minimum and ● = weak attractive forces.

Most bond lengths will lie at an optimum distance that can be described using Hooke's law, equation (4) where E_{pot} is the potential energy, k is a defined constant, R is the interatomic distance and R_0 is the equilibrium bond length.

$$E_{pot} = \frac{k(R - R_0)^2}{2}$$ (2.4)

This term becomes relatively easy to fit, but inaccuracies arise when bond lengths are extended due to steric effects and charge distribution. The bond lengths then lie outside the optimum region, though extra terms could be added to the equation, making it more and more complicated to carry out the calculations. Other terms such as bond angle related terms could also be treated using equation (4) by replacing the term for (R - R_0) by (θ - θ_0).

Extensive parameterization of MM methods is required to enable treatment of different types of systems (e.g. proteins, flexible hydrogen bonded systems, carbohydrates [2.137 – 2.141]). The generic Dreiding Force Field [2.126], chosen for this study, has calculated lattice energy expressed as a sum over all atom pairs with three independent terms in the summation the form is shown in equation (5).

$$E = 1/2 \sum_{k=1}^{N} \sum_{i=1}^{n} \sum_{j=1}^{n'} V_{kij}$$ (2.5)

where;

N is number of surrounding molecules, n is number of atoms in the central molecule, n' is number of atoms in each of the N surrounding molecules, V_{kij} is interaction between atom i in the central molecule and atom j in the kth surrounding molecule. The valence term in Dreiding consists of equation (6).

$$E_{val} = E_B + E_A + E_\Upsilon + E_l$$ (2.6)

where E_B is the term for bond stretching, E_A for bond-angle bend, E_Υ describes dihedral angle torsion and E_l for inversion terms.

Prior to this research, it was considered for small molecules that charges could be calculated using a quantum mechanical 'Hartree-Fock' wave function. However, for

more complicated systems, this approach would not provide a timely solution. Consequently, at this time, in Dreiding, which is a force field which can be used with structures containing many atoms and more than one degree of flexibility, the charges on the atoms are either ignored completely or Gasteiger [2.142] estimates also termed 'point-charges' are employed. This is instead of using the 'population analysis' procedure proposed by Muliken [2.143] where overlapping charges were divided equally between the two contributing atoms, or various 'charge equilibrium' or distribution methods proposed by Politzer [2.144] and Rappé [2.145] which determined charge distributions based on geometry and the atomic environment.

The non-bonded interactions (attractive and repulsive short-range interactions) in Dreiding are expressed using a Lennard Jones 6-12 type potential energy term. The attractive forces include dipole/dipole, dipole/induced dipole and induced dipole/induced dipole (London or dispersion) interactions [2.146]. The repulsive forces tend to dominate when the atoms are brought closer than the Van der Waals radii (this is the sum of the radii of the two atoms). In the form used in equation (7), R equals the atom separation.

$$E_{VdW}^{LJ} = AR^{-12} - BR^{-6} \tag{2.7}$$

where AR^{-12} describes the repulsive forces and BR^{-6} the attractive forces.

Secondly, a specific term with a 10-12 functional form is used to represent hydrogen bonds in the crystal lattice where; R_{DH} is the distance between the donor and the acceptor, θ_{DHA}, the angle between the donor, hydrogen and acceptor atoms (D, H and A).

$$E_{hb} = D_{hb}[5(R_{hb})/R_{DA})^{12} - 6(R_{hb}/R_{DA})^{10}]\cos^4(\theta_{DHA}) \tag{2.8}$$

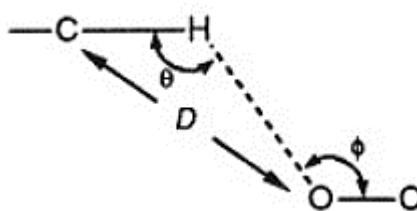This hydrogen bond expression is depicted in figure 10.

Figure 2.10; expression of the hydrogen bond in equation (2.8) taken from [3.47].

The simplistic nature of the point charge approach assumes that the charge distribution around an atom is spherical. This would seem to be an unrealistic approach [2.147] which is addressed with the more recent use of multipoles to describe the electrostatic contribution of a system [2.148– 2.150, 2.154]. Multipolar electrostatics can describe both the directionality and the strength of a bond. Since hydrogen bonds are both directional and are often a strong interaction, this approach would seem to offer an improvement on the point charge method [2.151].

Finally, Dreiding uses a coulombic energy term [2.152] to describe the polar interactions.

$$E_Q = (322.0637)Q_iQ_j/\varepsilon R_{ij} \tag{2.9}$$

Q_i, Q_j are the charges assigned to the atoms in electron units, R_{ij} is the interatomic distance in Angstroms, ε is the dielectric constant (normally = 1) and 322.0637 is a constant that converts the energy into kcal/mol.

2.5.2   Semi-Empirical Potential Functions

Electronic wave-functions for fixed atomic positions contain information about a given system and, for simple systems, can be calculated by solving equation (10), the time-independent Shrödinger equation.

$$E\Psi = \hat{H}\Psi \tag{2.10}$$

where;
$\hat{H}$ is the Hamiltonian Operator corresponding to the sum of the Potential and Kinetic Energy of all the atoms in the system, Ψ is the wavefunction and E is the energy of the wavefunction.

This calculation becomes impossible for many bodied systems and approximations are required. Empirical atom-atom force fields alone are useful when considering small systems, however, these are not accurate enough to use for larger flexible molecules [2.153, 2.154]. One compromise has been to use the faster empirical force fields coupled with the slower but more accurate ab-initio calculations [2.141, 2.155]. A second approach has been to calculate charge density rather than relative energies of crystal structures [2.156 – 2.157].

2.5.3 Ab-initio functions

Ab-initio methods only use approximations in their equations derived from theoretical consideration and do not contain any experimental data [2.169]. The simplest form of ab-initio calculation use the Hartree-Fock Method, which is an extension of the Molecular Orbital Theory. Hartree-Fock averages the electron-electron repulsion effects over a system in order to average the wave function of that system. Hartree-Fock methods assume the system is in a stationary state and in earlier articles it is referred to as the Self-Consistent Field Theory (SCFT).

Many post-Hartree-Fock methods begin with the Hartree-Fock approximation and use an electron-correlation addition to the equation. This is more accurate since in Hartree-Fock the electron-electron repulsions are only averaged. Examples of Post Hartree-Fock methods are configuration interaction, coupled cluster, Møller-Plesset perturbation theory (for example MP2, MP3, MP4), Quadratic Configuration Interaction and quantum chemistry composite methods (G2, G3, CBS, T1) [2.158, 2.159].

Despite the cost in time when using ab-initio calculations within crystal structure prediction strategies, it is still recognised as a more accurate and desirable approach [2.153]. Compromises in search strategies to minimise computational time have been proposed especially when treating flexible molecules. Ab-initio calculations are used to propose a set of rigid configurations which are used to search for low energy crystal structures. Accurate intermolecular potentials are used to optimise the results from this initial 'exploratory' search [2.160]. Reasonable results from the initial search are then used in subsequent cycles of the same strategy. Recently a Quantum Mechanically Derived Force Field which provides a 'black box' approach to generate classical potential energy functions which are molecule-specific [2.161] has been presented. This would allow the computation of systems where standard parameterisation is not possible.

Density Functional Theory (DFT) is a technique to treat many bodied systems [2.162-2.165]. Where a function is a process which relates an input to an output, a functional is a process which takes a function as the input to produce an output. Hohenberg and Kohn stated that the energy of a polyatomic system is uniquely determined by a functional of the electron density [2.165]. Kohn and Sham then determined that the total ground state energy of a polyatomic system is a functional of the electron density. Their formulation which proposes the electrons do not interact with external potentials such as the nuclei is shown in equation (11).

$$E_0[\rho(r)] = \int E_{KE}[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)] \qquad (2.11)$$

where;

$E_0[\rho(r)]$ is the ground state energy as a functional of the electron density.

$E_{KE}[\rho(r)]$ is the kinetic energy term.

$E_H[\rho(r)]$ is the Hartree term (electron – electron repulsion)

$E_{XC}[\rho(r)]$ is the exchange-correlation term which is not known precisely but approximations may be calculated for it.

Methods to calculate the exchange-correlation energy functional, including approximating for vdW interactions [2.166 and 2.167], have improved the accuracy of results when simulating molecular energy surfaces.

In summary;

| Method	Advantages	Disadvantages	Best for
Ab initio • Uses quantum physics • Mathematically rigorous: no empirical parameters	• Useful for a broad range of systems • Does not depend on experimental data • Calculates transition states and excited states	• Computationally expensive	• Small systems (tens of atoms) • Electronic transitions • Systems without experimental data • Systems requiring high accuracy.
Semi-empirical • Uses quantum physics • Uses experimental parameters • Uses extensive approximations	• Less demanding computationally than *ab initio* methods. • Calculates transition states and excited states.	• Requires *ab initio* or experimental data for parameters. • Less rigorous than *ab initio* methods.	• Medium-sized systems (hundreds of atoms). • Electronic transitions.
Molecular Mechanics • Uses classical physics • Relies on force field with embedded empirical parameters.	• Computationally 'cheap:' fast and useful with limited computer resources. • Can be used for large molecules like enzymes.	• Does not calculate electronic properties • Requires *ab initio* or experimental data for parameters • Commercial software applicable to a limited range of molecules	• Large systems (thousands of atoms) • Systems or processes that do not involve bond breaking.

Table 2.2; some advantages and disadvantages of Molecular Modelling Methods [2.169]

2.6   Testing Crystal Structure Prediction Methods – Cambridge Crystallographic Data Centre Blind Tests from CSP1999 to CSP2016

Since 1999, the CCDC (Cambridge Crystallographic Data Centre) has arranged a series of collaborative workshops designed to gain insight into the variety of techniques available to the field of CSP, how well they perform and if there is one approach which suits all systems.
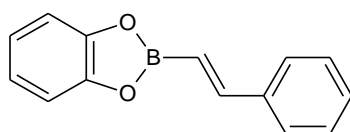
2.6.1   CSP1999 [2.168]

In 1999 for the first collaboration, eleven participants were given only the atomic-connectivity data of four organic compounds. Each group was allowed to provide up to three proposals for the correct structure of each compound. The molecules chosen for this workshop fell into one or more of the following categories;
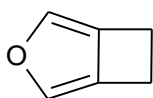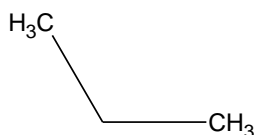
(I) A small rigid structure containing no more than 20 atoms when only C, H, N and O were allowed.
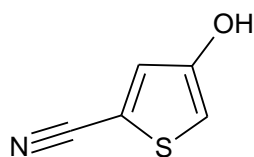
(II) A small rigid structure containing a less common element.

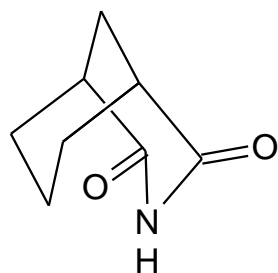(III)   A molecule with a small degree of conformational flexibility.

(VII)   Propane was also suggested by one of the participants.

Figure 2.11; Molecular diagrams of the structures investigated in CSP1999 [2.168]
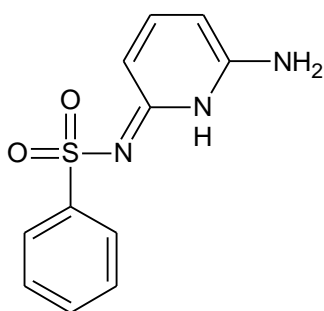
2.6.2 CSP2001 [2.169]

The second blind test was carried out by seventeen participants in 2001. Again the groups were provided with only the chemical diagram of the three compounds of interest. A similar set of criteria as in CSP1999 was set out for the selection of molecular structures to study in this test. These were;

(IV)　　A small rigid structure containing no more than 20 atoms when only C, H, N and O were allowed.
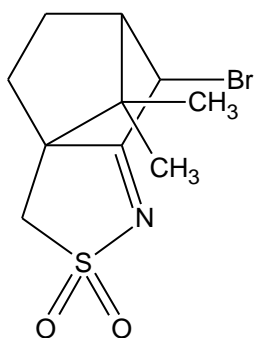
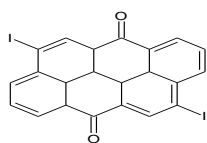(V) A small rigid structure containing a less common element.

(VI)　　A molecule with a small degree of conformational flexibility.

Figure 2.12; Molecular diagrams of the structures investigated in CSP2001 [2.169]

2.6.3   CSP2004 [2.170]

The third workshop was carried out in 2004 with eighteen participants taking up the challenge. The previous workshops had shown that when the asymmetric unit is a small rigid body (where Z' = 1) and when a common space group is used, CSP of the correct structure was no longer an impossible task. The major problems facing CSP at this time were less common space groups, more than one molecule in the asymmetric unit and large flexible molecular structures.
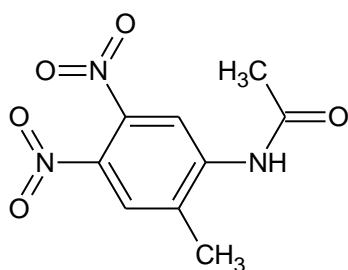
The chemical diagrams of the compounds investigated in this blind test are shown in figure 13.

(VIII) A small rigid structure containing no more than 25 atoms when only C, H, N and O were allowed.
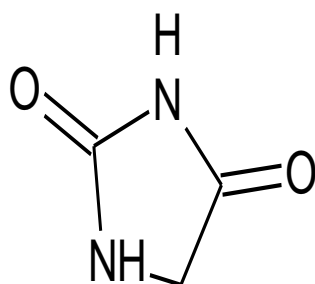
(IX)   Rigid structures containing elements or functional groups which provide problems for CSP methods.

(X)   Molecules with several degrees of conformational flexibility.

(XI)   A further example of a small rigid structure containing no more than 25 atoms when only C, H, N and O were allowed.

Figure 2.13; molecular diagrams of the structures investigated in CSP2004 [2.170]

2.6.4 CSP2007 [2.171]

There were fourteen participants in the fourth blind test in 2007 which included three single component systems and one 1:1 cocrystal. No significant changes were noted in the generation of the starting molecular conformation and most participants still ranked the generated structures based on static lattice-energy values. Two participants went further and considered lattice-dynamics contributions to the lattice energy calculation. The inclusion of a cocrystal added the possibility of hydrogen bonding analysis to aid ranking the structures generated.

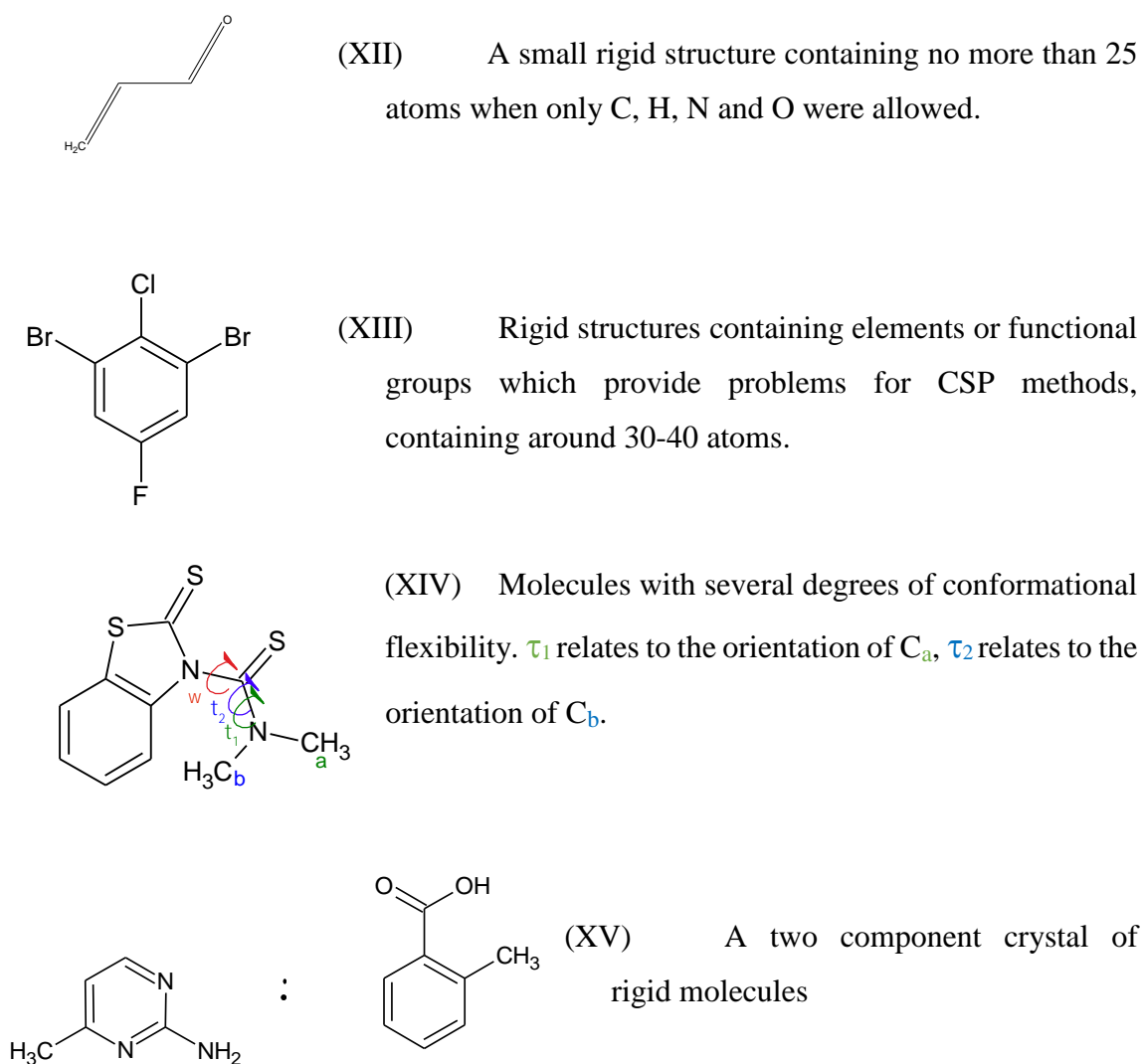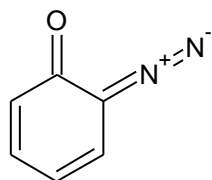The chemical diagrams of the compounds investigated in this blind test are shown in figure 14.

(XII)   A small rigid structure containing no more than 25 atoms when only C, H, N and O were allowed.

(XIII)   Rigid structures containing elements or functional groups which provide problems for CSP methods, containing around 30-40 atoms.

(XIV)   Molecules with several degrees of conformational flexibility. τ_1 relates to the orientation of C_a, τ_2 relates to the orientation of C_b.

(XV)   A two component crystal of rigid molecules

Figure 2.14; molecular diagrams of the structures investigated in CSP2001 [2.171]

2.6.5   CSP2010 [2.172]

Fourteen groups took part in this blind test in 2010 which included four molecules similar to those in CSP2007; two small rigid bodies, one semi-flexible molecule and one salt. In addition, one larger more flexible molecule and one hydrate with more than one polymorph were also included.
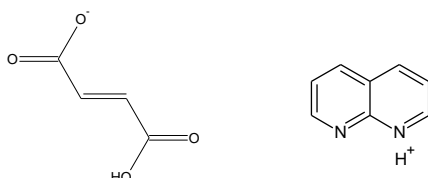
(XVI)   A small rigid structure containing no more than 25 atoms when only C, H, N and O were allowed.
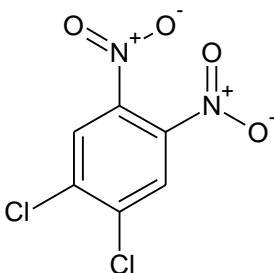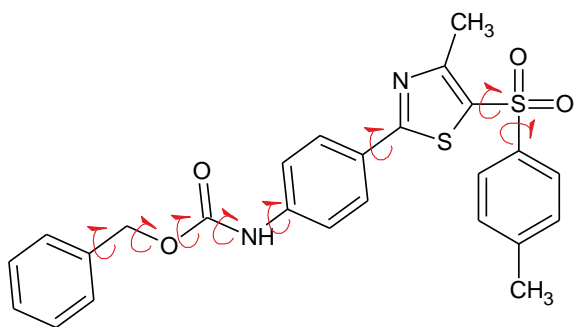
(XVII)   Rigid structures containing elements or functional groups which provide problems for CSP methods, containing up to 30 atoms.
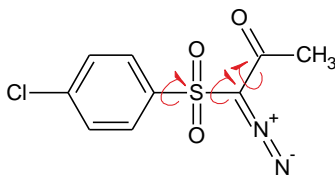
(XVIII)   Molecules with moderate conformational flexibility and containing up to 40 atoms.

(XIX)   Multiple independent rigid molecules containing up to 30 atoms, any space group.

(XX)   A molecule with 4-8 internal degrees of freedom, Z'≤2, any space group, 50-60 atoms.
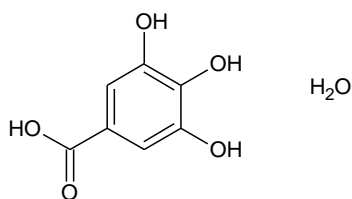
(XXI)   Any molecule which falls into any of the first four categories but has more than one known polymorph.

Figure 2.15; molecular diagrams of the structures investigated in CSP2010 [2.172]

2.6.6 CSP2016 [2.173]

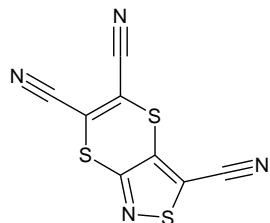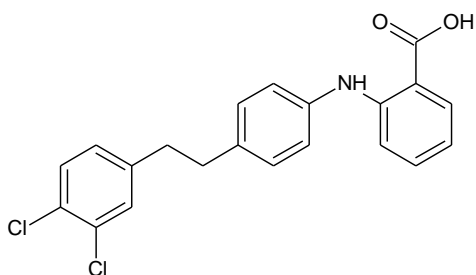Twenty-five groups took part in this blind test which included five different systems; a small nearly rigid molecule, a polymorphic former, a chloride salt hydrate, a cocrystal and large flexible molecule.

(XXII) Rigid molecules with functional groups up to 30 atoms. Restricted to atoms C, H, N, O, Halogens, S, P and B.

(XXIII)   Partially flexible molecules up to 40 atoms with between 2-4 internal degrees of freedom.
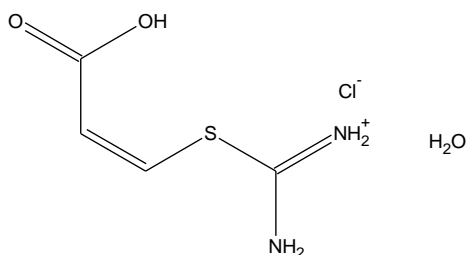
(XXIV)   Partially flexible molecule, up to 40 atoms, as a salt, with up to 2 internal degrees of freedom, in any space group.
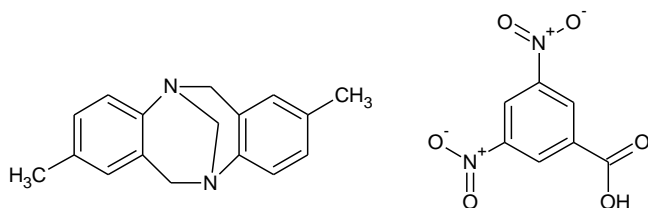
(XXV) Multiple independent molecules in the asymmetric unit as a cocrystal or salt. Up to 40 atoms, up to 2 internal degrees of freedom.
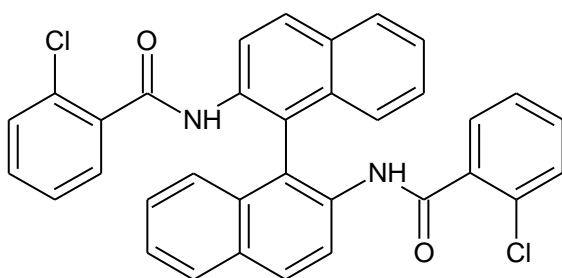
(XXVI)   4-8 internal degrees of freedom, any space group. Between 50-60 atoms and no more than two molecules in the asymmetric unit.

Figure 2.16; Molecular diagrams of the structures investigated in CSP2016 [2.173]

2.6.7   The Methods of CSP throughout the Blind Test Series

At the time of the first Blind Test in 1999, the most straightforward approach to solving the molecular packing problem was to search for the global minimum with energy being the most likely function chosen for minimization. From the first blind test in 1999 the majority of groups calculated the lattice energy of a stationary system with the lowest energy result being taken as the most thermodynamically stable form. In these tests, free-energy contributions from the vibrational enthalpy and entropy were ignored. Statistical analysis of data extracted from the CSD was also being considered as a fitness function, the idea being that the most likely polymorph would compare well with the most frequently seen patterns from extracted data.

Commonly the molecular model was described as a rigid structure which was calculated using *ab initio* calculations. When no flexibility was considered in the molecular structure such as with DMAREL and MOLPAK, the teams carried out multiple searches using different starting conformations. The choice of method used to determine a rigid molecular structure is important since the effect of molecular structure can have a significant effect on the overall crystal energy calculation. Programs such as CRYSCA and MPA were able to add conformational flexibility into the search by defining rotations of rigid fragments around selected bonds. UPACK and MSI-PP however were relaxing the full molecular structure within the crystal structure once the search had found a packing arrangement at an energy minima.

There were a variety of methods utilized to generate crystal structures from the optimized molecular structure. These mostly involved searching space using random steps (Genetic Algorithms such as *Rancel*, Simulated Annealing such as *PackStar*), building 2D grids and 3D lattices using co-ordination geometries (*Promet*), however a grid-based search method (UPACK) was also employed.

Of the methods used, seven predicted structures which were considered as the correct result. These were (I) MPA, MSI-PP, UPACK and *Zip-Promet*, (II) MSI-PP, (III) UPAK and (VII) UPACK. Overall, the results of this workshop were considered to be encouraging. The success of a method was concluded to be dependent more on the search and generation algorithm and the fitness function rather than minimization. The summary of methods used are shown in table 3.

Table 2.3; summary of methods used in CSP1999 [2.168]

Name	Molecular model	Crystal structure generation and search	Fitness function	Minimization by	References
CRYSCA	Flexible	Random	Empirical energy	Derivatives	Schmidt & Englert (1996)
DMAREL	Rigid	Pregenerated *MOLPAK* starting set	Electrostatic multipole plus empirical exp-6	Derivatives	Willock *et al.* (1995)
MOLPAK	Rigid	Patterns of coordination	Empirical energy	Derivatives	Holden *et al.* (1993)
MPA	Flexible	Random or rotational grid	Empirical energy	Derivatives	Williams (1999)
MSI-PP	Flexible	Random (Monte Carlo)	Empirical energy	Derivatives	Karfunkel *et al.* (1993), Verwer & Leusen (1998)
UPACK	Flexible	Grid or random	Empirical energy	Derivatives	van Eijck & Kroon (1999)
UPACK + ab initio	Flexible	Pre-generated starting set	*Ab initio* derived energy	Derivatives	Mooij, van Duijneveldt *et al.* (1999)
Zip-Promet	Rigid	Building into one-dimensional, two-dimensional and three-dimensional	Empirical energy	Simplex	Gavezzotti (1999)
FlexCryst	Rigid	Building into one-dimensional, two-dimensional and three-dimensional	Statistical potentials	Grid refinement	Hofmann & Lengauer (1997)
PackStar	Rigid	Random (SA)	Statistical fit	Simulated annealing	Lommerse & Motherwell (2000)
Rancel	Rigid	Random (GA)	Statistical fit	Genetic algorithm	Motherwell (1999)

At the time of the second blind test in 2001, the main problem for solving crystal structures had shifted from generation of stable crystal structures to selection of a likely structure from a large group of nearly equi-energetic possibilities. An optional secondary test of prediction was set at this time. The groups were also supplied with simulated X-ray powder diffraction patterns for the test molecules and submissions from this data were called 'powder-assisted predictions'.

The 2001 test methods used molecular mechanics or analogy with other CSD models to construct a 3D molecular model. The crystal structures were generated using various space groups and likely structures were selected using a preferred criterion, usually the calculated lattice energy. Volume/chemical intuition, density, morphology and elastic constraints were alternative selection criterion used in conjunction with CRYSTCA, MOLPAK and DMAREL programs respectively. The variety of force fields was more diverse in this workshop, ranging from simple atom-atom potentials to time intensive intermolecular potentials with some including polarization effects.

The summary of methods used are shown in table 4. The results supplied by each group were the top three energy rankings. From the searches, two of fifteen submissions predicted IV correctly with those results ranked 2 out of 3 and 3 out of 3 respectively. Four of fifteen submissions correctly predicted V with these results ranked 1, 1, 1 and 3 out of 3. None of the eleven submissions supplied a correct result for VI in their top three results. The correct structure was found sometimes but in a much higher ranked structure even though the energy difference was still very low. Hydrogen bonding potentials and

multipolar electrostatics were affected by small movements in molecule VI and this is presumably where the difficulties arose.

Contributor	Molecules attempted	Program/approach	Reference	Molecular model	Search generation
Methods employing lattice-energy minimization for generation of structures					
Gavezzotti	III, V	*ZIP-PROMET*	a	Rigid	Stepwise construction of dimers and layers
Schweizer & Dunitz	I, IV	*ZIP-PROMET*	a	Rigid	Stepwise construction of dimers and layers
Williams	I–VII	*MPA*	b	Flexible	Lattman grid systematic
Erk	IV–VI	*SySe* and *PP*	c	Flexible	Grid-based systematic
van Eijck	I, III–VII	*UPACK*	d	Flexible	Grid-based and random
Dzyabchenko	IV–VI	*PMC*	e	Flexible	Symmetry-adapted grid systematic
Schmidt	I–VI	*CRYSCA*	f	Flexible	Random plus steepest descent
Ammon	I–VI	*MOLPAK*	g	Rigid	Grid-based systematic
Price	I–V	*DMAREL*	h	Rigid	Using *MOLPAK*
Scheraga	IV–VI	*CRYSTALG*	i	Flexible	Conformation family Monte Carlo
Verwer & Leusen	I–III, VII	*Polymorph Predictor (PP)*	j	Flexible	Monte Carlo simulated annealing
Leusen	IV–VI	*Polymorph Predictor (PP)*	j	Flexible	Monte Carlo simulated annealing
Verwer	IV–VI	*Polymorph Predictor (PP)*	j	Flexible	Monte Carlo simulated annealing
Mooij	I, III, VII	Multipole crystal optimizer	k	Flexible	By van Eijck (*UPACK*)
Mooij	IV–VI	Multipole crystal optimizer	k	Flexible	By Leusen & Verwer (*PP*)
Methods based on statistical data from CSD					
Hofmann	I–III	*FlexCryst*	l	Rigid	Grid-based systematic
	IV–VI	*FlexCryst*	m	Rigid	Grid-based systematic
Lommerse	I–V, VII	*Packstar*	n	Rigid	Monte Carlo simulated annealing
Motherwell	I–V, VII	*Rancel*	o	Rigid	Genetic algorithm

Contributor	Lattice energy/fitness function		Other features used to select three submissions
	Electrostatic	Other	
Methods employing lattice-energy minimization for generation of structures			
Gavezzotti	None	Empirical	
Schweizer & Dunitz	Atom charges	6-exp	
Williams	Atom charges + extra sites	6-exp	
Erk	Atom charges	6-exp	
van Eijck	Atom charges	6-exp or 6–12	Free Energy
Dzyabchenko	Atom charges	6-exp or 6–12	
Schmidt	Atom charges	6-exp	Volume, chemical intuition
Ammon	Atom charges	6-exp	Density
Price	Atom multipoles	Empirical /derived	Morphology and elastic constants
Scheraga	Atom charges	6-exp or 6–12	
Verwer & Leusen	Atom charges	Dreiding 6–12	
Leusen	Atom charges	CVFF 6–12	
Verwer	Atom charges	Dreiding 6–12	
Mooij	Atom multipoles	*Ab initio* 6-exp + polarization	
Mooij	Atom multipoles	Dreiding 6-exp	
Methods based on statistical data from CSD			
Hofmann	Statistical potentials		
	Trained potentials		
Lommerse	CSD group contacts		
Motherwell	None	6-exp	Energy plus fitting of CSD contacts

Table 2.4; summary of methods used in CSP2001 [2.169]

Following the collaboration in 2001, the test in 2004 showed no significant changes in the first step of building a molecular structure from the chemical bonding diagram. The same two approaches were still being used. The first; treating the molecule as a rigid structure and calculating the starting geometry using gas-phase quantum chemical calculations. This assumes the crystal environment has no effect on bond lengths, angles and torsions. The second; adding conformational flexibility in the search, varying bond lengths, angles and torsions described within a suitable force-field.

This test did highlight some new algorithms for generating crystal structures were being employed. These new algorithms were taking advantage of the increased computing capability which had become available such as parallel computing. In general, and as in

previous tests, the methodologies included building up of dimers and layers, grid-based systematic searches, genetic algorithms, Monte Carlo and random searches. By their nature, random searches often exhibit gaps and clustering of points. Replacing the random number generator with a low-discrepancy sequence, these effects are reduced resulting in more efficient sampling (Sobol' sequences [2.174]). This smoothing is shown in Figure 17 and this method was followed by some participants in CSP2004.
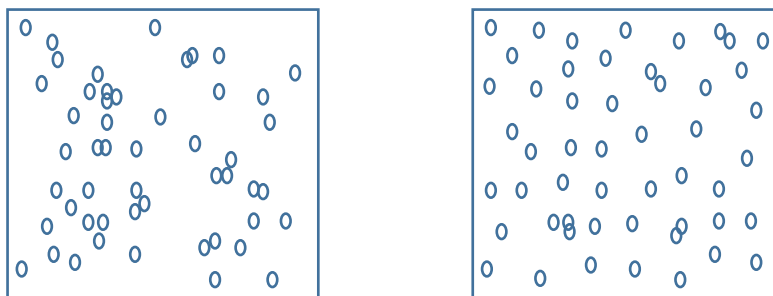

Figure 2.17; effects of gaps and clustering in a truly random search (left) compared with the more even effects of a quasi-random search (right).

In order to reduce necessary computing time, space-group symmetry if often used. Each space group and each number of molecules in the asymmetric unit would be treated as a unique search. Ideally, it would be possible to search all space groups with different numbers of molecules within the asymmetric unit. As this is too time consuming both to run and to process, clearly the choice of space group is an important consideration in the search methodology. Fortunately, structures of organic molecules tend to be found in only a handful of space groups and it is these that are commonly used. This is where parallel computing becomes a powerful tool. Networks of computers can be used in order to search more than the usual number of space-groups without taking an unrealistic amount of time. Alternatively, it is possible to carry out searches without considering space-group symmetry, keeping the space-group as P1 but varying the number of molecules in the asymmetric unit and adding space-group information in the final structures. In CSP2004, the majority of groups used around ten to fourteen space-groups with Z'=1, and 2. Three out of the eighteen participants only used P1 symmetry varying the number of molecules in the asymmetric unit and one group searched fifty-nine space-groups with Z'=1 or 2.

The differences in ranking the resulting structures was usually still based on lattice energy calculations and these vary between participants depending on the choice of model for inter- and intra-molecular energies. The entropic contributions to the free energy are

generally neglected although at this point, certain groups were introducing estimates of vibrational contributions to the overall free energy (Price et al [2.175]) and free energy minimization was also considered by Day [2.176], Della Valle and Venuti [2.177]. While some improvement to the thermodynamic model was achieved with the above considerations, it is clear that thermodynamics are not the only factors controlling the crystallization process. Kinetic contributions will also play a part. In CSP2004, two groups included mechanical properties and morphological calculations to re-rank the lattice energy predictions. Unfortunately these did not improve on the final result and in fact reduced the ranking of a submission which would otherwise have been correct.

Overall in CSP2004, no significant progress was found since the previous blind test CSP2001. Confidence had grown in the methodologies being developed, with significant interest in the PIXEL program [2.178] which quantified intermolecular interactions, however improvements were required in order to achieve better results for flexible molecules and structures with more than one molecule in the asymmetric unit.

In the blind tests before CSP2007, no single method showed success for general application. In the fourth test however, one method did achieve success for all four test structures with each submitted result being their top ranked result too. The review of CSP2007 also highlighted that the success rate for the flexible molecule was up on CSP2004 which in itself was a promising sign and that, as anticipated, the cocrystal had the lowest successful structures proposed.

The two methods of deriving the molecular structure had undergone little refinement compared to CSP2004. In the case of rigid molecules, it is considered highly unlikely that current CSP methods will fail due to a poor choice in the starting molecular structure. The success of a method tends to rely more on the generation of all possible crystal structures and the subsequent evaluation and ranking of these results. Therefore starting structures for CSP were still derived from either a force-field or from quantum-mechanics electronic structure calculations. The molecular models were treated as either rigid bodies, rigid bodies during searches with flexibility added during the minimization step or as flexible structures throughout the process.

There have been many different approaches to searching the energy landscape in order to generate trial crystal structures. Within CSP2007 the most popular method was to use

random or quasi-random variables which generated large numbers of trial structures. Other methods included Monte Carlo, genetic algorithms and systematic grid-based searches. The construction of crystal structures using dimers, chains and layers was also used by one group within CSP2007.

All but two participating groups used the most commonly found space groups for the crystal structure searches – the number of space groups used ranging from four to twenty. One of the two alternative approaches did not consider symmetry at all but varied the number of molecules in the asymmetric unit. The final approach was to consider all two hundred and thirty space groups for all four of the trial structures under investigation.

Ranking of results was based on the calculated lattice energy of structures generated from the searches. In this step, variations arose due to differences in the choice of model for calculating the crystal energy. Most participants used atomic charges in conjunction with an empirical force field, while others used quantum mechanics atomic multipoles with an empirical repulsion term. The limitations of using empirically fitted parameter sets is that a large dataset is usually required, which is fine for commonly found atoms but becomes more problematic for less common examples. Also, structural data in empirical parameterization is usually calculated at temperatures other than absolute zero. An alternative method for lattice energy calculations is periodic (solid state) DFT where all electron density and nuclear positions are optimized together. This allows the molecular conformation, and electron distribution, to change when changes are made to the crystalline environment.

Two of the participants also included lattice-dynamics contributions to their energy calculations. Molecule XV in this test (the cocrystal) was capable of forming strong hydrogen bonds. This non-energetic factor was also included in the assessment of possible results by comparing the predicted crystal structures with known structures of similar cocrystals.
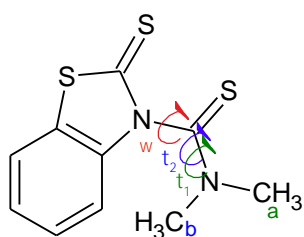
The molecular diagram of molecule (XIV) was shown in Figure 14 taken from [2.171], and is repeated here.

Two of the fourteen participants did not attempt molecule (XIV) and of the remaining twelve, three groups used the gas-phase minimum structure which was kept rigid throughout the search. Five participants used a combination of optimizing τ_1 and τ_2 choosing different angles of ω for multiple starting conformations. The conformations were kept rigid throughout the search but were then allowed to move during minimization steps. Finally, four groups treated all three torsions as flexible throughout searches and minimization. Intramolecular energy was calculated using either force fields such as Dreiding (Boerringer; Desiraju) or GAFF (Facelli), by derived torsional potentials (van Eijck, Price, Schmidt) or included in DFT calculations (Neumann, Scheraga).

Three participants (van Eijck; Neumann, Leusen & Kendrick; Price, Karamertzanis, Misquitta and Welch) predicted the observed crystal structure within their three submitted results for molecule (XIV). The observed crystal structure was also present in the extended lists of six other participating groups and three groups did not find the observed structure within their search. The rate of success for this flexible molecule was now comparable to the success rates for rigid molecules in previous tests and a significant improvement for previous results with flexible molecules.

Cocrystals had not been included in previous blind tests and of the fourteen participants, twelve attempted to predict the cocrystal structure with two groups (Neumann, Leusen & Kendrick and van Eijck) predicting the observed cocrystal structure. There were two different approaches to the searches. The first being to treat the positions and orientations of each molecule independently which will significantly increase the required search time. Alternatively the molecules may be considered as a dimer where intermolecular interactions influencing molecular orientations are considered prior to trial structure generation. This approach works well when there is the possibility of strong interactions between the two molecules, as in the case of molecule (XV). Five groups took the first approach, six groups the latter and one group performed a combination of the two ideas with starting points of dimers or independent molecules. The two groups finding the observed structure in their official submitted predictions treated both molecules independently in their search strategies. A further three groups also predicted the observed structure in their extended lists. Two of these groups had used the dimer approach and the third treated the molecules independently. It is not possible to say which approach is more successful however, since the submitted predictions and extended lists

of the seven remaining groups, also using either strategy, did not contain the observed structure.

CSP2007 had shown for the first time that a single method was capable of predicting the observed crystal and cocrystal structures of the rigid and flexible molecules set in this blind test. This was the approach used by Neumann, Leusen and Kendrick [2.179] who used random searches in all two hundred and thirty space groups, with a flexible molecular model. Plane-wave DFT was used to calculate lattice energy. Plane-waves are another way of representing electron wavefunctions. They consider all areas of space equally and are therefore independent of the type of crystal under consideration. Prior to this, basis sets had used localized functions, dependent on ion positions, such as Gaussians to represent atomic orbitals. It was known that when using DFT in energy calculations, the model used for the exchange correlation functional was very sensitive and was often poorly described. The two models most frequently used were the Local Density Approximation (LDA) or generalized gradient functionals (such as PW91). The first model would lead to over binding in crystal structures while the second would tend to under bind. In an attempt to improve the description of long range dispersion contributions to intermolecular interaction energies, the solution by Neumann and Perrin [2.180] was to augment DFT with an empirically derived additional dispersion (C_6R^{-6}) term. This was the DFT+D model which could be adjusted for the structure under investigation. Neumann, Leusen and Kendrick used a "Tailor Made Force Field" (TMFF), fitting an atom-atom force field specifically for each molecule during structure generation. The generated trial structures were then re-optimized using DFT+D. This method does come at a significant computational cost compared to other methodologies in CSP2007 and the process is shown in Figure 18.
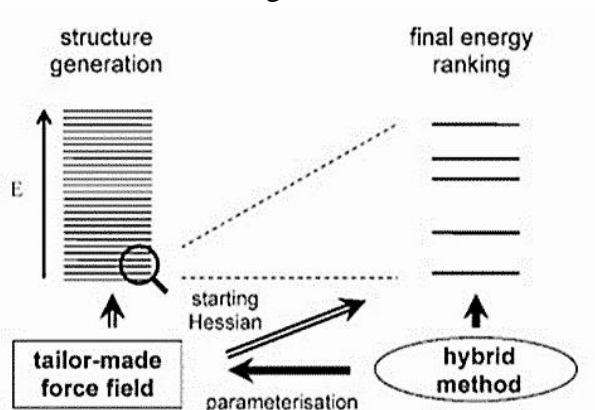

Figure 2.18; using the hybrid method with a TMFF for structure generation and subsequent ranking [2.181].

In CSP2007 crystal structures were generally still being ranked by calculated energies alone. The main use of non-energetic data for ranking purposes was the synthon-based approach (Desiraju and Thakur [2.182]) for the cocrystal, molecule (XV). Data from the CSD and from an in-house database was used to compare likely hydrogen bond motifs for the dimer. Using data from known crystal structures could be a way of including some kind of kinetic information in order to rank likely predicted structures. The predicted hydrogen bond motif was the same as the observed structure, however the choice of electrostatic potential was considered to be inadequate and the final prediction was not the observed cocrystal structure.

CSP2010 continued to test the methodologies employed for CSP to see if similar success rates could be obtained. Two new categories were introduced; a larger more flexible molecule and a hydrate with more than one known polymorph. Flexibility of the starting molecular structure was considered by the majority of participants whether it be in the crystal structure search and/or for energy minimization.

In CSP2010, similar methods for crystal structure generation were applied including random searches, quasi-random searches, genetic algorithms and modified Monte Carlo methods.

There was an increase in the number of participating groups using more than the most commonly found space groups in their crystal structure searches and more than one group in CSP2010 attempted to search all two hundred and thirty space groups. One participant did not consider symmetry at all, using P1 with varying Z (= 2, 4, 8).

Ranking of results was again based on the calculated lattice energy of structures generated from the searches. Extremely accurate calculations were required to differentiate between many structures which all lay within a few kJmol^{-1} to the global minimum. The results from this test showed that generic force fields could not provide this level of accuracy.

The DFT-D method had previously successfully produced accurate lattice energies (Neumann [2.181]). Improved force fields including distributed multipoles were also capable of improving energy results on minimization (Stone, [2.183]). However the limitations of these methods were highlighted during this test. Use of plane-wave *ab initio* methods for isolated ions in a vacuum lead to problems with force-field

parameterization. Generally, exp^{-6} potentials are fitted to crystal structures of neutral ions. It is this fitting which limits their use with molecular salts. Clearly it is important to design a methodology which suits the nature of the molecule under investigation.

In CSP2010, two participants successfully predicted the observed structure of molecule (XX). This was a large molecule with more degrees of flexibility than had been attempted in previous tests. The first method was referred to as Flexible CrystalPredictor-CrystalOptimizer (FCC) method; the second, referred to as Rigid CrystalPredictor-Molecular Mechanics method (RCM). The key features of these two programs are listed in table 5.

	FCC Method	RCM Method
Step 1. Conformational Analysis	Cambridge Structural Database (CSD) and DFT Specify 8 molecular models based on conformational regions, defined by ranges for the main torsions	Specify 48 molecular models, defined by distinct rigid conformations
Step 2. Crystal Structure Generation	CrystalPredictor and Clustering Flexible-molecule search in each conformational region, 12 common space groups, interpolated intramolecular energy, FIT+ESP charges for intermolecular forces 2,800,000 lattice energy minimizations 800,000 distinct structures generated 2000 CPU hours for intramolecular grid generation 16,000 CPU hours for search	Rigid-body search for each distinct conformation in 21 common space groups DFT(PCM, $\varepsilon = 3$)/W99 + ESP 9,600,000 lattice energy minimizations 93,000 distinct structures generated 580 CPU hours for DFT geometry optimization and ESP(PCM, $\varepsilon = 3$) charges 41,500 CPU hours for search
Step 3. Refinement of Energy Models and Clustering	CrystalOptimizer Re-minimization of stable structures using DFT-accuracy intramolecular energy and conformation-dependent DMA. 14 torsions and 5 bond angles explicitly optimized during lattice energy minimization; the rest of intramolecular degrees of freedom optimized within isolated-molecule wavefunction calculations. 1500 structures 100,000 CPU hours	MM and DMACRYS Re-minimization of stable structures using two different MM intramolecular energy models and Gasteiger-derived charges 10 torsions explicitly optimized 1: DREIDING 2: COMPASS Rigid-body minimization using DMA model Intramolecular energy obtained from DFT calculations using Polarizable Continuum Model ($\varepsilon = 3$) DFT(PCM, $\varepsilon = 3$)/W99+DMA 1500 structures 12,000 CPU hours
Step 4. Identification of 3 Structures for Submission for Blind Test	The submission structures were the two lowest in energy. The third submission was the lowest energy structure with a cis-amide.	The three lowest structures were chosen for submission. The ranking of the submitted structures was determined from their energies and a hydrogen bond propensity model.
Total CPU requirements	~120,000 CPU hours	~54,000 CPU hours
Further Computational Resource Details	Intel Xeon 5150 2.66 GHz processor, 1500 MB of memory DFT—GAUSSIAN 09 (Frisch et al., 2009) DMA—GDMA (Stone, 2005) Repulsion–dispersion potential—FIT (Coombes et al., 1996)	AMD Opteron 285 2.60 GHz processor, 256 MB of memory DFT—GAUSSIAN 03 (Frisch et al., 2003) DMA—GDMA (Stone, 2005) with (Stone, 1999) solver. Repulsion–dispersion potential—W99 (Williams, 2001)

Table 2.5; Details of the two methods applied to CSP2010 molecule (XX). [2.184]

As the table shows, each method follows the four key steps in CSP;
1) Conformational analysis.
2) Extensive crystal structure search.
3) Lattice energy minimization.
4) Examination of the lowest energy structure and other energetically feasible crystal structures.

FCC uses quantum mechanical scans during the first step of conformational analysis. This reduces the amount of search space to include only energetically meaningful regions.

This is combined with an analysis using the CSD for two purposes. Firstly, feasible torsion angles could be identified for the unknown molecule by comparing fragments of this molecule with similar structures from the CSD. Secondly, the likely number of molecules in the asymmetric units and common space groups could be examined through statistical analysis. During the crystal structure search, intramolecular energy was assessed using *ab initio* calculations for the flexible torsions and the semi-empirical AM1 for the rest of the molecule. Intermolecular electrostatic interactions were modeled using the charges derived from the intramolecular calculations and all intermolecular energy terms were derived using 'FIT' [2.185] an empirical exp-6 potential. For the refinement stage, CrystalOptimizer [2.186] was used to re-minimize all stable structures within 10kJmol^{-1} of the global minimum. Finally a clustering process was used to eliminate and reduce the number of possible solutions.

The RCM method also started by using the CSD for molecular structure assessment. However, since statistical data for the angle distribution of torsion R-SO$_2$ was insufficient, a DFT constrained geometry scan was carried out. This produced forty-eight starting conformations for the rigid search. Polarization of the charge density was considered which had previously been shown to have an influence on conformation energies and electrostatic interactions for polar, flexible molecules [2.187]. The forty-eight starting conformations were searched in the twenty-one most common space groups using CrystalPredictor with one rigid molecule in the asymmetric unit for all searches. The refinement method was carried out using two stages. First, using a Molecular Mechanics force field to optimize the molecular geometry. In this investigation, two separate force fields were used to compare the parameters and charges; COMPASS, using its own charge set, and DRIEDING, using Gasteiger derived charges. The second stage of re-optimization was to calculate accurate electrostatic interactions without changing molecular geometry. Single point DFT calculations and polarization effects were assessed to derive the intramolecular energy and atomic multipole moments for each conformation. Intermolecular terms were calculated using the empirical W99 potential [2.188]. Clustering was employed to remove duplicate structures and reduce the number of possible solutions. Further DFT and PCM minimization steps were carried out on the remaining top structures keeping the torsion geometries constant. Hydrogen bond motifs in the final structural models were assessed using an LHP (logit hydrogen-bonding propensity) model [also discussed in chapter 4 section 4.4.3] which resulted in a re-ordering of the final structures selected for submission.

Overall in CSP2010 there a slight dip in success rates compared with previous blind tests for the smaller, slightly flexible structures. This was offset however with the continued successes reported when using dispersion-corrected DFT. For the larger more flexible molecule however, the success of this method was limited with the observed structure being seventh in the submission list. The more successful methods for this molecule considered the problem in smaller more manageable pieces and mined the CSD for significant statistical data. The most considerable challenge in this test was in developing models to predict the hydrate (structure XXI). None of the groups submitted the observed structure in their top three predictions, the best ranked submission was outside the top three in twelfth place.

Going into the sixth blind test, CSP2016, it was summarized that the tests so far had highlighted many challenges which included the ranking of possible structures, the computational costs of CSP and whether there was a methodology to successfully predict all types of solid-form types.

Since the fifth blind test in 2010, methodologies had moved away from predicting 'the' crystal structure of a molecule and towards understanding the solid-form landscape of the generated structures [2.189]. Much work had also been invested in using CSP to aid solid-form screening of pharmaceuticals [2.190] and other effect chemicals. Using density-functional approximations to rank the results of crystal structure searches based on their stability had proved successful and progress had also been seen with these methods especially for molecular materials [2.180, 2.191]. The differences in the augmented DFT methods were usually found in the origin of the C_6 coefficients and which higher order terms were included. Progress had also been made with the CSP algorithms for flexible molecules, hydrates, multicomponent complexes and salts [2.192]. As such, in 2014, the sixth blind test was launched with a group of more realistic structures to highlight the current challenges still facing the developing methods of CSP.

CSP2016 contained five different categories which are shown in Figure 16. The main difference in submissions for this test was instead of providing the top three structures, the participants were required to submit a list of the top one hundred structures. The top structures were considered the most likely structures based on a chosen fitness function. Participants were also allowed to submit a second list containing the top one hundred structures generated from using a second, alternative fitness function. In order to test re-

ranking methods being developed by different groups, it was also permitted for participants to re-rank a list of submitted structures from a completely different group. There were two main approaches to generating the molecular starting geometry. The first was via *ab initio* calculations of the isolated molecule in the gas-phase. Torsions were investigated and predefined using information from the CSD. The second method was to employ force fields to explore molecular geometries and to optimize the resulting conformations using *ab initio* methods.

Along with the usual crystal structure generation methods (e.g. Monte Carlo simulated annealing, parallel tempering, systematic grid searches and quasi-random searches) a shape matching method was also used in CSP2016. This method matched target systems to known experimental structures in the CSD to generate possible results [2.193].

Most of the participating groups imposed crystallographic symmetry in their search methods, selecting the most likely space groups from the CSD.

The final ranking of the resulting structures was always carried out using energy data in this test. In a number of submissions, the initial evaluation of energy is carried out using a less intensive, and therefore faster, algorithm with the final minimization step using a more intensive method for the structures closest to the global minimum [2.179, 2.180]. Distributed multipole electrostatics potentials [2.183, 2.194], *ab initio* intramolecular energies [2.186, 2.192(a)] as well as dispersion-repulsion potentials were also utilized to rank generated structures. Force fields were also used as part of a ranking protocol. Here the generic force fields were improved with *ab initio* data [2.195, 2.196]. Further submissions used DFT calculations to derive potentials using symmetry-adapted perturbation theory [SAPT(DFT)], [2.197].

Normally, structures are re-ranked based on lattice energy but this does not take into account thermal motion or disorder. It has been shown that these effects may affect the overall ranking order of structures [2.198] and one of the groups did consider kinetic aspects in their methodology.

Post-processing techniques are common in CSP methods and often this results in many initial results converging to the same structure. This effect has been used to determine the completeness of the search [2.199] but in order to prevent repetition of processing,

methods are employed to assess and remove duplication of results. These include powder pattern similarity and fingerprint functions [2.200]. The CSD has also been used to filter results down to prevent duplicates appearing in the final list of results.

In general, CSP2016 search methods were seen to be using more exhaustive searches of the conformational space as well as being able to consider more space groups than in previous tests. The use of DFT in ranking results has increased significantly even since CSP2010 and the potentials being used in calculations are much more computationally demanding. It was concluded that when methods failed to find the observed structure it was down to using incorrect molecular conformations in the initial structure. By placing limitations or assumptions of intermolecular interactions also limited the success of the method. Searches of the conformational space which were not exhaustive enough were also found to miss reporting the observed structure. Lattice dynamics are routinely used to estimate vibrational contributions in the harmonic limit and were used by a handful of groups in this blind test, significantly in the reordering of results for molecule (XXIII). However, when the system deviates from harmonic oscillations, they become anharmonic vibrations which along with thermal expansion are not widely understood. Anharmonic lattice dynamics methods have been investigated [2.201] but are still in the relatively early stages of use. Disorder in crystals could be investigated further too as this could also assist in the estimation of thermodynamic stability [2.202, 2.190]. The additional use of these calculations could start to become computationally expensive however, and these may not be methods included in everyday use.

2.7 Cerius^2 – A Molecular Modelling Program

Cerius^2 is a molecular simulation package, which allows numerous calculations to be performed on molecules and crystals by applying parameters from predefined force fields. In this study, Cerius^2 was used in order to visualize molecules, build and examine crystal structures and to generate powder diffraction data. It was also used to calculate lattice energies and optimize these results using the crystal packer module.

MOPAC (Molecular Orbital PACkage) [2.203] is the semi-empirical molecular orbital package within Cerius^2. There are five distinct self-consistent field methods available within MOPAC, these are MINDO/3, MNDO, AM1, PM3 and MNDO-*d*. They all take into account electrostatic repulsion and exchange stabilization. MINDO/3 contains sets

of parameters for atoms and diatoms, whereas MNDO, AM1, PM3 and MNDO-*d* only use single-atom parameters [2.204 - 2.206].

As previously mentioned, force fields play an important role within the simulation of crystal structures and therefore within the Cerius^2 simulation package. In fact, it is the base from which most calculations take place. A force field provides various structural parameter constants like bond lengths, angles and energies that other modules draw from to perform the various operations and calculations. Quantum mechanics calculations are then able to perform geometry optimization on sketched drawings of the molecule. In this study, the AM1 method was used in order to calculate atomic positions.

The unit cell parameters may be defined and a crystal built using the crystal builder options, and symmetry operations applied using space group assignment. In this way, an entire unit cell may be built from scratch with lattice energy calculated using the crystal packer module. Crystal packer sums all Van der Waals interactions, coulombic charges, hydrogen bonding, internal rotations and hydrostatic pressure. However, in this study only the first three terms were considered. Before any calculations are carried out, the unit cell is 'initialised', which means that all rigid groups are defined.

Another option in this module is to minimise the packing arrangement. This is carried out using the minimiser option, defining any constraints before commencing the operation. Constraints include the unit cell parameters, translational and rotational variables and finally sub-rotations. Throughout this study, only translations and rotations were varied using the crystal packer option. Any other adjustments were either dealt with in the search itself (torsions) or in the refinement stage (unit cell parameters).

2.8   Crystal Morphology

Morphology is a term used to describe the sets of faces and edges enclosing a crystal. Crystal faces lie parallel to lattice planes and crystal edges are parallel to lattice lines. Miller indices may be used to describe a crystal face and the lattice point co-ordinates [uvw] used to describe a crystal edge. Crystal growth can be considered in two stages:

- Nucleation
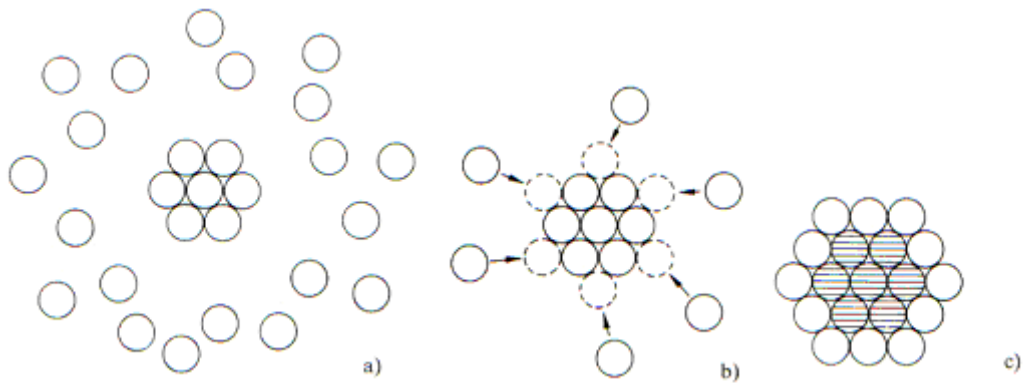
Figure 2.18; the initial stages of crystal growth when a few molecules have come together to form a nucleus (a), a three-dimensional structure (b). The crystal faces may already be observed but they are only a few unit cells in size (c).
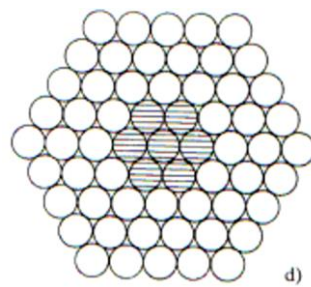
- Development of a nucleus into a crystal

Figure 2.19; the periodicity of the nucleus determines how attracted molecules are positioned, thus forming new lattice planes (d).

Temperature, pressure and the saturation of solution can have an effect on the rate of crystal growth [2.207]. Crystal faces which grow at faster rates than other faces will appear smaller.
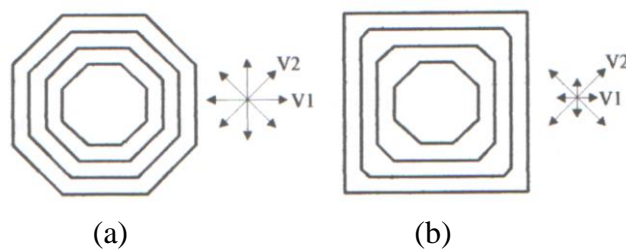
Figure 2.20; crystal growth – (a) the rates of different faces are similar and (b) where the rates vary. The growth rates are labelled v_1 and v_2.

Crystal shapes are defined by their 'habits', i.e. the relative sizes of the crystal faces. Essentially, there are three types of crystal habit. From left to right; equant, planar or tubular and prismatic or acicular (needle shaped).
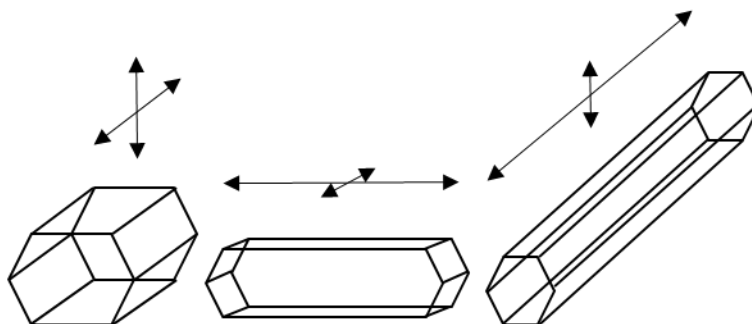
Figure 2.21; the three basic crystal habits. The arrows show the different growth rates and directions.

Although a number of methods to calculate crystal morphology exist [2.208], four are available in Cerius^2 and these are the Bravais-Friedel-Donnay-Harker (BFDH) method, the Attachment Energy method, the Surface Energy method and the Hartman-Perdok method [2.209]. The first of these two methods deduces the likely growth planes and their respective growth rates. It does not however take any consideration of the energetics of the system and is therefore simply an approximation. This method was also useful for predicting morphologies of ionic compounds but when it came to considering detailed interaction energies, like hydrogen bonds for example, the method failed to produce satisfactory results. Hartman and Perdok made some adjustments to the BFDH method by adding an intermolecular interaction component. Further modifications to this method resulted in the Attachment Energy method. Surface attachment energy was defined as the fraction of the total lattice energy released on the attachment of a growth slice to a growing crystal surface.

The crystal lattice energy was therefore defined using the attachment and slice energies;

$$E_{latt} = E_{att} + E_{sl} \tag{2.11}$$

In the HABIT95 program [2.210], an updated version of the Attachment Energy method, morphology predictions are allowed in the presence of impurity or solvent molecules [2.211 – 2.214]. Both can have an effect on the on the crystal surface during crystal growth. Lattice energies of molecular crystals are calculated using a summation

(averaging) of all atom-atom interactions up to a range where interactions become negligible. In other words interactions are summed until a convergence point is reached. In the HABIT95 program, the slice energy (the sum of all interactions lying within a defined layer of thickness d_{hkl}) is optimised until the most negative value is reached. The slice energy is optimised by moving the position of the slice but keeping the value of d_{hkl} and the position of origin constant.

2.9   Refinement of Trial Structure

In 1968, Rietveld published a method for the refinement of calculated crystal structures [2.215], which involved using the whole X-ray diffraction pattern instead of reducing the powder data into a 'single-crystal'-like pattern. This meant the problem of determining intensities of overlapping peaks was removed and all the information in the experimental profile could be utilized. In this study Rietveld refinement was carried out using the computer suite GSAS. As the name of the method implies, Rietveld is a refinement method and as such, a reasonable trial structure close to the global energy minimum is required. GSAS is used to optimize peak shape, peak positions and overall background of the simulated diffractogram to ensure the profile has as close fit as possible to the experimental data [2.216 – 2.218].

The refinement parameters are made up of two sets of data, the profile parameters, defining peak positions, widths, asymmetry and properties of the powder, and the structural parameters, defining the contents of the asymmetric unit cell.

The contributions of the profile parameters give rise to a near Gaussian peak shape for each Bragg peak. The Bragg peak shape in a profile y_i at position $2\theta_i$ can therefore be described as in equation (12) and is compared to a measured diffraction peak as shown in figure 22.

$$ y_i = I_k e^{[-4\ln\left(2/H_k^2\right)(2\theta_i - 2\theta_k)^2]} \tag{2.12}$$

where H_k is the 'full width at half maximum (FWHM) of the peak. $2\theta_k$ is the calculated position of the peak after it has been corrected for the zero-point shift of the counter. I_k is the calculated intensity of the peak, which includes the sum of the structure factor, Lorentz factor and reflection multiplicity. At any one point on y_i, multiple peaks may contribute to the profile. So I_k is really the sum of all peaks at the point $2\theta_i$
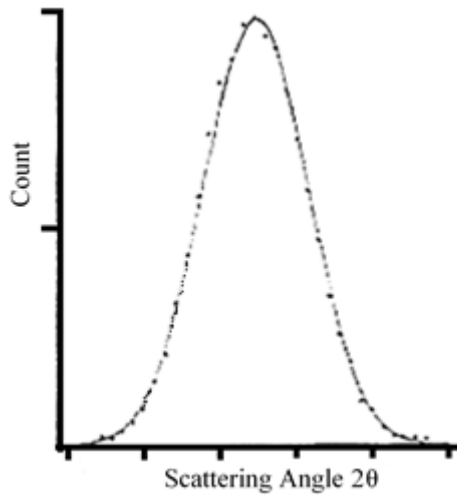
Figure 2.22; comparing a Gaussian peak shape distribution (solid line) with a measured diffraction peak (dotted line).

The Bragg peak can be affected by asymmetry at very low diffraction angles. This is known as a 'vertical divergence effect' causing the peak maximum to shift to a lower angle. This effect can be corrected in the Rietveld method using the asymmetry factor.

$$A_s = 1 - \left[\frac{sP(2\theta_i - 2\theta_k)^2}{\tan\theta_k} \right] \tag{2.13}$$

where P is the asymmetry factor and s is +1, 0, or –1 depending on the outcome of $2\theta_i$ - $2\theta_k$.
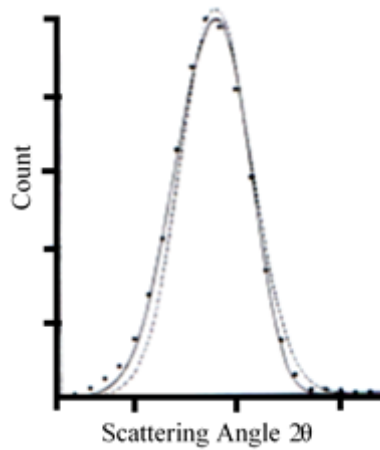
Figure 2.23; comparison of a measured diffraction peak showing asymmetry effects (dotted line), a symmetric Gaussian peak (solid line) and a corrected calculated peak (dashed line).

Bragg peaks have a tendency to broaden at higher Bragg angles. This angular dependence of Bragg peak half-widths has been defined by Caglioti, Paoletti and Ricci (referenced in [2.215]) and is defined in equation (14).

$$H_k^2 = U \tan^2 \theta_k + V \tan \theta_k + W \qquad (2.14)$$

U, V and W are the half-width parameters. This equation has also taken the effects caused by particle size (peak broadening) and scattering angle (variation of half-width) into consideration.
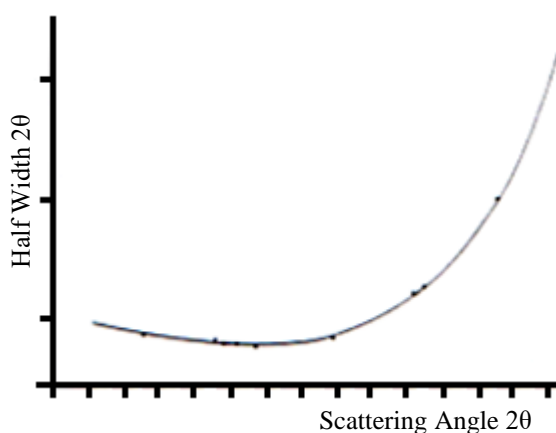
Figure 2.24; graph to show the variation of peak half width with Bragg angle 2θ. Measured half widths (dotted line) and calculated curve (solid line).

Preferred orientation in crystalline samples can cause variations in peak intensities that would be expected in true randomly distributed samples [2.3]. When these effects are not too pronounced, Rietveld determined a correction factor.

$$I_{corr} = I_{obs} e^{\left(-G\alpha^2\right)} \qquad (2.15)$$

I_{obs} is the intensity from random distribution, G is the preferred orientation parameter and α is the angle between the Bragg reflection and the preferred orientation direction.

Included in the terms for the structural parameters are terms for the structure factor, temperature dependency factors and the fractional co-ordinates.

For profile refinement, the quantity to be minimised using a least-squares method is described as a function 'M'. Approximate starting parameters are used for the initial cycles until a convergence criterion has been met [2.215].

$$M = \sum_i w_i \left[\left(y_{i(obs)} \right) - \left(\frac{1}{c} \right) y_{i(calc)} \right]^2 \tag{2.16}$$

Σ_i is the sum over the independent observations, w_i is the statistical weight factor and c is the overall scale factor to make $y_{calc} \equiv y_{obs}$. The function used in the Rietveld method compares the integrated intensities of calculated and observed peaks instead of profile intensities. The function is described by a weighted R-factor.

$$R_{wp} = \sqrt{\frac{\left(\sum_{i=1}^{N} w_i (y_{i(obs)} - y_{i(calc)})^2 \right)}{\left(\sum_{i=1}^{N} w_i (y_{i(obs)})^2 \right)}} \tag{2.17}$$

In programs supporting Rietveld refinement methods, this calculation is often adjusted to remove any background effects [2.219]. Recently, GSAS (including EXPGUI) has been extended into GSAS-II with changes to many Rietveld refinement parameters to improve peak fitting and rigid body fitting [2.220]. A further development for Rietveld refinement is in making it more accessible to more users with the toolkit SrRietveld. This has made the software more automated, faster and easier to use [2.221].

2.10   Differential Scanning Calorimetry

Differential Scanning Calorimetry is often used to detect phase transitions such as melting points in chemicals. An inert reference sample is heated alongside an experimental sample and both are heated to the same degree i.e. the same amount of energy is put into both samples. When the experimental sample undergoes a phase change like melting where it becomes a liquid, more energy or heat will be required for this sample than for the reference sample. The difference in applied heat is calculated and recorded [2.222].
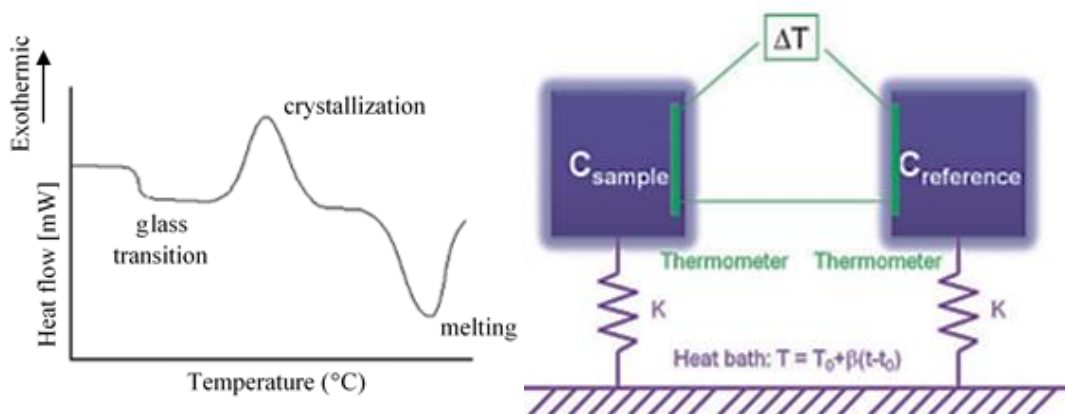
Figure 2.25; From left to right; showing the features recorded on a typical DSC graph and the measurement of sample phase change compared to a reference sample.

2.11 References

[2.1]   (a) R.B. Hammond, *Computationally Assisted Structure Determination for Molecular Materials from X-ray Powder Diffraction Data,* J. Phys. Chem. B., **101,** 6532-6536 (1997); (b) J. Brüning, *The Determination of Crystal Structures of Active Pharmaceutical Ingredients from X-ray Powder Diffraction Data: A Brief, Practical Introduction, with Fexofenadine Hydrochloride as an Example,* Journal of Pharmacy and Pharmacology, **67**, 773-781 (2015); (c) K. D. M. Harris, *Contemporary Advances in the Use of Powder X-Ray Diffraction for Structure Determination,* Angewandte Chemie Int. Ed., **40**(9), 1626-1651 (2001); (d) A. Meden, *Crystal Structure Solution from Powder Diffraction Data – State of the Art and Perspectives,* Croatica Chemica Acta, **71**(3), 615-633 (1998); (e) K. D. M. Harris, *Crystal Structure Determination from Powder Diffraction Data,* Chem. Mater., **8**, 2554-2570 (1996)

[2.2]   D. K. Smith, *X-ray Diffraction*, Anal. Chem., **54,** 156R-165R (1982)

[2.3]   (a) W. A. Dollase, *Correction of Intensities for Preferred Orientation in Powder Diffractometry: Application of the March Model,* J. Appl. Cryst., **19**, 267-272 (1986) (b) C. J. Gilmore, *X-ray Diffraction,* Solid State Characterization of Pharmaceuticals, R. A. Storey, I. Ymén, (Ed.), Chapter 2, 35-70 (2011); (c) G. H. Stout, *X-ray Structure Determination 2nd Edition A Practical Guide,* Wiley and Sons Inc. (1989)

[2.4]   P. M. De Wolff, *A Simplified Criterion for the Reliability of a Powder Pattern Indexing*, J. Appl. Cryst., **1,** 108-113 (1968)

[2.5]   G. S. Smith, *F_N: A Criterion for Rating Powder Diffraction Patterns and Evaluation the Reliability of Powder Pattern Indexing*, J. Appl. Cryst., **12,** 60-65 (1979)

[2.6]   P. E. Werner, *On the Determination of Unit-Cell Dimensions from Inaccurate Powder Diffraction Data*, J. Appl. Cryst., **9,** 216-219 (1976)

[2.7]   J. W. Visser, *A Fully Automatic Program for Finding the Unit Cell from Powder Data*, J. Appl. Cryst., **2,** 89-95 (1969)

[2.8]   F. Kohlbeck, *Indexing Program for Powder Patterns Especially Suitable for Triclinic, Monoclinic and Orthorhombic* Lattices, J. Appl. Cryst., **9,** 28-33 (1976)

[2.9]   A. Boultif, *Indexing of Powder Diffraction Patterns for Low-Symmetry Lattices by the Successive Dichotomy Method*, J. Appl. Cryst., **24,** 987-992 (1981)

[2.10] (a) P. E. Werner, *TREOR, a Semi-Exhaustive Trial-and-Error Powder Indexing Program for all Symmetries,* J. Appl. Cryst., **18**, 367-370 (1985); (b) P. E. Werner, *Documentation for TREOR90*, TREOR90.txt, January 1990

[2.11] D. Louer, *DICVOL91,* General Instructions, January 1992

[2.12] B. M. Kariuki, *A New Approach for Indexing Powder Diffraction Data Based on Whole-Profile Fitting and Global Optimization Using a Genetic Algorithm*, J. Synchrotron Rad. **6,** 87-92 (1999)

[2.13] Checkcell 2000 http://www.ccp14.ac.uk/tutorial/lmgp/

[2.14] A. A. Coelho, *Indexing of Powder Diffraction Patterns by Iterative use of Singular Value Decomposition,* J. Appl. Cryst., **36**, 86-95 (2003)

[2.15] M. A. Neumann, *X-Cell: A Novel Indexing Algorithm for Routine Tasks and Difficult Cases,* J. Appl. Cryst.,

[2.16] A. Le Bail, *Monte Carlo Indexing with McMaille,* Powder Diffraction, **19**(3), 249-254 (2004)

[2.17] (a) A. Boultif, *Powder Pattern Indexing with the Dichotomy Method,* J. Appl. Cryst., **37**, 724-731 (2004); (b) D. Louèr, *Some Further Considerations in Powder Diffraction Pattern Indexing with the Dichotomy Method,* Powder Diffraction, **29**(S2), S7-S12 (2014)

[2.18] S. Habershon, *Powder Diffraction Indexing as a Pattern Recognition Problem: A New Approach for Unit Cell Determination Based on an Artificial Neural Network*, J. Phys. Chem. A., **108,** 711-716 (2004)

[2.19] R. Oishi-Tomiyasu, *Robust Powder Auto-indexing Using Many Peaks,* J. Appl. Cryst., **47**, 593-598 (2014)

[2.20] B. H. Toby, *CMPR – A Powder Diffraction Toolkit,* J. Appl. Cryst., **38,**1040-1041 (2005)

[2.21] Crysfire 2006 http://www.ccp14.ac.uk/tutorial/crys/program/crysfire.htm

[2.22] M. Bortolotti, *ReX.Cell: A User-friendly Program for Powder Diffraction Indexing,* J. Appl. Cryst., **46**, 259-261 (2003)

[2.23] J. Bergmann, *Renewed Interest in Powder Diffraction Data Indexing,* Z. Kristallogr., **219**, 783-790 (2004)

[2.24] R. B. McClurg, *X-ray Powder-Diffraction Pattern Indexing for Pharmaceutical Applications,* Pharm. Tech., **January,** 56-61 (2013)

[2.25] W. I. F. David, *DASH: A Program for Crystal Structure Determination From Powder Diffraction Data,* J. Appl. Cryst., **39**, 910-915 (2006)

[2.26] E. Pidcock, *A Database Survey of Molecular and Crystallographic Symmetry,* Acta Cryst., **B59**, 634-640 (2003)

[2.27] G. Taylor, *The Phase Problem,* Acta Cryst., **D59**, 1881-1890 (2003)

[2.28] J. Jansen, *On the Determination of Accurate Intensities from Powder Diffraction Data. I. Whole-Pattern Fitting with a Least-Squares Procedure*, J. Appl. Cryst., **25**, 231-236 (1992)

[2.29] J. Jansen, *On the Determination of Accurate Intensities from Powder Diffraction Data. II. Estimation of Intensities of Overlapping Reflections*, J. Appl. Cryst., **25**, 237-243 (1992)

[2.30] G. Cascarano, *About the Treatment of Weak Reflections in Direct Procedures*, Acta Cryst., **A47,** 698-702 (1991)

[2.31] A. Altomere, *E-Map Improvements in Direct Procedures*, Acta Cryst., **A47,** 744-748 (1991)

[2.32] L. D Marks, *A Feasible Set Approach to the Crystallographic Phase Problem,* Acta Cryst., **A55**, 601-612 (1999)

[2.33] G. Cascarano, *Figures of Merit in Direct Methods: A New Point of View*, Acta Cryst., **A43,** 22-29 (1987)

[2.34] A. Altomare, *EXPO: A Program for Full Powder Pattern Decomposition and Crystal Structure Solution,* J. Appl. Cryst., **32**, 339-340 (1999)

[3.25] D.A. Langs, *Ab Initio Direct Methods. Practical Advice for Getting Beyond the First 300 Atoms*, Acta Cryst., **D49,** 158-167 (1993)

[2.36] H. Yonggeng, *STEP – a Trial-and-Error Procedure for Crystal Structure Determination. I. A Description of the Program System SYSTEM90*, Acta Cryst., **A50,** 748-753 (1994)

[2.37] A. Altomare, *Solving Crystal Structures from Powder Data. I. The Role of the Prior Information in the Two-Stage Method*, J. Appl. Cryst., **29**, 667-673 (1996)

[2.38] (a) C. C. Wilson, *Crystal Structure Determination from Low-Resolution X-ray Powder Diffraction Data*, Acta Cryst., **A46**, 258-262 (1990); (b) J. Rius, *Application of*

Patterson-function Direct Methods to Materials Charactersizzation, IUCrJ, **1**, 291-304 (2014)

[2.39] W. I. F. David, *Extending the Power of Powder Diffraction for Structure Determination*, Nature, **346** August, 731-734 (1990)

[2.40] G. Bricogne, *Maximum Entropy and the Foundations of Direct Methods*, Acta Cryst., **A40,** 410-445 (1984)

[2.41] C. J. Gilmore, *Applications of the Maximum Entropy Method to Powder Diffraction and Electron Crystallography*, Proc. R. Soc. London A, **442,** 97-111 (1993)

[2.42] G. Bricogne, *A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. I. Theory, Algorithms and Strategy*, Acta Cryst., **A46,** 284-297 (1990)

[2.43] G. Bricogne, *A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. III. Extension to Powder Diffraction Data*, Acta Cryst., **A47,** 803-829 (1991)

[2.44] M. Tremayne, *Application of the Combined Maximum Entropy and Likelihood Method to the Ab Initio Determination of an Organic Crystal Structure from X-ray Powder Diffraction Data*, J. Mater. Chem., **2**(12), 1301-1302 (1992)

[2.45] R. J. Read, *New Ways of Looking at Experimental Phasing,* Acta Cryst., **D59,** 1891-1902 (2003)

[2.46] (a) J. Wu, *Ab Initio Phasing of X-ray Powder Diffraction Patterns by Charge Flipping,* Nature Materials, Advance Online Publication, 9 July 2006 (1-6); (b) G. Oszlányi, *Ab Initio Structure Solution by Charge Flipping,* Acta Cryst., **A60**, 134-141 (2004), (c) G. Oszlányi, *Ab Initio Structure Solution by Charge Flipping II. Use of Weak Reflections,* Acta Cryst., **A61**, 147-152 (2005); (d) L. Palatinus, *The Charge-flipping Algorithm in Crystallography,* Acta Cryst., **B69**, 1-16 (2013)

[2.47] H. Xu, *Statistical Approach to the Phase Problem,* Acta Cryst., **A60**, 153-157 (2004)

[2.48] (a) R. Caliandro, *Protein Phasing at Non-atomic Resolution by Combining Patterson and VLD Techniques*, Acta Cryst., **D70**, 1994-2006 (2014); (b) M. C. Burla, *SIR2011: A New Package for Crystal Determination and Refinement,* J. Appl. Cryst., **45**, 357-361 (2012); M. C. Burla, *Crystal Structure Determination and Refinement via SIR2014,* J. Appl. Cryst., **48**, 306-309 (2015)

[2.49] R. Černý, *Direct Space Methods of Structure Determination from Powder Diffraction: Principles, Guidelines and Perspectives,* Z. Kristallogr., **222**, 105-113 (2007)

[2.50] C. W. Padgett, *Crystal Structures Elucidated from X-ray Powder Diffraction Data without Prior Indexing,* Cryst. Growth and Des., **7**(2), 367-372 (2007)

[2.51] S. Habermehl, *Structure Determination from Powder Data Without Prior Indexing, Using a Similarity Measure Based on Cross-Correlation Functions,* Acta Cryst., **B70**, 347-359 (2014)

[2.52] N. Metropolis, *Equation of State Calculations by Fast Computing Machines*, The Journal of Chemical Physics, **21**(6), 1087-1092 (1953)

[2.53] K. D. M. Harris., *Crystal Structure Determination from Powder Diffraction Data by Monte Carlo Methods*, J. Am. Chem. Soc., **116**, 3543-3547 (1994)

[2.54] E. Laurent, *Topochemical Rationalization of the Solid-State Polymerization Reaction of Sodium Chloroacetate: Structure Determination from Powder Diffraction Data by the Monte Carlo Method*, J. Phys. Chem. B., **101**, 8827-8831 (1997)

[2.55] M. Tremayne, *Solution of an Organic Crystal Structure from X-ray Powder Diffraction Data by a Generalized Rigid-Body Monte Carlo Method: Crystal Structure Determination of 1-methylfluorene*, J. Mater. Chem., **6**(9), 1601-1604 (1996)

[2.56] M. Tremayne, *The Development of Monte Carlo Methods for Crystal Structure Solution from Powder Diffraction Data: Simultaneous Translation and Rotation of a Structural Fragment within the Unit Cell*, J. Appl. Cryst., **29**, 211-214 (1996)

[2.57] C. R. A. Catlow, *Simulating and Predicting Crystal Structures,* Proc. R. Soc. Lond. A., **442**, 85-96 (1993)

[2.58] R. S. Payne, *Generation of Crystal Structures of Acetic Acid and its Halogenated Analogs,* J. Comput. Chem., **19**(1), 1-20 (1998)

[2.59] A. A. Coelho, *Whole-Profile Structure Solution from Powder Diffraction Data using Simulated Annealing,* J. Appl. Cryst., **33**, 899-908 (2000)

[2.60] H. Putz, *Combined Method for Ab-Initio Structure Solution from Powder Diffraction Data,* J. Appl. Cryst., **32**, 864-870 (1999)

[2.61] M. W. Deem, *Determination of 4-Connected Framework Crystal Structures by Simulated Annealing,* Nature, **342**, 260-262 (1989)

[2.62] Y. G. Andreev, *Ab Initio Solution of a Complex Crystal Structure from Powder-Diffraction Data using Simulated-Annealing Method and a High Degree of Moleculat Flexibility,* Phys. Rev., **B55**(18), 12011-12017 (1997)

[2.63] G. E. Engel, *PowderSolve – A Complete Package for Crystal Structure Solution from Powder Diffraction Data,* J. Appl. Cryst., **32**, 1169-1179 (1999)

[2.64] R. B. Hammond, *Applications of a Computational Systematic Search Strategy to Study Polymorphism in Phenazine and Perylene*, J. Phys. Chem. B., **103**, 7762-7770 (1999)

[2.65] E. D. L. Smith, *The Determination of the Crystal Structure of Anhydrous Theophylline by X-ray Powder Diffraction with a Systematic Search Algorithm, Lattice Energy Calculations, and ^{13}C and ^{15}N Solid-State NMR: A Question of Polymorphism in a Given Unit Cell*, J. Phys. Chem. B., **105**, 5818-5826 (2001)

[2.66] A. Gavazzotti, *Generation of Possible Crystal Structures from the Molecular Structure for Low-Polarity Organic Compounds*, J. Am. Chem. Soc., **113**, 4622-4629 (1991)

[2.67] A. Gavezzotti, *Computer Prediction of Organic Crystal Structures Using Partial X-ray Diffraction Data*, J. Am. Chem. Soc., **118,** 7153-7157 (1996)

[2.68] A. M. Chaka, *Predicting the Crystal Structure of Organic Molecular Materials*, Acta Cryst. **B52,** 165-183 (1996)

[2.69] F. H. Allen, *Comparison of Conformer Distributions in the Crystalline State with Conformational Energies Calculated by ab initio Techniques*, Journal of Computer-Aided Molecular Design, **10**, 247-254 (1996)

[2.70] B. M. Kariuki, *The Application of a Genetic Algorithm for Solving Crystal Structures from Powder Diffraction Data,* Chem. Phys. Lett., **280**, 189-195 (1997)

[2.71] K. Shankland, *Crystal Structure Determination from Powder Diffraction Data by the Application of a Genetic Algorithm,* Z. Kristallogr., **212**, 550-552 (1997)

[2.72] S. M. Woodley, *Crystal Structure Prediction from First Principles,* Nature Materials, **7**(December), 937-946 (2008)

[2.73] R. Orlando, *A New Massively Parallel Version of CRYSTAL for Large Systems on High Performance Computing Architectures,* J. Comput. Chem., **33**, 2276-2284 (2012)

[2.74] H. Nowell, *Validation of a Search Technique for Crystal Structure Prediction of Flexible Molecules by Application to Piracetam,* Acta Cryst., **B61,** 558-568 (2005)

[2.75] S. Goedecker, *Global Minimum Determination of the Born-Oppenheimer Surface within Density Functional Theory,* Phys. Rev. Lett., **95**, 55501-1 – 55501-4 (2005)

[2.76] V. Brodski, *Organa – A Program Package for Structure Determination from Powder Diffraction Data by Direct-Space Methods,* J. Appl. Cryst., **38**, 688-693 (2005)

[2.77] D. J. Earl, *Parallel Tempering; Theory, Applications and New Perspectives,* Phys. Chem. Chem. Phys., **7**, 3910-3916 (2005)

[2.78] R. Wehrens, *Small Molecule Geometry Optimization and Conformational Search,* Chapter 2 from Evolutionary Algorithms in Molecular Design, Ed D. E. Clark, 15-29 (2000)

[2.79] A. E. Eiben, *From Evolutionary Computation to the Evolution of Things,* Nature, **521**, 476-482 (2015)

[2.80] K. D. M. Harris, *Applications of Evolutionary Computation in Structure Determination from Diffraction Data,* Structure and Bonding, **110**, 55-94 (2004)

[2.81] Z. Zhou, *Multiple-fragment Representations of Molecular Geometry in Direct-space Structure Solution from Powder X-ray Diffraction Data using Genetic Algorithms,* Comput. Mat. Sci., **45**, 118-121 (2009)

[2.82] K. D. M. Harris, *Fundamentals and Applications of Genetic Algorithms for Structure Solution from Powder X-ray Diffraction Data,* Comput. Mat. Sci., **45**, 16-20 (2009)

[2.83] K. D. M. Harris, *Genetic Algorithms: A Universal Tool for Solving Computational Tasks in Materials Science,* Comput. Mat. Sci., **45**, ix-x (2009)

[2.84] A. Hossenian, *Genetic Algorithms for Structure Prediction of New Bithiazole Molecular Crystals: Methodology and Applications,* J. Math. Comput. Sci., **13**, 124-129 (2014)

[2.85] M. Valle, *Crystal Fingerprint Space – A Novel Paradigm for Studying Crystal-Structure Sets,* Acta Cryst., **A66**, 507-517 (2010)

[2.86] A. R. Oganov, *Crystal Structure Prediction Using ab initio Evolutionary Techniques,* J. Chem. Phys., **124**, 244704-1 – 244704-15 (2006)

[2.87] A. O. Lyakhov, *Crystal Structure Prediction Using Evolutionary Approach,* Chapter 7 from Modern Methods of Crystal Structure Prediction, Ed. R. R. Oganov, 147-169 (2011)

[2.88] G. Jones, *Genetic and Evolutionary Algorithms,* from Encyclopedia of Computational Chemistry, DOI: 10.1002/0470845015, (2002)

[2.89] (a) C. W. Glass, *USPEX – Evolutionary Crystal Structure Prediction,* Comput. Phys. Commun., **175**, 713-720 (2006); (b) A. O. Lyakhov, *New Developments in Evolutionary Structure Prediction Algorithm USPEX,* Comput. Phys. Commun., **184**, 1172-1182 (2013)

[2.90] Q. Zhu, *Contrained Evolutionary Algorithm for Structure Prediction of Molecular Crystals: Methodology and Applications,* Acta Cryst., **B68**, 215-226 (2012)

[2.91] S. Bahmann, *EVO – Evolutionary Algorithm for Crystal Structure Prediction,* Comput. Phys. Commun., **184**, 1618-1625 (2013)

[2.92] D. C. Lonie, *XTALOPT: An Open-Source Evolutionary Algorithm for Crystal Structure Prediction,* Comput. Phys. Commun., **182**, 372-387 (2011)

[2.93] R. Judson, *Genetic Algorithms and Their Use in Chemistry,* Chapter 1 from Reviews in Computational Chemistry, Volume 10, Ed. K. B. Lipkowitz and D. B. Boyd, 1-73 (1997)

[2.94] W. Paszkowicz, *Properties of a Genetic Algorithm Equipped with a Dynamic Penalty Function,* Comput. Mat. Sci., **45**, 77-83 (2009)

[2.95] (a) E. Akhmatskaya, *New Hybrid Monte Carlo Methods for Efficient Sampling: from Physics to Biology and Statistics,* Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo, October 17-21 (2010) (b) E. C. Neyts, *Combining Molecular Dynamics with Monte Carlo Simulations: Implementations and Applications,* Theor. Chem. Acc., **132**, 1320-1 – 1320-12 (2013)

[2.96] B. Escribano, *Multiple-Time-Stepping Generalized Hybrid Monte Carlo Methods,* J. Comput. Phys., **280**, 1-20 (2015)

[2.97] A. A. Javadi, *A Hybrid Intelligent Genetic Algorithm,* Adv. Eng. Informatics, **19**, 255-262 (2005)

[2.98] (a) V. Favre-Nicolin, *A Better FOX: Using Flexible Modelling and Maximum Likelihood to Improve Direct-Space ab initio Structure Determination from Powder Diffraction*, Z. Kreistallogr., **219**, 847-856 (2004); (b) P. Derollez, *Structure Determination of L-arabinitol by Powder X-ray Diffraction,* Acta Cryst., **B68**, 407-411 (2012)

[2.99] X. Lin, *Protein Structure Prediction with Local Adjust Tabu Search Algorithm,* BMC Bioinformatics, **15**(15), 1-11 (2014)

[2.100] S. Goedecker, *Global Minimum Determination of the Born-Oppenheimer Surface within Density Functional Theory,* Phys. Rev. Lett., **95**, (July 055501), 1-4 (2005)

[2.101] S. Goedecker, *Global Optimization with the Minima Hopping Method,* Chapter 6 from Modern Methods of Crystal Structure Prediction, Ed., A. R. Oganov, 131-145 (2011)

[2.102] S. Goedecker, *Minima Hopping: Searching for the Global Minimum of the Potential Energy Surface of Complex Molecular Systems without Invoking Thermodynamics,* J. Chem. Phys., **120**, 9911-9917 (2004)

[2.103] K. H. Sutherland, *Free Energy Basin-Hopping,* Chem. Phys. Lett., **625**, 1-4 (2015)

[2.104] L. Zhan, *Multicannonical Basin Hopping: A New Global Optimization Method for Complex Systems,* J. Chem. Phys., **120**(12), 5536-5542 (2004)

[2.105] D. J. Wales, *Energy Landscapes and Structure Prediction Using Basin-Hopping,* Chapter 2 from Modern Methods of Crystal Structure Prediction, Ed., A. R. Oganov, 29-54 (2011)

[2.106] L. Fábián, *Hydrogen-bond Motifs in the Crystals of Hydrophobic Amino Acids,* Acta Cryst., **B64**, 504-514 (2008)

[2.107] R. Taylor, *Rules Governing the Crystal Packing of Mono- and Dialcohols,* Acta Cryst., **B57**, 815-827 (2001)

[2.108] D. W. M. Hofmann, *Data Mining in Organic Crystallography,* Struct. Bond., **134**, 89-134 (2010)

[2.109] J. Apostolakis, *An Introduction to Data Mining,* Struct. Bond., **134**, 1-35 (2010)

[2.110] S. Habershon, *Development of a Multipopulation Parallel Genetic Algorithm for Structure Solution from Powder Diffraction Data,* J. Comput. Chem., **24**(14), 1767-1774 (2003)

[2.111] S. Kim, *Crystal Structure Prediction of Flexible Molecules Using Parallel Genetic Algorithms with a Standard Force Field,* J. Comput. Chem., **30**(13), 1973-1985 (2009)

[2.112] R. Storn, *Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces,* J. Global Opt., **11**, 341-359 (1997)

[2.113] T. Weber, *Determination and Refinement of Disordered Crystal Structures Using Evolutionary Algorithms in Combination with Monte Carlo Methods,* Acta Cryst., **A58,** 526-540 (2002)

[2.114] Y. Wang, *Crystal Structure Prediction via Particle-swarm Optimization,* Phys. Rev., **B82**, 94116-1 – 94116-8 (2010)

[2.115] P. Willett, *Chemical Structure Handling,* Chapter 7 from Evolutionary Algorithms in Molecular Design, Ed. D. E. Clark, 115-135 (2000)

[2.116] V. A. Saleev, *Hybrid Topology-Based Computational Approach for Crystal Structure Prediction,* Information Technology and Nanatechnology, 368-371 (2015)

[2.117] V. A. Blatov, *Periodic-Graph Approaches in Crystal Structure Prediction,* Chapter 1 from Modern Methods of Crystal Structure Prediction, Ed., A. R. Oganov, 1-28 (2011)

[2.118] P. Chang, *Dynamic Diversity Control in Genetic Algorithm for Mining Unsearched Solution Space in TSP Problems,* Expert Systems with Applications, **37**, 1863-1878

[2.119] Z. Liu, *MUSE: Multi-Algorithm Collaborative Crystal Structure Prediction,* Comp. Phys. Commun., **185**, 1893-1900 (2014)

[2.120]   F. A. Momany, *Intermolecular Potentials from Crystal Data. III. Determination of Empirical Potentials and Application to the Packing Configurations and Lattice Energies in Crystals of Hydrocarbons, Carboxylic Acids, Amines and Amides,* J. Phys. Chem., **78**(16), 1595-1620 (1974)

[2.121]   F. A. Momany, *Energy Parameters in Polypeptides. VII. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids,* J. Phys. Chem., **79**(22), 2361-2381 (1975)

[2.122]   G. Némethy, *Energy Parameters in Polypeptides. 9. Updating of Geometrical Parameters, Nonbonded Interactions, and Hydrogen Bond Interactions for the Naturally Occurring Amino Acids,* J. Phys. Chem., **87**(11), 1883-1887 (1983)

[2.123]   B. R. Brooks, *CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations,* J. Comput. Chem., **4**(2), 187-217 (1983)

[2.124]   (a) S. J. Weiner, *An All Atom Force Field for Simulations of Proteins and Nucleic Acids,* J. Comput. Chem., **7**(2), 230-252 (1986); (b) S. J. Weiner, *A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins,* J. Am. Chem. Soc., **106**, 765-784 (1984)

[2.125]   T. A. Halgren, *Maximally Diagonal Force Constants in Dependent Angle-Bending Coordinates. 2. Implications for the Design of Empirical Force Fields,* J. Am. Chem. Soc., **112**, 4710-4723 (1990)

[2.126]   S. L. Mayo, *DREIDING: A Generic Force Field For Molecular Simulations,* J. Phys. Chem., **94**(26), 8897-8909 (1990)

[2.127]   J. T. Sprague, *The MMP2 Calculational Model,* J. Comput Chem., **8**(5), 581-603 (1987)

[2.128]   A. K. Rappé, *UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations,* J. Am. Chem. Soc., **114**, 10024-10035 (1992)

[2.129]   C. J. Casewit, *Application of a Universal Force Field to Organic Molecules,* J. Am. Chem. Soc., **114**, 10035-10046 (1992)

[2.130]   T. A. Halgren, *Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters,* J. Am. Chem. Soc., **114**, 7827-7843 (1992)

[2.131]   K. D. Gibson, *Minimization of Polypeptide Energy, I. Preliminary Structures of Bovine Pancreatic Ribonuclease S-Peptide,* Proc. Natl. Acad. Sci., **58**, 420-426 (1967)

[2.132] A. T. Hagler, *Energy Functions for Peptides and Proteins. I. Derivation of a Consistent Force Field Including the Hydrogen Bond from Amide Crystals,* J. Am. Chem. Soc., **78**, 1595- (1974)

[2.133] L. S. Bartell, *Representations of Molecular Force Fields. 3. On Gauche Conformational Energy,* J. Am. Chem. Soc., **99**, 3279- (1977)

[2.134] L. Nilsson, *Empirical Energy Functions for Energy Minimization and Dynamics of Nucleic Acids,* J. Comput. Chem., **7**(2), 230- (1986)

[2.135] W. L. Jorgensen, *The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin,* J. Am. Chem. Soc., **110**, 1657-1665 (1988)

[2.136] J. C. Smith, *Empirical Force Field Study of Geometries and Conformational Transitions of Some Organic Molecules,* J. Am. Chem. Soc., **114**(3), 801-812 (1992)

[2.137] M. A. González, *Force Fields and Molecular Dynamics Simulations,* Collection SFN, **12**, 169-200 (2011) DOI: 10.1051/sfn/201112009

[2.138] K. Vanommeslaeghe, *Molecular Mechanics,* Curr. Pharm. Des., **20**(20), 3281-3292 (2014)

[2.139] E. O. Pyzer-Knapp, *An Optimized Intermolecular Force Field for Hydrogen-Bonded Organic Molecular Crystals Using Atomic Multipole Electrostatics,* Acta Cryst., **B72**, 477-487 (2016)

[2.140] A. Gavezzotti, *Geometry of the Intermolecular X-H···Y (X, Y = N, O) Hydrogen Bond and the Calibration of Empirical Hydrogen-Bond Potentials,* J. Phys. Chem., **98**, 4831-4837 (1994)

[2.141] P. Mörschel, *Prediction of Molecular Crystal Structures by a Crystallographic QM/MM Model with Full Space-Group Symmetry,* Acta Cryst., **A71**, 26-35 (2015)

[2.142] J. Gasteiger, *Iterative Partial Equalization of Orbital Electronegativity – A Rapid Access to Atomic Charges,* Tetrahedron, **36**, 3219-3288 (1980)

[2.143] (a) R. S. Muliken, *Electronic Population Analysis on LCAO-MO Molecule Wave Functions. I.* J. Chem. Phys., **23**, 1833-1840 (1955); (b) R. S. Muliken, *Electronic Population Analysis on LCAO-MO Molecular Wave Functions. II. Overlap Populations, Bond Orders, and Covalent Bond Energies,* J. Chem. Phys., **23**, 1841-1846 (1955); (c) R. S. Muliken, *Electronic Population Analysis on LCAO-MO Molecular Wave Functions. III. Effects of Hybridization on Overlap and Gross AO Populations,* J. Chem. Phys., **23**, 2338-2342 (1955); (d) R. S. Muliken, *Electronic Population Analysis on LCAO-MO Molecular Wave Functions. IV. Bonding and Antibonding in LCAO and Valence-Bond Theories,* J. Chem. Phys., **23**, 2343-2346 (1955)

[2.144] P. Politzer, *Properties of Atoms in Molecules. I. A Proposed Definition of the Charge on an Atom in a Molecule,* J. Am. Chem. Soc., **92**(22), 6451-6454 (1970)

[2.145] A. K. Rappé, *Charge Equilibrium for Molecular Dynamics Simulations,* J. Phys. Chem., **95**(8), 3358-3363 (1991)

[2.146] J. Homer, *Polyatomic London Dispersion Forces,* J. Chem Soc., Faraday Trans. 2, **83**(11), 1957-1974 (1987)

[2.147] D. E. Williams, *Failure of Net Atomic Charge Models to Represent the van der Waals Envelope Electric Potential of n-Alkanes,* J. Comput. Chem., **15**(7), 719-732 (1994)

[2.148] S. L. Price, *Applications of Realistic Electrostatic Modelling to Molecules in Complexes, Solids and Proteins,* J. Chem. Soc. Faraday Trans., **92**(17), 2997-3008 (1996)

[2.149] A. J. Stone, *Distributed Multipole Analysis: Stability for Large Basis Sets,* J. Chem. Theory Comput., **1**, 1128-1132 (2005)

[2.150] S. Cardamone, *Multipolar Electrostatics,* Phys. Chem. Chem. Phys., **16**, 10367-10387 (2014)

[2.151] J. Kong, *The Effects of Atomic Multipole Moments Obtained by the Potential-Derived Method on Hydrogen Bonding,* Int. J. Quantum Chem., **46**, 239-255 (1993)

[2.152] D. E. Williams, *Coulombic Interactions in Crystalline Hydrocarbons,* Acta Cryst., **A30**, 71-77 (1974)

[2.153] M. D. Beachy, *Accurate ab Initio Quantum Chemical Determination of the Relative Energetics of Peptide Conformations and Assessment of Empirical Force Fields,* J. Am. Chem. Soc., **119**, 5908-5920 (1997)

[2.154] S. L. Price, *Computational Prediction of Organic Crystal Structures and Polymorphism,* Int. Rev. Phys. Chem., **27**(3), 541-568 (2008)

[2.155] (a) H. M. Senn, *QM/MM Methods for Biomolecular Systems,* Angew. Chem. Int. Ed., **48**, 1198-1229 (2009); (b) D. Bakowies, *Semiempirical Treatment of Electrostatic Potentials and Partial Charges in Combined Quantum Mechanical and Molecular Mechanical Approaches,* J. Comput. Chem., **17**(1), 87-108 (1996); (c) M. J. Field, *A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations,* J. Comput. Chem., **11**(6), 700-733 (1990); (d) C. F. Fan, *Molecular Modeling of Polycarbonate. 1. Force Field, Static Structure, and Mechanical Properties,* Macromolecules, **27**, 2383-2391 (1994)

[2.156] J. D. Dunitz, *Toward a Quantitative Description of Crystal Packing in Terms of Molecular Pairs: Application to the Hexamorphic Crystal System 5-Methyl-2-[(2-*

nitrophenyl)amino]-3-thiophenecarbonitrile, Cryst. Growth and Des., **5**(6), 2810-2189 (2005)

[2.157] (a) A. Gavezotti, *Towards a Realistic Model for the Quantitative Evaluation of Intermolecular Potentials and for the Rationalization of Organic Crystal Structures. Part I. Philosophy,* Cryst. Eng. Comm., **5**(76), 429-438 (2003); (b) A. Gavezotti, *Towards a Realistic Model for the Quantitative Evaluation of Intermolecular Potentials and for the Rationalization of Organic Crystal Structures. Part II. Crystal Energy Landscapes,* Cryst. Eng. Comm., **5**(77), 439-446 (2003)

[2.158] Y. Shao, *Advances in Methods and Algorithms in a Modern Quantum Chemistry Program Package,* Phys. Chem. Chem. Phys., **8**, 3172-3191 (2006)

[2.159] J. Han, *Quantum Mechanical Force Field for Water with Explicit Electronic Polarization,* J. Chem. Phys., **139**, 54503-1 – 54503-18 (2013)

[2.160] S. R. Pruitt, *The Fragment Molecular Orbital and Systematic Molecular Fragmentation Methods Applied to Water Clusters,* Phys. Chem. Chem. Phys., **14**, 7752-7764 (2012)

[2.161] S. Grimme, *A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations,* J. Chem. Theory Comput., **10**, 4497-4514 (2014)

[2.162] N. M. Harrison, *An Introduction to Density Functional Theory,* Lecture

[2.163] C. H. Pham, *Crystal Structure Prediction of Molecular Crystals from First Principles: Are We There Yet?,* arXiv:1605.00733 [cond-mat.mtrl-sci] 3 May 2016

[2.164] A. D. Becke, *Perspective: Fifty Years of Density-Functional Theory in Chemical Physics,* J. Chem. Phys., **140**, 18A301-1 – 18A301-18 (2014)

[2.165] K. Burke, *DFT in a Nutshell,* Int. J. Quantum Chem., **113**, 96-101 (2013)

[2.166] (a) K. Berland, *van der Waals Forces in Density Functional Theory: A Review of the vdW-DF Method,* Rep. Prog. Phys., **78**, 66501-1 – 66501-41 (2015); (b) J. Klimeš, *Perspective: 50 Years of Density-functional Theory in Physics,* J. Chem. Phys., **140**, 18A301 (2014); (c) J. Gräfenstein, *An Efficient Algorithm for the Density-Functional Theory Treatment of Dispersion Interactions,* J. Chem. Phys., **130**, 124105-1 – 124105-16 (2009); (d) J. van de Streek, *Validation of Experimental Molecular Crystal Structures with Dispersion-Corrected Density Functional Theory Calculations,* Acta Cryst., **B66**, 544-558 (2010); (e) J. Nyman, *Accurate Force Fields and Methods for Modelling Organic Crystals at Finite Temperatures,* Phys. Chem. Chem. Phys., **18**, 15828-15837 (2016)

[2.167]   A. J. Cohen, *Insights into Current Limitations of Density Functional Theory,* Science, **321**,792-794 (2008)

[2.168]   J. P. M. Lommerse, *A Test of Crystal Structure Prediction of Small Organic Molecules,* Acta Cryst., **B56**, 697-714 (2000)

[2.169]   (a) W. D. S. Motherwell, *Crystal Structure Prediction of Small Organic Molecules: A Second Blind Test,* Acta Cryst., **B58**, 647-661 (2002); (b) J. A. R. P. Sarma, *The Supramolecular Synthon Approach to Crystal Structure Prediction,* Cryst. Growth Des., **2**(2), 93-100 (2002)

[2.170]   (a) G. M. Day, *A Third Blind Test of Crystal Structure Prediction,* Acta Cryst., **B61**, 511-527 (2005); (b) G. M. Day, *An Assessment of Lattice Energy Minimization for the Prediction of Small Organic Crystal Structures,* Cryst. Growth and Des., **4**(6), 1327-1340 (2004)

[2.171]   G. M. Day, *Significant Progress in Predicting the Crystal Structures of Small Organic Molecules – A Report on the Fourth Blind Test,* Acta Cryst., **B65**, 107-125 (2009)

[2.172]   D. A. Bardwell, *Towards Crystal Structure Prediction of Complex Organic Compounds – A Report on the Fifth Blind Test,* Acta Cryst., **B67**, 535-551 (2011)

[2.173]   A. M. Reilly, *Report on the Sixth Blind Test of Organic Crystal Structure Prediction Methods,* Acta Cryst., **B72**, 439-459 (2016)

[2.174]   W. H. Press, *Numerical Recipes in Fortran 77, The Art of Scientific Computing, Second Edition, Volume 1 of Fortran Numerical Recipes*, Cambridge University Press (1992)

[2.175]   http://www.ucl.ac.uk/~ucca17p/dmarelmanual/dmarel.html

[2.176]   G. M. Day, *A Nonempirical Anistropic Atom-Atom Model Potential for Chlorobenzene Crystals,* J. Am. Chem. Soc., **125,** 16434-16443 (2003)

[2.177]   R. G. D. Valle, *Inherent Structures of Crystalline Pentacene,* J. Chem. Phys., **118(2),** 807-815 (2003)

[2.178] (a) A. Gavezzotti, *Calculation of Intermolecular Interaction Energies by Direct Numerical Integration over Electron Densities. 1. Electrostatic and Polarization Energies in Molecular Crystals,* J. Phys. Chem., **106**, 4145-4154 (2002) (b) A. Gavezzotti, *Calculation of Intermolecular Interaction Energies by Direct Numerical Integration over Electron Densities. 2. An Improved Polarization Model and the Evaluation of Dispersion and Repulsion Energies,* J. Phys. Chem., **107**, 2344-2353 (2003)

[2.179]   M. A. Newmann, *A Major Advance in Crystal Structure Prediction,* Angew. Chem. Int. Ed., **47**, 2427-2430 (2008)

[2.180] M. A. Neumann, *Energy Ranking of Molecular Crystals Using Density Functional Theory Calculations and an Empirical van der Waals Correction,* J. Phys. Chem., **B109**, 15531-15541 (2005);

[2.181] M. A. Neumann, *Tailor-Made Force Fields for Crystal-Structure Prediction,* J. Phys. Chem. **B112**, 9810-9828 (2008)

[2.182] T. S. Thakur, *Crystal Structure Prediction of a Co-Crystal Using a Supramolecular Synthon Approach: 2-Methylbenzoic Acid – 2-Amino-4-methylpyrimidine,* Cryst. Growth and Des., **8(11),** 4031-4044 (20080

[2.183] A. J. Stone, *Distributed Multipole Analysis: Stability for Large Basis Sets,* J. Chem. Theory. Comput., **1,** 1128-1132 (2005)

[2.184] A.V. Kazantsev, *Successful Prediction of a Model Pharmaceutical in the Fifth Blind Test of Crystal Structure Prediction,* Int. J. Pharm., **418,** 168-178 (2011)

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

[2.186] A. V. Kazantsev, *Efficient Handling of Molecular Flexibility in Lattice Energy Minimization of Organic Solids,* J. Chem. Theory Comput., **7,** 1998-2016 (2011)

[2.187] T. G. Cooper, *Molecular Polarization Effects on the Relative Energies of the Real and Putative Crystal Structures of Valine,* J. Chem. Theory Comput., **4,** 1795-1805 (2008)

[2.188] D. E. Williams, *Improved Intermolecular Force Field for Molecules Containing H, C, N, and O Atoms, with Application to Nucleoside and Peptide Crystals,* J. Comput. Chem., **22(11),** 1154-1166 (2001)

[2.189] (a) S. L. Price, *Why Don't we Find More Polymorphs?,* Acta Cryst., **B69,** 313-328 (2013); (b) S. Z. Ismail., *Evaluating a Crystal Energy Landscape in the Context of Industrial Polymorph Screening,* Cryst. Growth and Des., **13,** 2396-2406 (2013); (c) S. L. Price, *Predicting Structures of Organic Compounds,* Chem. Soc. Rev., **43,** 2098-2111 (2014); (d) S. L. Price, *Can Computed Crystal Energy Landscapes Help Understand Pharmaceutical Solids?,* Chem. Commun., **52,** 7065-7077 (2016); (e) H. P. G. Thompson, *Which Conformations make Stable Crystal Structures? Mapping Crystalline Molecular Geometries to the Conformational Energy Landscape,* Chem. Sci., **5,** 3173-3182 (2014)

[2.190] M. A. Neumann, *Combined Crystal Structure Prediction and High Pressure Crystallization in Rational Pharmaceutical Polymorph Screening,* Natur. Comm., DOI: 10.1038/ncomms8793 (July 2015)

[2.191] (a) A. M. Reilly, *van der Waals Dispersion Interactions in Molecular Materials: Beyond Pairwise Activity,* Chem. Sci., **6,** 3289-3301 (2015); (b) S. Grimme, *Semiempirical GGA-Type Density Functional Constructed with a Long Range Dispersion Correction,* J. Comput. Chem., **15,** 1787-1799 (2006); (c) A. Tkatchenko, *Accurate and Efficient Method for Many-Body van der Waals Interactions,* Phys. Rev. Lett., **108,** 236402-1-236402-5 (2012); (d) J. Klimeš, *van der Waals Density Functionals Applied to Solids,* Phys. Rev., **B83,** 195131-1-195131-13 (2011); (e) J. G. Brandenburg, *Organic Crystal Polymorphism: A Benchmark for Dispersion-Corrected Mean-Field Electronic Structure Methods,* Acta Cryst., **B72,** 502-513 (2016)

[2.192] (a) M. Habgood, *Efficient Handling of Molecular Flexibility in Ab Initio Generation of Crystal Structures,* J. Chem. Theory Comput., **11,** 1957-1969 (2015); (b) J. Kendrick, *Crystal Structure Prediction of a Flexible Molecule of Pharmaceutical Interest with Unusual Polymorphic Behavior,* Cryst. Growth and Des., **13,** 581-589 (2013); (c) D. E. Braun, *Which, if any, Hydrates will Crystallize? Predicting Hydrate Formation of Two Dihydroxybenzoic Acids,* Chem. Commun., **47,** 5443-5445 (2011); (d) M. Habgood, *Isomers, Conformers, and Cocrystal Stoichiometry: Insights from the Crystal Energy Landscape of Caffeine with the Hydroxybenzoic Acids,* Cryst. Growth and Des., **10,** 3263-3272 (2010); (e) S. Mohamed, *Computational Prediction of Salt and Cocrystal Structures – Does a Proton Position Matter?,* Int. J. Pharm., **418,** 187-198 (2011)

[2.193] J. C. Cole, *Generation of Crystal Structures Using Known Structures as Analogues,* Acta Cryst., **B72,** 530-541 (2016)

[2.194] S. L. Price, *Modelling Organic Crystal Structures Using Distributed Multipole and Polarizability-Based Model Intermolecular Potentials,* Phys. Chem. Chem. Phys., **12,** 8478-8490 (2010)

[2.195] B. P. van Eijck, *Ab Initio Crystal Structure Predictions for Flexible Hydrogen-Bonded Molecules. Part II. Accurate Energy Minimization,* J. Comp. Chem., **22(8),** 805-815 (2001)

[2.196] (a) D. M. Elking, *Crystal Structure Prediction of Rigid Molecules,* Acta Cryst., **B72,** 488-501 (2016); (b) E. O. Pyzer-Knapp, *An Optimized Intermolecular Force Field for Hydrogen-Bonded Organic Molecular Crystals Using Atomic Multipole Electrostatics,* Acta Cryst., **B72,** 477-487 (2016); (c) A. Broo, *Transferable Force Field for Crystal Structure Predictions, Investigation of Performance and Exploration of Different Rescoring Strategies Using DFT-D Methods,* Acta Cryst., **B72,** 460-476 (2016)

[2.197] A. J. Misquitta, *Intermolecular Potentials Based on Symmetry-Adapted Perturbation Theory with Dispersion Energies from Time-Dependent Density-Functional Calculations,* J. Chem. Phys., **123,** 214103-1-214103-14 (2005)

[2.198] (a) B. P. van Eijck, *Crystal Structure for Six Monosaccharides Revisited,* J. Phys. Chem. B, **105,** 10573-10578 (2001); (b) A. M. Reilly, *Role of Dispersion Interactions in the Polymorphism and Entropic Stabilization of the Aspirin Crystal,* Phys. Rev. Lett., **113,** 055701-1-055701-5 (2014); (c) J. Nyman, *Static and Lattic Vibrational Energy Differences Between Polymorphs,* Cryst. Eng. Comm., **17,** 5154-5165 (2015)

[2.199] D. H. Case, *Convergence Properties of Crystal Structure Prediction by Quasi-Random Sampling,* J. Chem. Theory Compu., **12,** 910-924 (2016)

[2.200] (a) D. W. M. Hofmann, *New Similarity Index for Crystal Structure Determination from X-ray Powder Diagrams,* J. Appl. Cryst., **38,** 861-866 (2005); (b) A. R. Oganov, *How to Quantify Energy Lanscapes of Solids,* J. Chem. Phys., **130,** 104504-1-104504-9, (2009)

[2.201] B. Monserrat, *Anharmonic Vibrational Properties in Periodic Systems: Energy, Electron-Phonon Coupling and Stress,* Phys. Rev. B, **87,** 144302-1-144302-10 (2013)

[2.202] M. Habgood, *Substitutional and Orientational Disorder in Organic Crystals: A Symmetry-Adapted Ensemble Model,* Phys. Chem. Chem. Phys., **13,** 9590-9600 (2011)

[2.203] http://openmopac.net/

[2.204] H. Dorsett, *Overview of Molecular Modelling and Ab initio Molecular Orbital Methods Suitable for Use with Energetic Materials,* DSTO-GD-0253 (2000)

[2.205] M. J. S. Dewar, *Development and Status of MINDO/3 and MNDO,* J. Mol. Struc., **100**, 41-50 1983)

[2.206] M. S. J. Dewar, *AM1: A New General Purpose Quantum Mechanical Molecular Model,* J. Am. Chem. Soc., **107**, 3902-3909 (1985)

[2.207] Z. Berkovitch-Yellin, *Toward an ab Initio Derivation of Crystal Morphology,* **107**, 8239-8253 (1985)

[2.208] P. Bennema, *Crystal Growth and Morphology: New Developments in an Integrated Hartman-Perdok-Connected Net-Roughening Transition Theory, Supported by Computer Simulations,* Cryst. Growth and Des., **4**(5), 905-913 (2004)

[2.209] P. Hartman, *On the Relationship between Structure and Morphology of Crystals. I*, Acta Cryst., **8**, 49-52 (1954)

[2.210] G. Clydesdale, HABIT98 – A Program for Predicting the Morphology of Molecular Crystals as a Function of the Growth Environment, Program Manual

[2.211] C. Stoica, *Understanding the Effect of a Solvent on the Crystal Habit,* Cryst. Growth and Des., **4**(4), 765-768 (2004)

[2.212] H. M. Cuppen, *Sensitivity of Morphology Prediction to the Force Field: Paracetamol as an Example,* Cryst. Growth and Des., **4(6),** 1341-1349 (2004)

[2.213] R. I. Ristic, *Macro- and Micromorphology of Monoclinic Paracetamol Grown from Pure Aqueous Solution,* J. Phys. Chem., **B105**, 9057-9066 (2001)

[2.214] G. Clydesdale, *A Molecular Modeling Study of the Crystal Morphology of Adipic Acid and its Habit Modification by Homologous Impurities,* Cryrt. Growth and Des., **5**(6), 2154-2163 (2005)

[2.215] H. M. Rietveld, *A Profile Refinement Method for Nuclear and Magnetic Structures,* J. Appl. Cryst., **2**, 65-71 (1969)

[2.216] L. B. McCusker, *Rietveld Refinement Guidelines,* J. Appl. Cryst., **32**, 36-50 (1999)

[2.217] H. Manohar, *The New Potential of Powder Diffraction. Structure Determination and Refinement Using Profile Analysis Technique,* Curr. Sci., **52**(2), 39-44 (1983)

[2.218] K. Shankland, *Global Rietveld Refinement,* J. Res. Nat. Inst. Stand. Technol., **109**, 143-154 (2004)

[2.219] B. Toby, *R Factor in Rietveld Analysis: How Good is Good Enough?,* Powder Diffr., **21**(1), 67-70 (2006)

[2.220] B. Toby, *What's New In GSAS-II,* Powder Diffr., **29**(S2), S2-S6 (2014)

[2.221] P. Tian, *SrRietveld: A Program for Automating Rietveld Refinements for High-Throughput Powder Diffraction Studies,* J. Appl. Cryst., **46**, 255-258 (2013)

[2.222] B. Bushan (ed.) *Differential Scanning Calorimetry,* Encyclopedia of Nanotechnology, 1496-1497 (2012) DOI 10. 1007/978-90-481-9751-4

CHAPTER 3 – USING TWO POLYMORPHS OF TWO KNOWN CRYSTAL STRUCTURES, CARBAMAZEPINE AND PRIMIDONE TO VALIDATE THE SYSTEMATIC AND SIMULATED ANNEALING SEARCH METHODS.

3.1 Introduction

A systematic search method of small unit cells for rigid bodies had already proved successful [3.1, 3.2]. Its success lay mostly in the fact that the volumes being searched were small, ensuring a step-wise search method would be thorough without becoming excessive in terms of computational time. The random nature of alternative search methods might not find the same number of energy minima as a systematic search and, initially, have the potential to miss the global energy minimum. However, when unit cell volumes increase, the advantages of random search methods outweigh the systematic search method as the latter becomes too time-consuming.

Prior to this research, the systematic search program had not been tested using flexible molecules. This additional component of the search had the potential to increase computational time but it was not known to what extent.

At this time, a new 'in-house' simulated annealing search method had been developed [3.3] and required validation. The simulated annealing method is described in chapter 2. This method is essentially a random search method using temperature ramps to 'melt' and 'freeze' a molecule in a certain conformation based on how acceptable the energy of that structure is.

The aim of the work in this chapter was threefold:

- To extend the systematic search to include flexible structures and evaluate its success.
- To validate a new simulated annealing method using the same flexible structures.
- A comparison of the results generated from both searches to determine which may be the preferred method.

The pharmaceutical compounds carbamazepine and primidone were chosen to use in this study. Carbamazepine has one degree of freedom and primidone has two. The structures also have the ability to form intermolecular hydrogen bonds within the unit cell. This feature would also provide a good test of the Dreiding Force Field which was used in all searches.

3.1.1   *Carbamazepine*

Carbamazepine (5*H*-Dibenz[*b,f*]azepine-5-carboxamide) was first synthesized by Schindler at Geigy in 1953 [3.4]. It was designed to alleviate the effects of seizures by blocking the sodium channels [3.5, 3.6]. Clinical trials on carbamazepine began in the early 1960's and it was finally introduced as a prescribed drug by the end of that decade. The molecular structure of carbamazepine is shown in figure 1.
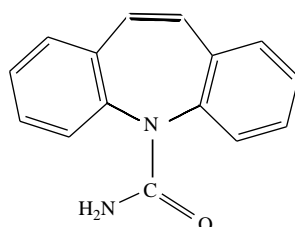
Figure 3.1; the molecular structure of carbamazepine.

In 1997, two polymorphs of carbamazepine already had full crystal structure data recorded. These were the monoclinic [3.7, 3.8] and the trigonal [3.9] forms. The unit cell parameters of the triclinic form had also been calculated at this stage, but the crystals were too thin for the atomic positions to be determined [3.10]. In 2003 however, the full crystal structure of triclinic carbamazepine was published from single crystal data [3.11]. A C-centred monoclinic structure of carbamazepine was also reported in 2002 [3.12]. The structures of numerous pseudopolymorphs have been solved including the acetone and dioxane solvates [3.13, 3.14] and dihydrate modifications 1 and 2 [3.15 - 3.17]. The decomposition product iminostilbene also had structure data published [3.18]. There have been many studies carried out on the physicochemical properties of the various forms of carbamazepine, which have helped understand the relationship between chemical form and subsequent effect on such properties [3.19 – 3.27].

Figure 2 shows how the four anhydrous polymorphic forms of carbamazepine can be prepared from commercial carbamazepine [3.11].
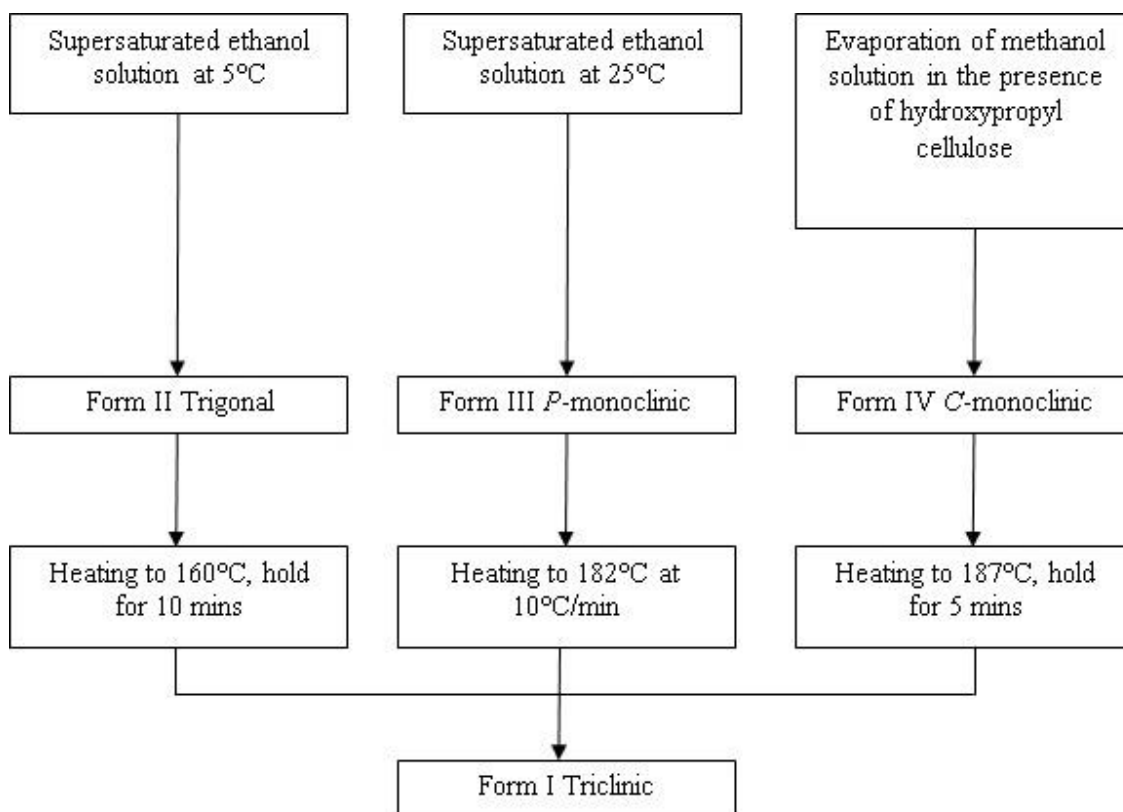
Figure 3.2; preparation methods of the anhydrous forms of carbamazepine.

In this study, the monoclinic and trigonal polymorphs of carbamazepine were used. It was found that the trigonal molecular configuration differed quite significantly with the monoclinic form and this is described in more detail later in this chapter.

3.1.2 Primidone

Primidone (5-ethyl-5-phenyl-hexahydro-pyrimidine-4,6,dione) has a molecular structure similar to the barbiturate phenobarbitone. In the 1950's, Phenobarbitone was an established anti-epileptic drug when primidone was introduced as an anticonvulsant [3.28, 3.29]. Over time however, the introduction to the market of more effective alternatives with fewer side effects reduced the demand of this product [3.30]. The molecular structure of primidone is shown in figure 3.
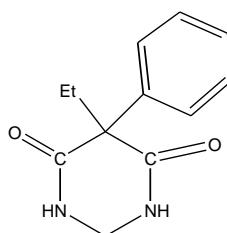
Figure 3.3; the molecular structure of primidone.

The structure of the most stable polymorph of primidone was recorded in 1975 by Yeates and Palmer [3.31] using direct methods and refined using full-matrix least-squares calculations. Primidone B, the less stable form, was also isolated and characterised in the early 1970's by Daley 1973 [3.32], Summers and Enver 1976 [3.33]. In 1996, the mechanical properties of the two polymorphs were discussed [3.34] and in 1999, the full crystal structure of the second polymorph, primidone B, was published [3.3, 3.35].

3.2 The Hydrogen Bond

The nature of the hydrogen bond has been debated since the early 1900's. Initial research attempted to describe the hydrogen bond as a purely electrostatic interaction [3.36]. However, subsequent examinations from the 1930's onwards have modified this approach to incorporate the possibility of bonding to exist that is more covalent in nature depending on the distance between the contributing atoms [3.37]. Valence bond theories put forward around the 1950's and 1960's [3.38] stated that when distances between the two contributing atoms are long, the interactions are predominantly electrostatic, but when the distances are shorter, repulsion and delocalization also have an effect. More recent research however, has stated that evidence now exists in some cases for describing the hydrogen bond as almost a covalent bond [3.37, 3.39 - 3.40]. In 2011, IUPAC issued a provisional report on the definition of the hydrogen bond [3.41].

For the purpose of this study however, the hydrogen bond is understood to be an electrostatic interaction between a slightly positive donor atom and a slightly negatively charged acceptor atom (δ^+ -- δ^-). As the strength of the covalent bond determines the stability of the molecule, the strength of the hydrogen bond determines the stability of the molecular assembly.
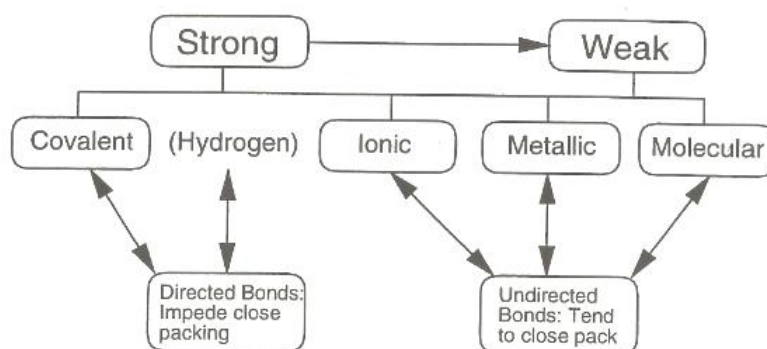
Figure 3.4 (a); representation of the relative strengths of different bonds.
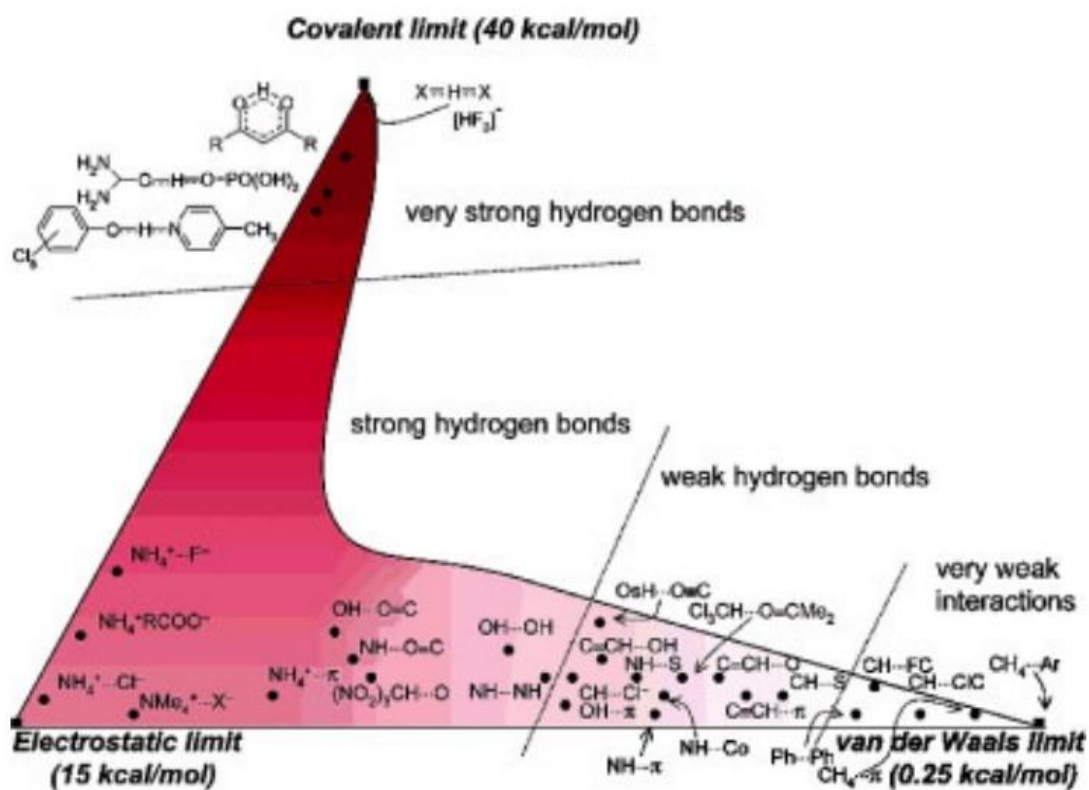
Figure 3.4 (b); representation of the relative strengths and composite nature of different interactions. This image is referenced from [3.53 and 3.55]. Shading of the diagram aims to give a visual guide to the energy scale.

The nature of different hydrogen bonds environments have been examined within literature [3.42-3.55] and various forms of hydrogen bonding dimers have been found [3.43].
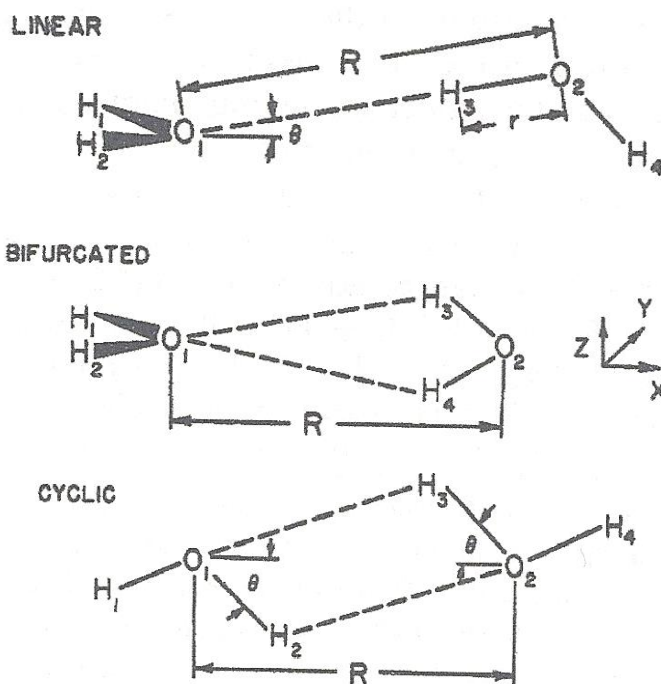
Figure 3.5; Reference from [3.55](b).

Linear: H_1 and H_2 are in the *xy* plane while H_3 and H_4 are in the *xz* plane.

Bifurcated: H_1 and H_2 are in the *xy* plane while H_3 and H_4 are in the *xz* plane.

Cyclic: All atoms are in the *xz* plane.

R is the distance between the oxygen atoms and ranges from 2.53Å to 3.00Å.

θ ranges from 0° to 57°.

The data stored within the Cambridge Structural Database (CSD) provides an opportunity to retrieve and examine vast quantities of information [3.56] concerning the geometry of the hydrogen bond which lead to the view that some directionality exists. These statistics have affected the design of some force fields in that geometrical restrictions have been included in the hydrogen bonding term. Describing hydrogen bond systems can become very complicated when numerous donor/acceptor sites are available. Attempts to isolate the basic geometric and energetic features of single hydrogen bonding systems have led to a classification system being developed by Etter, and MacDonald [3.56]. In both studies it was noted, however, that the exact position of hydrogen atoms using x-ray powder diffraction was problematic purely because of the nature of the hydrogen atom having very little electron density. This does not pose much of a problem in rigid structures as the hydrogen positions may be calculated using geometric calculations. The issue becomes considerable when examples are used that will tend to assume positions depending on their environment. Neutron analysis has in the past proved much more useful for such studies [3.57] but at the time of this work, it was not readily available.

As would be expected, it is the nature of the non-hydrogen atom which has a significant effect on the bond length, but in molecular crystals, the lattice also has a considerable input.

In both carbamazepine and primidone, there are multiple opportunities for hydrogen bonding to occur. Hydrogen bonds can form between the carbonyl group and the hydrogens of the primary amine in the case of carbamazepine and between the carbonyl groups and the secondary amines in the case of primidone. As a result of this feature, it may be seen that the observed crystal structures are not the structure that has the lowest lattice energy value, as has often been the case [3.41]. In addition, the maximum number of possible hydrogen bonds predicted to occur may not always form in practice. This

could be due to steric hindrance or due to the low energy value of the hydrogen bond itself [3.51].

3.3 Crystal Structures of Carbamazepine

Table 1 lists published crystal structure data and measurements calculated using Cerius[2], while figure 6 shows the crystal structures of the (a) monoclinic and (b) trigonal polymorphs available from the Cambridge Structural Database.

Parameter	Monoclinic	Trigonal
a[Å]	7.54	35.45
b[Å]	11.16	35.45
c[Å]	13.91	5.25
α[º]	90.00	90.00
β[º]	92.86	90.00
γ[º]	90.00	120.00
Volume[Å^3]	1168.3	5718.30
Space Group	P2$_1$/n [3.8]	R-3
Torsion Angle [º]	178.00	-176.10
H-Bond Length [Å]	2.04	2.16
Z	4	18

Table 3.1; summary of published unit cell dimensions, space group and Z numbers for monoclinic and trigonal forms of carbamazepine [3.8, 3.9]. Cerius[2] was used to measure torsion angles and H-bond lengths
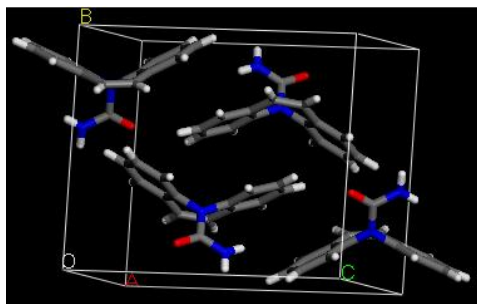
Figure 3.6 (a)
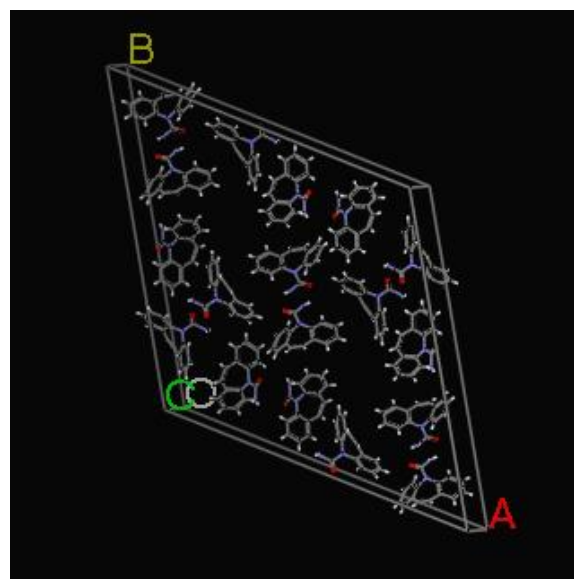
Figure 3.6 (b)

The crystal structure for both forms of carbamazepine used in this study were built in Cerius[2] using the fractional co-ordinates, bond lengths and bond angles previously published [3.8 and 3.9]. Hydrogen atoms were added in Cerius[2] once the main body of the molecule had been built.

Methods used in previous studies on both forms of carbamazepine include XRPD, DSC, NMR, FT-IR, TGA, SEM and thermo-microscopy [3.20 - 3.25]. The results show that the monoclinic polymorph is more stable at room temperature and is more frequently formed when carbamazepine is crystallized from solvents with high dielectric constants at slow cooling rates. The monoclinic polymorph is most dense at 1.35 Mgm^{-3}, the trigonal form at 1.23 Mgm^{-3}.

Amide groups are generally planar and this is seen in the monoclinic form of carbamazepine, in figure 7(a) as it aids the most beneficial way of packing the aromatic molecules within the unit cell (edge to face) [3.59].

The amide group can be distorted by rotation through the C-N bond or by pyramidalization of the amine group [3.58]. The latter is usually the more energetically feasible option although a combination of effects is normally observed.

The pyramidalization of the amine group is seen in trigonal carbamazepine, in figure 7(b), as this also aids the packing of the aromatic molecules within the unit cell (offset pi stacking).
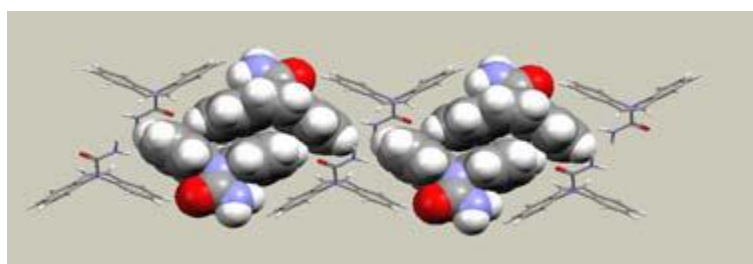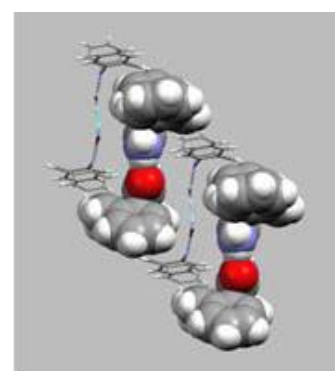
Figure 3.7(a)   **Figure 3.7(b)**

Both images referenced from [3.59].

As the morphologies of the two forms will show, the monoclinic polymorph crystallizes into well-defined prisms, and the trigonal form into needle like hollow hexagonal tubes.

Carbamazepine has good hydrogen bonding capabilities and although possessing a reasonably rigid backbone, it does comprise one degree of conformational flexibility. The introduction of a flexible side group would test the ability of the search methods to cope with more than just rigid structures. Figure 8 shows the degree of conformational flexibility in carbamazepine.
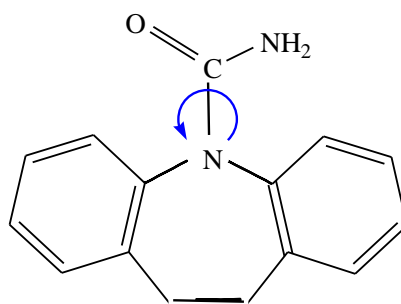
Figure 3.8; the flexible group within a molecule of carbamazepine.

3.3.1 Monoclinic Carbamazepine

The unit cell dimensions of monoclinic carbamazepine are already listed in table 1.

The carbamazepine molecule in the monoclinic structure has the following properties:

- The central ring adopts a boat conformation relative to the C(1), C(6), C(9), C(14) plane.
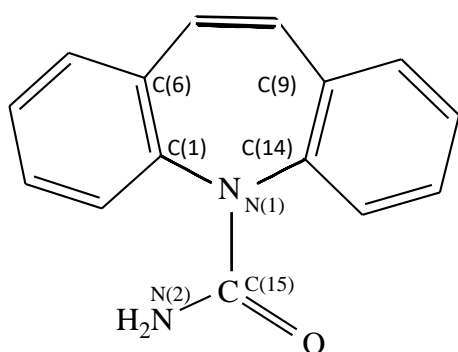
Figure 3.9; numbering the atoms within the plane in the central ring and in the torsion angle described in Figure 10.
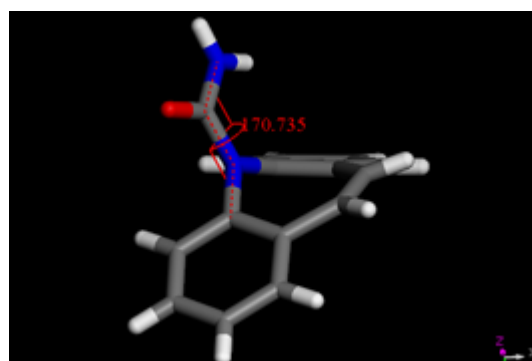
Figure 3.10; showing the torsion angle between the central ring and the amide group involving atoms C(14), N(1), C(15), N(2). Numbering taken from Figure 9.

- The two benzene moieties are planar and form a 126.6° dihedral angle.
- The torsion angle between the central ring and the amide group is 170.7°. Measured in Cerius^2 and referenced in [3.8].

94

- Centrosymmetric dimers are formed, which are held together by hydrogen bonds. This means the dimer, comprising two carbamazepine molecules linked together via hydrogen bonds, has a symmetrical arrangement radiating around a central point.
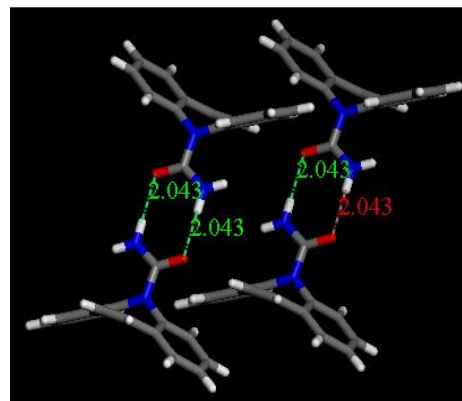
Figure 3.11; dimers of monoclinic carbamazepine with hydrogen bond length of 2.043Å
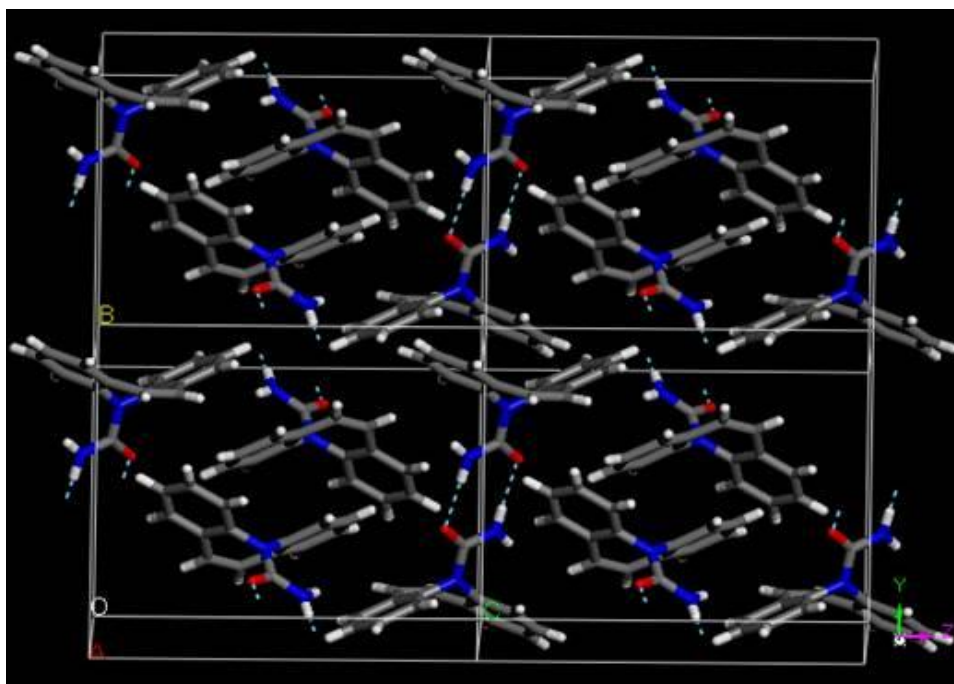
Figure 3.12; dimers of monoclinic carbamazepine built up to show the crystal packing motif.

Figure 13 shows (a) Kofler hot stage microscopy – a polarised light image at room temperature of prismatic crystals of monoclinic carbamazepine [3.25]; (b) the crystal habit, the description of the visible external shape of a crystal, for monoclinic carbamazepine. This was calculated using the *BFDH* method in the Morphology menu of Quantum Mechanics in Cerius[2], and (c) S.E.M. of monoclinic carbamazepine [3.9].

3.13(a)   **3.13(b)**   **3.13(c)**
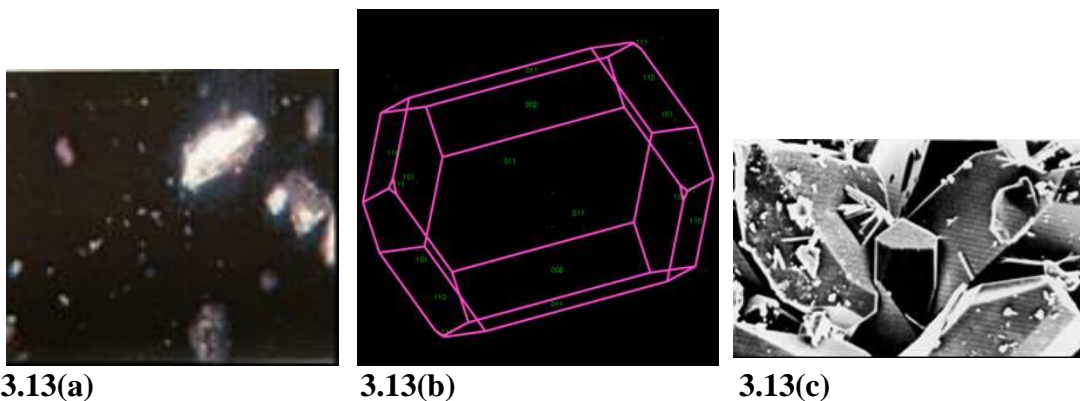
Figure 3.13 (a): Image of monoclinic carbamazepine; **Figure 3.13 (b):** Crystal habit of monoclinic carbamazepine predicted using Cerius[2]; **Figure 3.13 (c):** S.E.M. image of the prismatic crystal in monoclinic carbamazepine.

3.3.2 Trigonal Carbamazepine

The unit cell dimensions of trigonal carbamazepine are already listed in table 1. The carbamazepine molecule in the trigonal structure has the following properties;

- The central ring adopts the boat conformation relative to the C(1), C(6), C(9), C(14) plane.
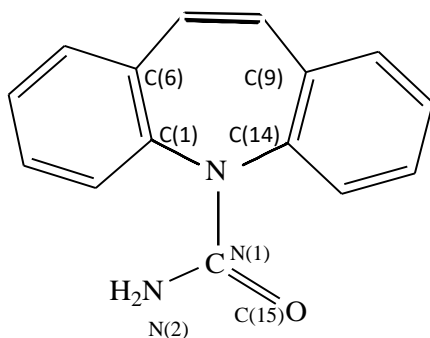
Figure 3.14; numbering the atoms within the plane in the central ring and in the torsion angle described in Figure 15.
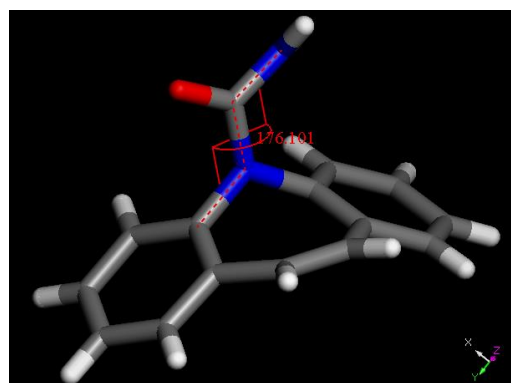
Figure 3.15; showing the torsion angle between the central ring and the amide group involving atoms C(14), N(1), C(15), N(2). Numbering taken from Figure 14.

- The two benzene moieties are planar and form a dihedral angle of 124.7°.
- The torsion angle made by the amide group with the central ring is 176.1°.
- Centrosymmetric dimers are formed, which are held together by hydrogen bonds in a similar way to the monoclinic form. The hydrogen bonds are longer in the

96

trigonal form of carbamazepine and this creates voids in molecular packing (figure 17), resulting in a lower density value.
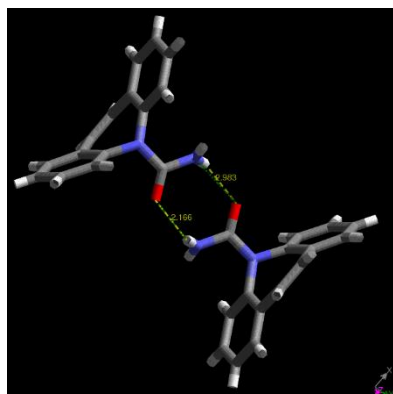
Figure 3.16; showing the hydrogen bonded dimers in trigonal carbamazepine

- The two nitrogen-hydrogen bonds on the amine group are different lengths. The length of the N-H bond involved in hydrogen bonding is 0.75Å. The remaining free N-H bond measures 0.96Å [3.9].

- The distance between $N(2)\cdots O(1)$ was measured in Cerius^2 as 2.98Å. In literature this distance is reported as 2.89Å. The distance measured in Cerius^2 between $N(2)$-$H(29)\cdots O(1)$ was 2.16Å, whereas in literature it is reported as 1.97Å. These difference will most likely arise from the different force fields being used which are paramaterised differently. In the literature [3.61] the Momanay-Carruthers force field was utilized whereas in this study, the Dreiding force field was used.
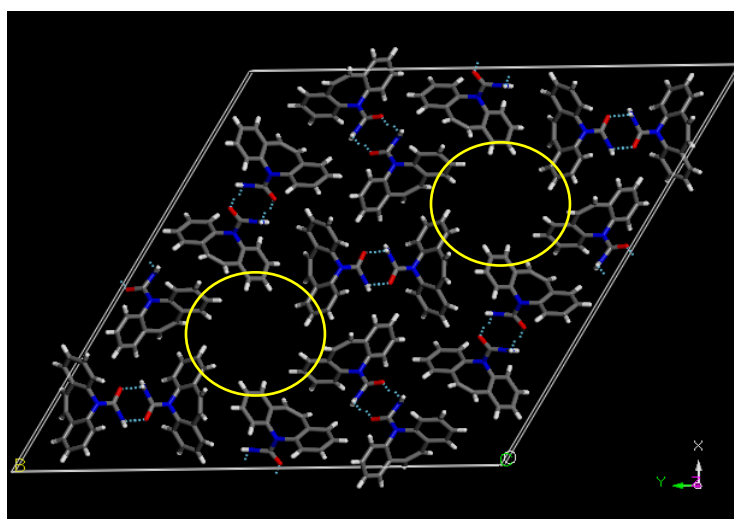
Figure 3.17; showing the voids within the molecular packing in trigonal carbamazepine resulting in the lower density.

Figure 18 shows (a) Kofler hot stage microscopy – a polarised light image at room temperature of needle-like crystals of trigonal carbamazepine [3.25]; (b) the crystal habit, the description of the visible external shape of a crystal, for trigonal carbamazepine. This was calculated using the *BFDH* method in the Morphology menu of Quantum Mechanics in Cerius[2], and (c) S.E.M. of trigonal carbamazepine [3.9].
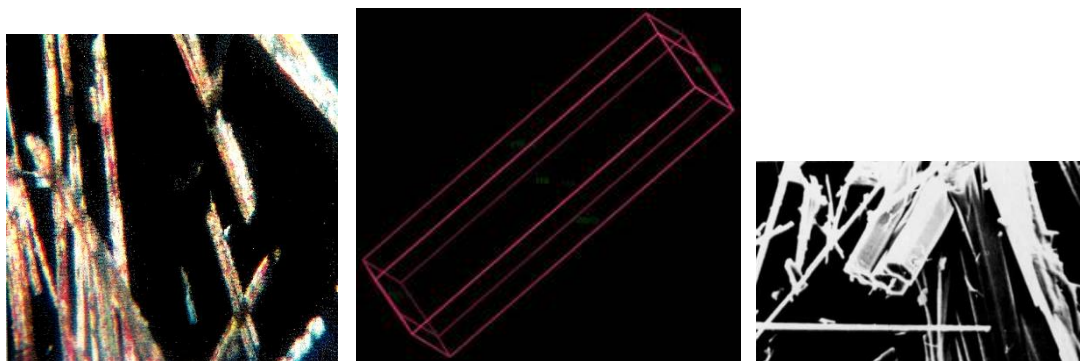
3.18(a)   **3.18(b)**   **3.18(c)**

Figure 3.18 (a); image of trigonal carbamazepine; **Figure 3.18 (b);** crystal habit of trigonal carbamazepine predicted using Cerius[2]; **Figure 3.18 (c);** S.E.M. image of the needle-like crystals in trigonal carbamazepine.

3.4 Crystal Structures of Primidone

Table 2 lists published unit cell dimensions, space groups and Z numbers [3.31, 3.34] alongside measurements of torsion angles and H-bond lengths calculated using Cerius^2. Figure 19 shows the crystal structures of polymorph (a) A (monoclinic) (b) B (orthorhombic) available from the Cambridge Structural Database.

Parameter	Structure	
	Primidone A	Primidone B
	MOPAC	MOPAC
Unit Cell Vol [Å^3]	1284.96	2237.42
a[Å]	12.25	10.27
b[Å]	7.09	7.92
c[Å]	14.81	27.54
α[º]	90.00	90.00
β[º]	117.82	90.00
γ[º]	90.00	90.00
H-Bond Length [Å]	2.04, 1.94	1.96, 2.00
Torsion Angle [º] (ethyl, phenyl)	-61.5, -40.1	-59.5, 10.6
Z	4	8
Space Group	P 2$_1$/c	P b c a

Table 3.2; published and calculated crystal structure data for Primidone A and B.
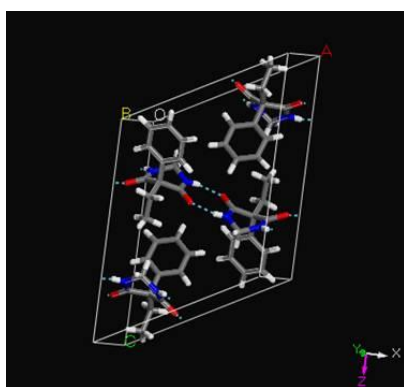
Figure 3.19 (a); crystal structure of Primidone, Monoclinic Form A
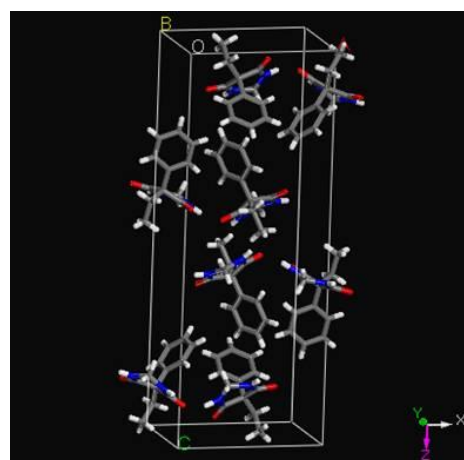
Figure 3.19 (b); crystal structure of Primidone, Orthorhombic Form B

Previous studies on both forms of primidone include XRPD, IR, NMR, GCMS and TLC [3.32]. Research shows that the monoclinic polymorph is formed by slow evaporation of a solution of commercial primidone in ethanol. The monoclinic polymorph has a measured density of 1.26gcm^{-3} the orthorhombic form of 1.29gcm^{-3}.

Primdione has good hydrogen bonding capabilities and comprises two degrees of conformational flexibility. The introduction of a second flexible group would test the ability of the search methods further. Figure 20 shows the two degrees of conformational flexibility in primidone.
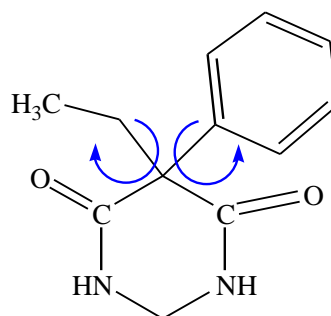
Figure 3.20; flexibility defined in the molecule for the unit cell searches of both primidone A and B.

In this study, it was found that the starting molecular configurations of primidone A and B differed quite significantly and this is described in more detail later.

3.4.1 Primidone A (Monoclinic)

The following structural details were noted for primidone A:

- The pyrimidine ring adopts the boat conformation where C(5) and C(2) are elevated from the plane N(1), N(3), C(4), C(6).
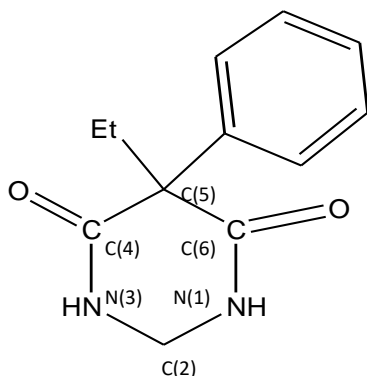
Figure 3.21: Numbering the atoms within the plane in the central ring
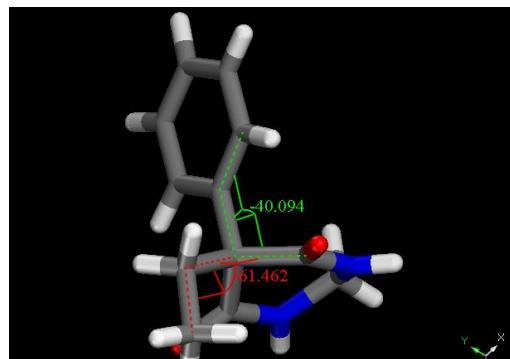
Figure 3.22: Showing the torsion angles between the pyrimidine ring and the phenyl and ethyl groups.

- Bond lengths and angles on either side of the median line C(2)-C(5), whilst not always identical, are very similar and only slight differences in bond lengths and angles are seen. This gives the appearance of a symmetrical structure. Atomic numbering is shown in figure 23 which corresponds to the tabulated bond lengths and angles in table 3. The bond lengths and angles are taken from [3.31].
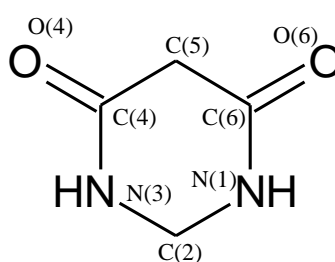
Figure 3.23; Atomic numbering of the pyrimidine ring.

Atoms	Bond Lengths (Å)	Atoms	Bond Angles (°)
C(2)-N(3)	1.441	C(2)-N(3)-C(4)	122.1
C(2)-N(1)	1.441	C(2)-N(1)-C(6)	121.8
N(3)-C(4)	1.329	N(3)-C(4)-O(4)	122.2
N(1)-C(6)	1.329	N(1)-C(6)-O(6)	122.9
C(4)-C(5)	1.540	N(3)-C(4)-C(5)	115.8
C(6)-C(5)	1.543	N(1)-C(6)-C(5)	116.0
C(4)-O(4)	1.223	O(4)-C(4)-C(5)	122.0
C(6)-O(6)	1.230	O(6)-C(6)-C(5)	121.0

Table 3.3; Bond lengths and angles of the pyrimidine ring from [3.31].

- The phenyl ring is planar and the torsion angle between the pyrimidine ring and the phenyl ring is –40.1° shown in figure 3.22.

- The torsion angle between the pyrimidine ring and the ethyl group is 61.5° also shown in figure 3.22.

- The molecules within the unit cell are connected by two types of hydrogen bonds although both types are very similar in length. Type one (slightly longer) creates a dimer, type two (slightly shorter) links the dimers into sheets of molecules shown in figure 24.
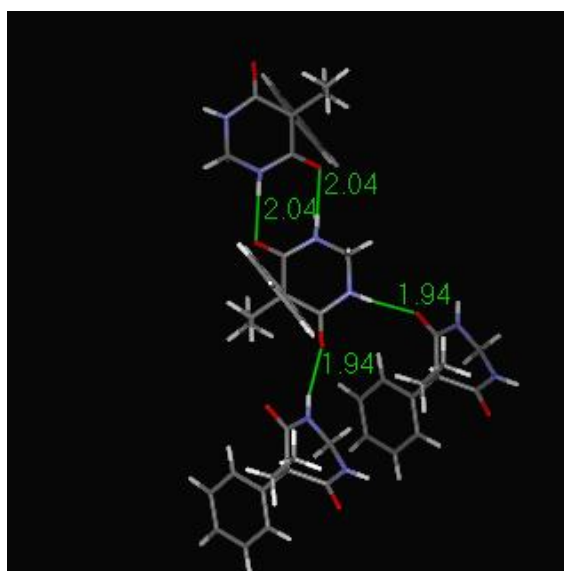
Figure 3.24; hydrogen bonding motif in the monoclinic form of primidone.
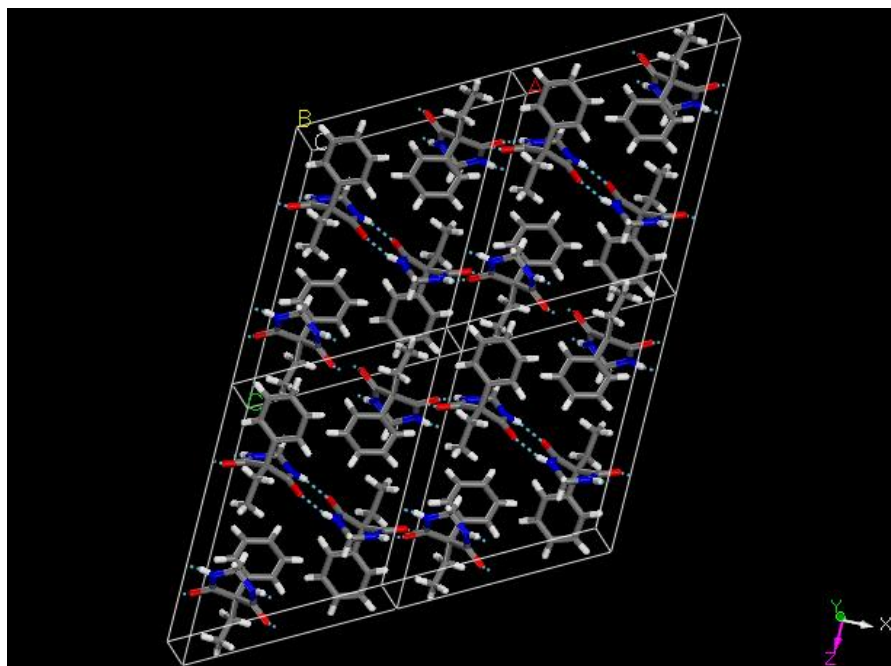
Figure 3.24; molecular packing in the monoclinic form of primidone.

The crystal habit shown in figure 25 was calculated using the *BFDH* method in the quantum mechanics package of Cerius[2].
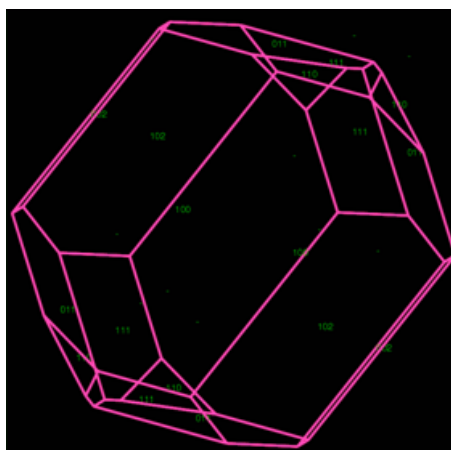
Figure 3.25; crystal habit of monoclinic primidone.

3.4.2 Primidone B (Orthorhombic)

The following structural details were noted for primidone B:

- The pyrimidine ring adopts the boat conformation where C(5) and C(2) are elevated from the plane N(1), N(3), C(4), C(6).
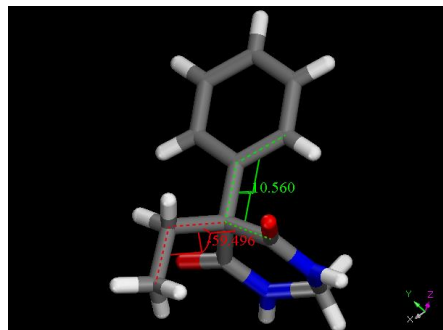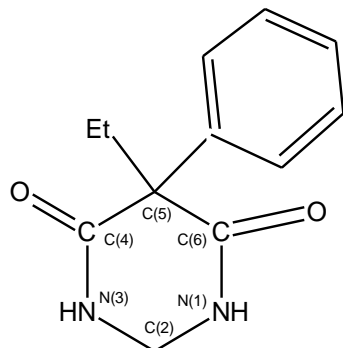
Figure 3.26: Numbering the atoms within the plane in the central ring

Figure 3.27: Showing the torsion angles between the pyrimidine ring and the phenyl and ethyl groups.

- Bond lengths and angles on either side of the median line C(2)-C(5) in the orthorhombic form are very similar to those measured in the monoclinic form which again gives the pyrimidine ring the appearance of a symmetrical structure.

- The phenyl ring is planar and the torsion angle between the pyrimidine ring and the phenyl ring is 10.6º shown in figure 3.27.

- The torsion angle between the pyrimidine ring and the ethyl group is 59.5° also shown in figure 3.27.

- Each molecule in the unit cell is attached to four others via hydrogen bonds which are very similar in length, shown in figure 3.28.
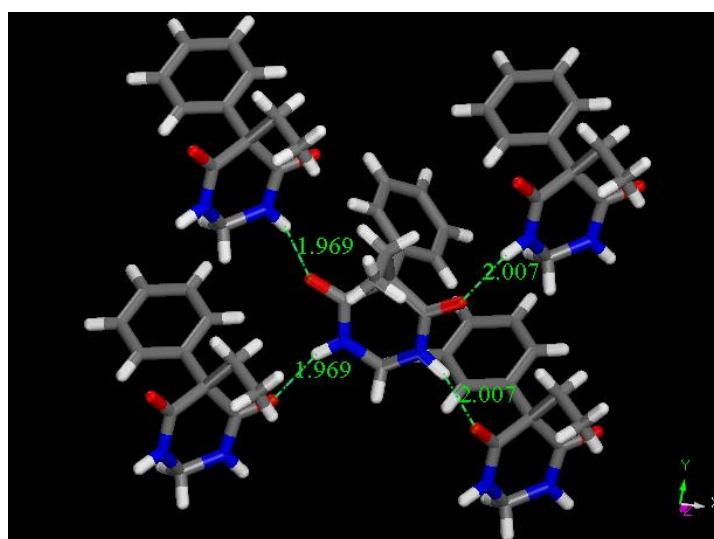
Figure 3.28: Hydrogen bonding in the orthorhombic form of primidone.
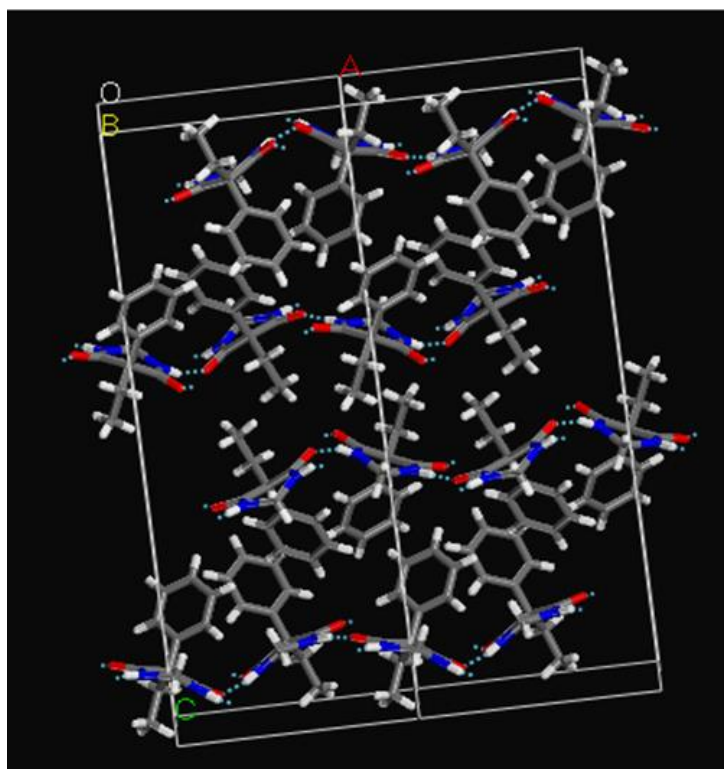
Figure 3.29: Molecular packing in the orthorhombic form of primidone.

- The hydrogen bond network created sheets of molecules in the orthorhombic polymorph. This is shown in figure 3.29.
- The crystal habit, figure 3.30, was again calculated using the BFDH method in Cerius[2].
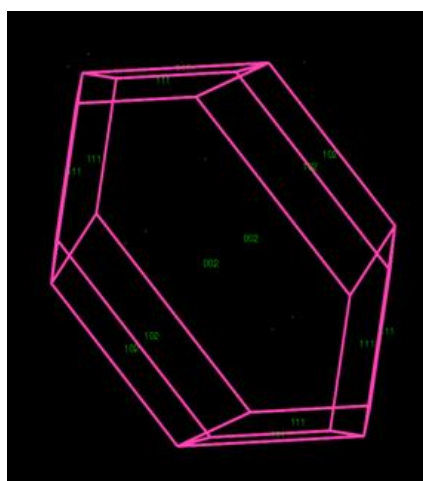
Figure 3.30; crystal habit of orthorhombic primidone.

3.5 Validating the Search methods

3.5.1 Systematic Searches

The systematic search uses a pre-defined step size to search the entire unit cell. The search is carried out using translational and rotational steps. For unit cells of a small volume and for rigid body systems, this appears a thorough and useful tool. Problems arise in optimizing the computational time against the step sizes when the unit cell volume increases significantly or when the molecule under investigation contains flexible groups. A further limitation of the systematic method is that it can only search a unit cell using one molecule in the asymmetric unit.

For the purposes of this study, the search and acceptance criteria for all the searches are shown in table 3. In the searches using primidone B the fraction translated along the x axis was reduced. This was to compensate for the difference in length of this third axis and ensure a reasonably even step size for all three directions.

	Step Sizes			Cut-off Parameters	
	Translations (fractional co-ordinates) x, y, z	Rotations [º] a, b, c	Torsions [º] 1*, 2**	Energy [kcal/mol]	R_{wp}
Primidone A	0.05 for all	10 for all	10.00, 10.00	-10.00	0.95
Primidone B	0.05 for y+z, 0.03 for x	10 for all	10.00, 10.00	-10.00	0.95
Monoclinic Carbamazepine	0.05 for all	10 for all	10.00	-10.00	0.95
Trigonal Carbamazepine	0.05 for all	10 for all	10.00	-20.00	0.95

Table 3.4; acceptance criteria used in all the searches. For the rigid body searches, torsions = 0.

*for primidone, phenyl torsion, **for primidone ethyl torsion

*for carbamazepine, amide group, **for carbamazepine, not applicable

3.5.2 Simulated Annealing Searches

Due to the random nature of the simulated annealing method, each search was repeated a number of times to ensure confidence that a sufficient area of the unit cell had been covered. Also, the best result from the previous search was used as the starting point for the following search. If consecutive searches produced results that were very similar to previous searches, it could be assumed that the method did not need to be repeated and was capable of quickly finding a useful energy minimum. A second view is that the search had become trapped in a local minima and so a new, random starting point was created for the following search.

	Maximum Step Sizes Allowed			Cut-off Parameters	
	Translations (fractional co-ordinates) x, y, z	Rotations [°] a, b, c	Torsions [°] 1*, 2**	Energy [kcal/mol]	R_{wp}
Primidone A	0.5 for all	10 for all	10, 10	-10.00	0.90
Primidone B	0.5 for all	10 for all	10, 10	-10.00	0.90
Monoclinic Carbamazepine	0.5 for all	10 for all	10	-10.00	0.75
Trigonal Carbamazepine	0.5 for all	10 for all	10	-10.00	0.75

Table 3.5: acceptance criteria used in all the searches. For rigid body searches, torsions = 0.

*for primidone, phenyl torsion, **for primidone ethyl torsion

*for carbamazepine, amide group, **for carbamazepine, not applicable

3.6 Post Analysis Optimization and Refinement Methods

Once a trial structure had been generated, the symmetry was reduced to P1. The molecule was treated as a rigid body while it was moved to seek out 'better' positions in the near surrounding area. Sometimes this move was quite large and good results were generated from unlikely starting points. When the optimized structure was calculated, the immediate area around this energy minimum was searched, giving a small indication as to the energy surface at that point. The gradients calculated were plotted against the distance actually moved by the unit.

Finally, a refinement of the calculated data was briefly attempted on the simulated powder data to further optimize the position of the molecules in the unit cell.

3.7 Results

3.7.1 Carbamazepine

Lattice energies for the published monoclinic and trigonal forms of carbamazepine are listed in table 5. This lattice energy was calculated in Cerius^2 using the molecular geometry taken from the Cambridge Structural Database, with hydrogen atoms being added to the molecule in Cerius^2, and a MOPAC charge set applied also using Cerius^2.

In summary, the aim of these searches was to assess how successfully systematic and simulated annealing searches would be in finding the correct structure with the corresponding lattice energy.

Carbamazepine Polymorph	Lattice Energy [kcal/mol]
Monoclinic	-130.92
Trigonal	-553.55

Table 3.6; Lattice energy in kcal/mol for two polymorphs of carbamazepine.

Systematic Searching and Simulated Annealing with Rigid Bodies

The results generated from these searches were very promising for future investigations. Even though it was quite a straightforward example where the starting configuration was known, it was encouraging that the Dreiding force field provided the process with information accurate enough to determine the global minimum. The results are shown in table 6 for both lattice energies and then x-ray powder diffraction pattern fits.

Carbamazepine Polymorph	Lattice Energy [kcal/mol]		R_{wp}
	Un-optimized	Optimized	
Systematic Searches			
Monoclinic	-78.26	-130.94	0.91
Trigonal	-125.55	-567.53	0.98
Simulated Annealing			
Monoclinic	-119.38	-130.93	0.74
Trigonal	-431.49	-553.18	0.98

Table 3.7; results for the rigid body searches of carbamazepine

Monoclinic carbamazepine: The following two figures show an overlay of the crystal structure and the top optimized structures found by the two different search methods.
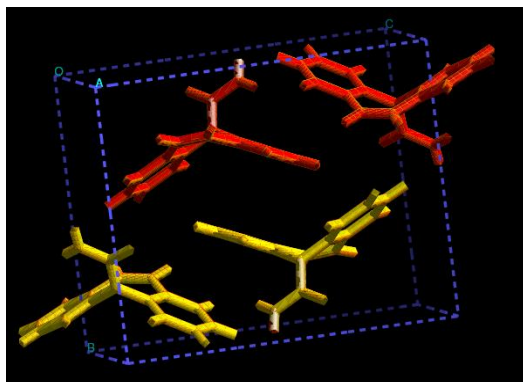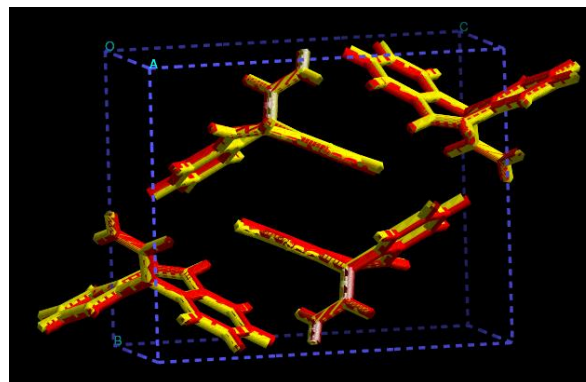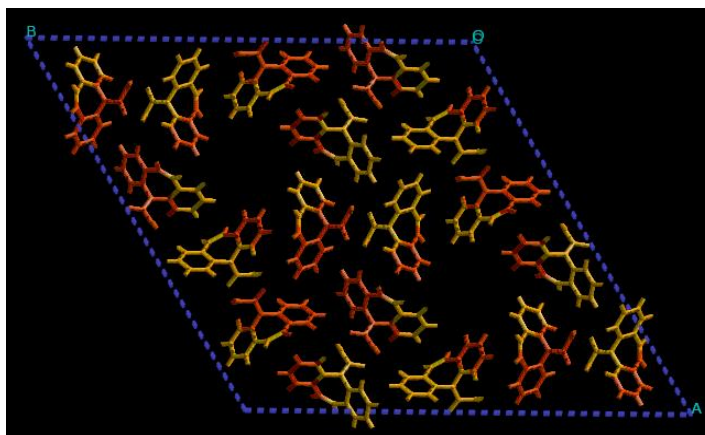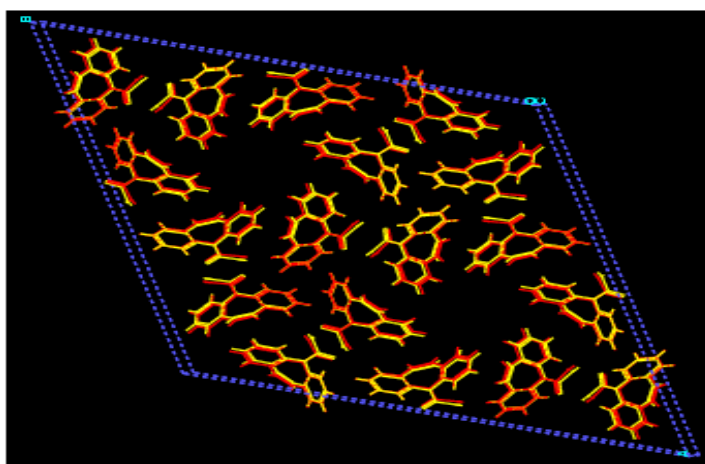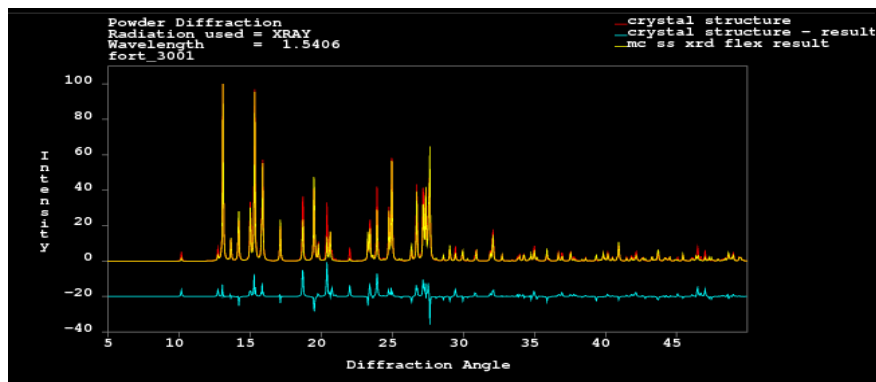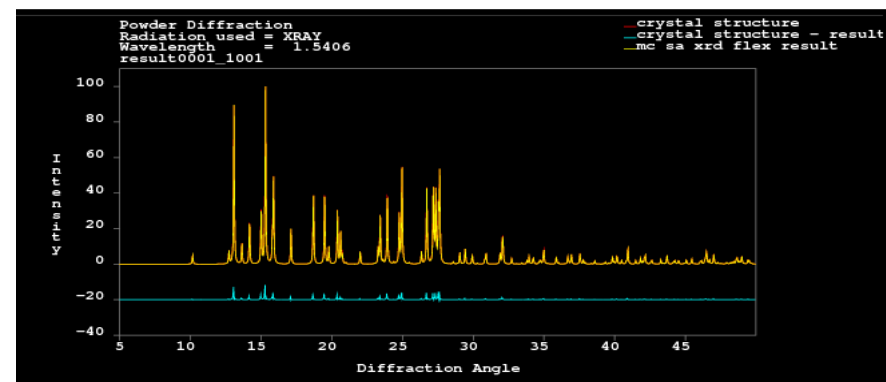

3.31(a)　　　　　　　　　　　　**3.31(b)**

Figure 3.31; rigid body searches of monoclinic carbamazepine. The crystal structure (red) compared with the top optimized lattice energy result (yellow) from (a) the systematic search (b) simulated annealing search.

Trigonal carbamazepine: The following two figures show an overlay of the crystal structure and the top optimized structures found by the two different search methods.
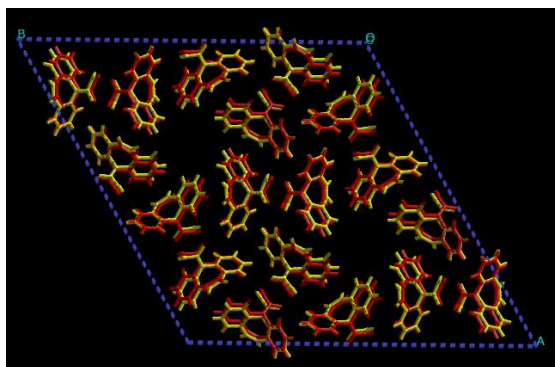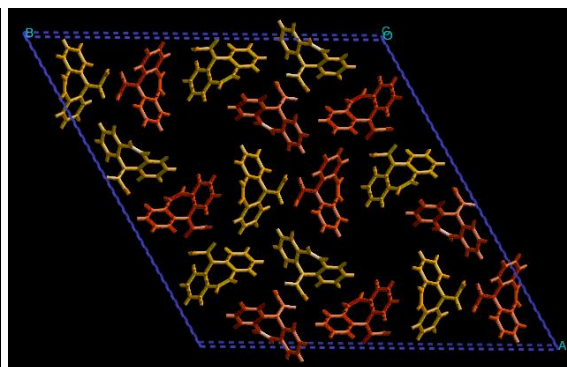
3.32 (a)　　　　　　　　　　　　**3.32 (b)**

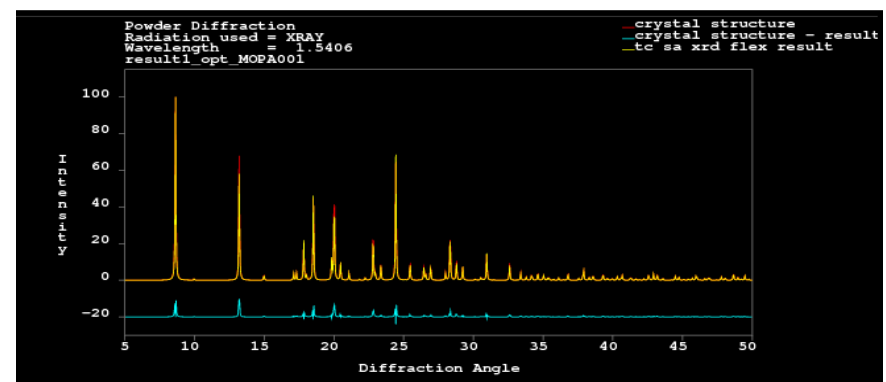**Figure 3.32;** rigid body searches of trigonal carbamazepine. The crystal structure (red) compared with the top optimized lattice energy result (yellow) from (a) the systematic search (b) simulated annealing search.
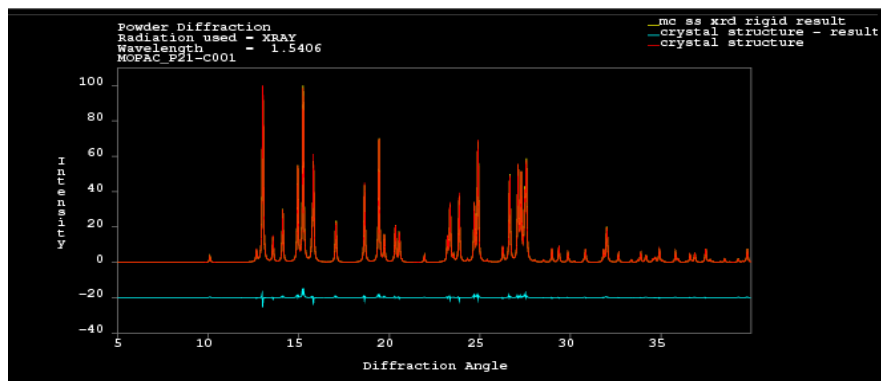
3.33 (a)
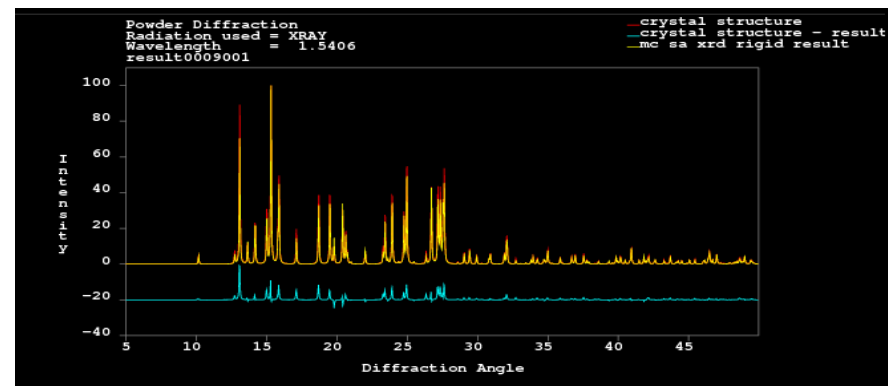
3.33 (b)

Figure 3.33: rigid body searches of monoclinic carbamazepine. The crystal structure (red) compared with the top xrd result (yellow) from (A) the systematic search (B) simulated annealing. The difference plots for both sets are in blue.
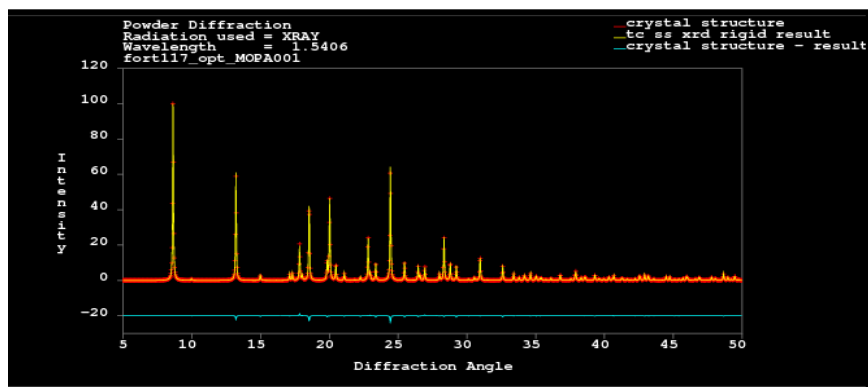
3.34 (a)
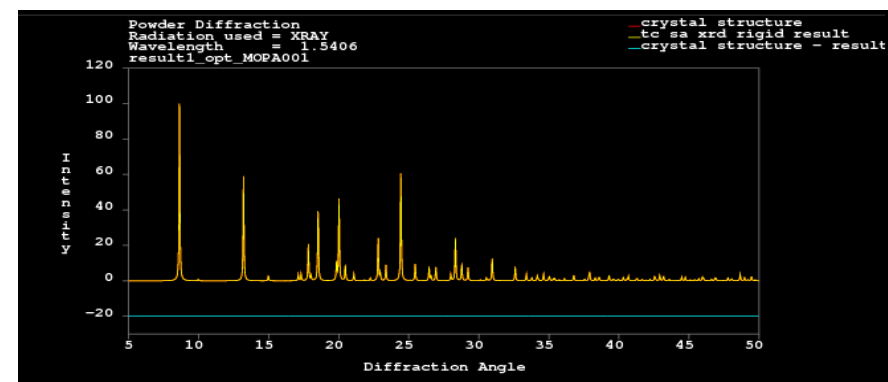
3.34 (b)

Figure 3.34; rigid body searches of trigonal carbamazepine. The crystal structure (red) compared with the top xrd result (yellow) from (A) the systematic search (B) simulated annealing. The difference plots for both sets are in blue.

Systematic Searching and Simulated Annealing with Conformational Flexibility

The degree of freedom was introduced in these searches and again, reasonable results were obtained prior to using the Rietveld refinement program. The results are shown in table 3.7.

	Lattice Energy [kcal/mol]		R_{wp}
Polymorph	Un-optimized	Optimized	
Systematic Searching			
Monoclinic Carbamazepine	-128.59	-130.92	0.95
Trigonal Carbamazepine	-476.78	-552.09	0.98
Simulated Annealing			
Monoclinic Carbamazepine	-122.82	-130.66	0.73
Trigonal Carbamazepine	-547.18	-557.75	0.98

Table 3.8; results for the conformationally flexible searches of carbamazepine

Monoclinic carbamazepine: The following two figures show an overlay of the crystal structure and the top optimized structures found by the two different search methods.
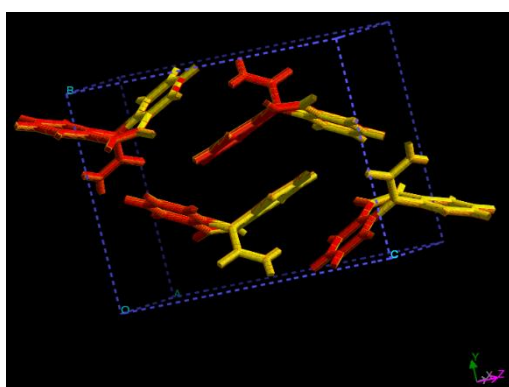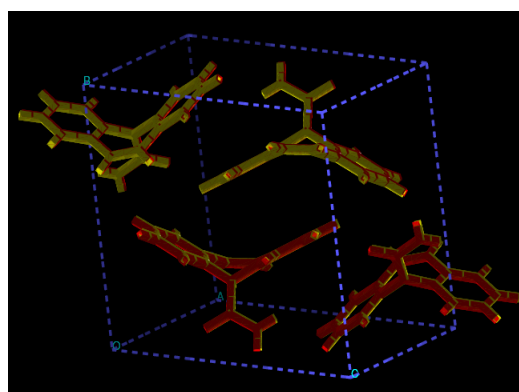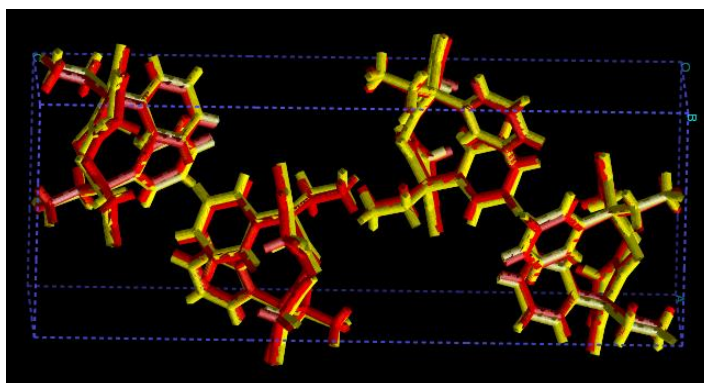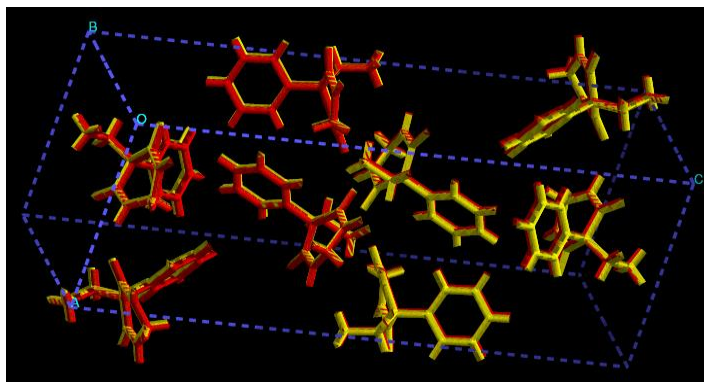

3.35 (a)   **3.35 (b)**

Figure 3.35; conformationally flexible searches of monoclinic carbamazepine. The crystal structure (red) compared with the top lattice energy result (yellow) from (a) the systematic search (b) simulated annealing search.

Trigonal carbamazepine: The following two figures show an overlay of the crystal structure and the top optimized structures found by the two different search methods.
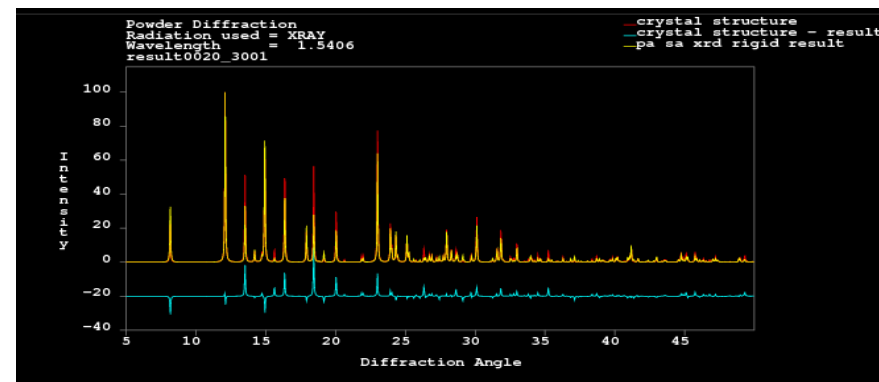

3.36 (a); the systematic search

3.36 (b); simulated annealing

Figure 3.36; conformationally flexible searches of trigonal carbamazepine. The crystal structure (red) compared with the top lattice energy result (yellow) from;

3.37 (a)   **3.37 (b)**

Figure 3.37: Conformationally flexible searches of monoclinic carbamazepine. The crystal structure (red) compared with the top xrd result (yellow) from (a) the systematic search (b) simulated annealing, the difference plots for both sets are in blue.
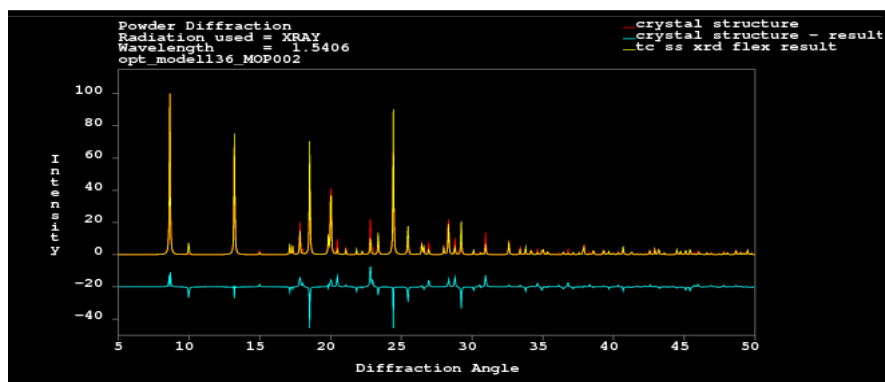
3.38 (a)   **3.38 (b)**

Figure 3.38; conformationally flexible searches of trigonal carbamazepine. The crystal structure (red) compared with the top xrd result (yellow) from (a) the systematic search (b) simulated annealing, the difference plots for both sets are in blue
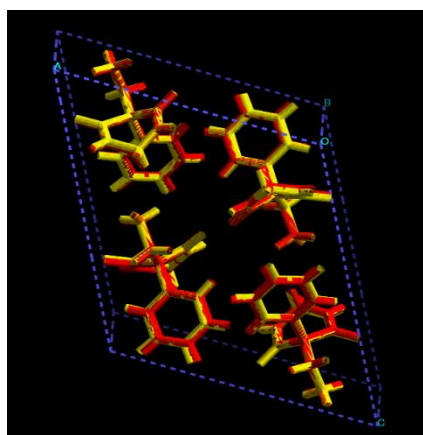
3.7.2 Primidone

Systematic Searching and Simulated Annealing with Rigid Bodies

Carbamazepine Polymorph	Lattice Energy [kcal/mol]
Monoclinic (A)	-130.06
Orthorhombic (B)	-260.80

Table 3.9; published lattice energy in kcal/mol for two polymorphs of primidone.

	Lattice Energy [kcal/mol]		R_{wp}
Polymorph	**Un-optimized**	**Optimized**	
Systematic Searching			
Primidone A	-109.99	-128.64	0.84
Primidone B	-204.64	-234.12	0.94
Simulated Annealing			
Primidone A	-115.36	-130.06	0.69
Primidone B	-257.60	-260.80	0.54

Table 3.10; showing the results for the rigid body searches of primidone.

Primidone A: The following two figures show an overlay of the crystal structure and the top optimized structures found by the two different search methods.
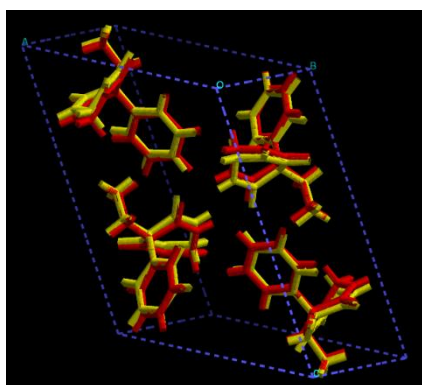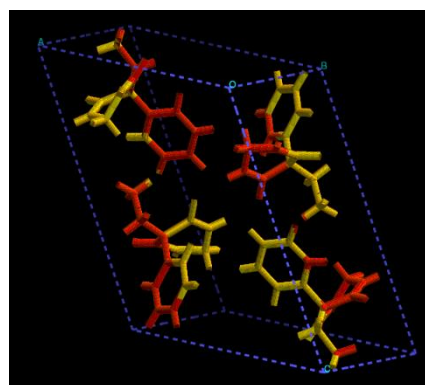
3.39 (a)   **3.39 (b)**

Figure 3.39; rigid body searches of primidone A. The crystal structure (red) compared with the top lattice energy result (yellow) from (a) the systematic search (b) simulated annealing.

Primidone B: The following two figures show an overlay of the crystal structure and the top optimized structures found by the two different search methods.
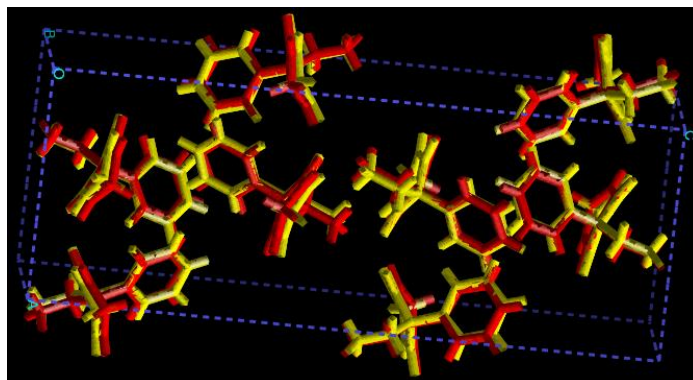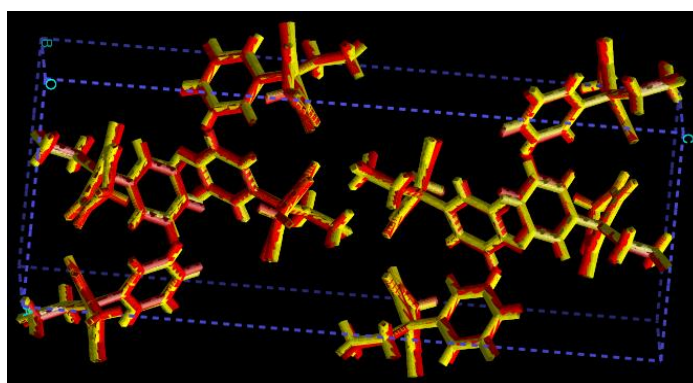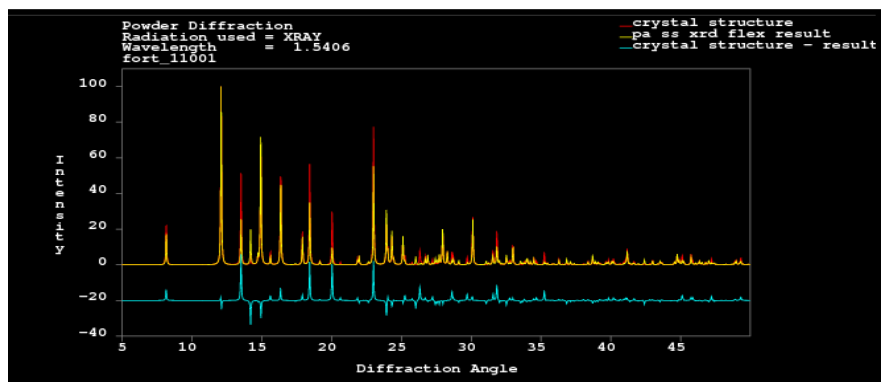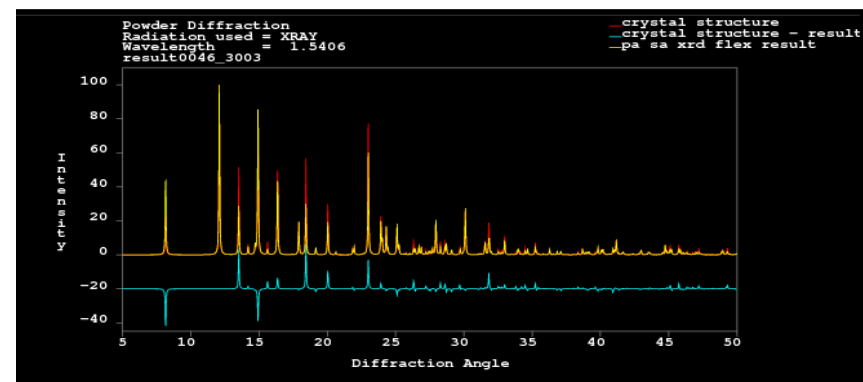

3.40 (a) the systematic search

3.40 (b) simulated annealing

Figure 3.40; rigid body searches of primidone b. The crystal structure (red) compared with the top lattice energy result (yellow).
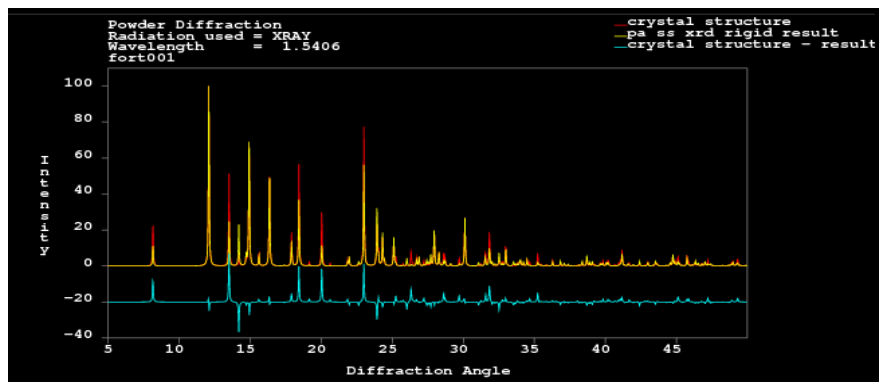
3.41 (a)

3.41 (b)

Figure 3.41; rigid body searches of primidone a. The crystal structure (red) compared with the top xrd result (yellow) from (a) the systematic search (b) simulated annealing, the difference plots for both sets are in blue.
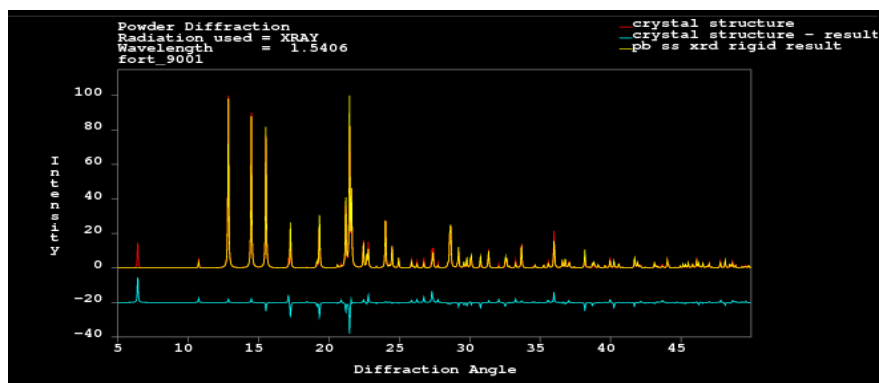
3.42 (a)
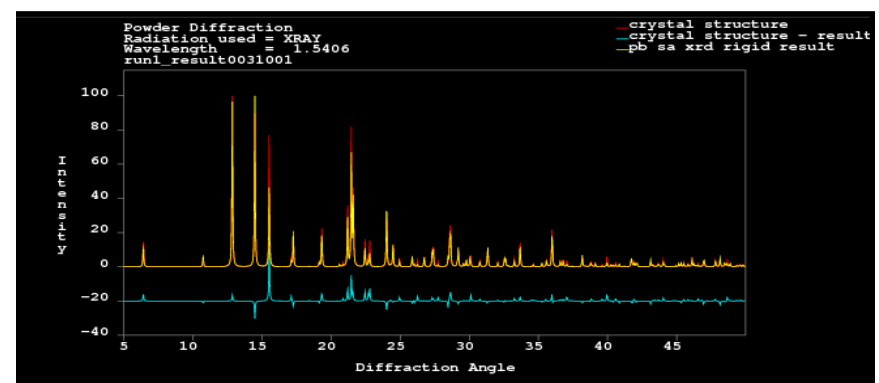
3.42 (b)

Figure 3.42; rigid body searches of primidone b. The crystal structure (red) compared with the top xrd result (yellow) from (a) the systematic search (b) simulated annealing, the difference plots for both sets are in blue.

Systematic Searching and Simulated Annealing with Conformational Flexibility

The two degrees of freedom were introduced in these searches and again, even with this added degree of difficulty, reasonable results were obtained. The results are tabulated in table 10 and figures 43 - 46 show the overlay of the crystal structure with the top optimized structures generated.

Polymorph	Lattice Energy [kcal/mol]		R_{wp}
	Un-optimized	Optimized	
Systematic Searching			
Primidone A	-125.41	-127.94	0.64
Primidone B	-199.33	-268.90	0.92
Simulated Annealing			
Primidone A	-114.33	-128.64	0.48
Primidone B	-242.46	-262.54	0.43

Table 3.11; the results for the primidone B conformationally flexible searches.

Primidone A: The following two figures show an overlay of the crystal structure and the top optimized structures found by the two different search methods.
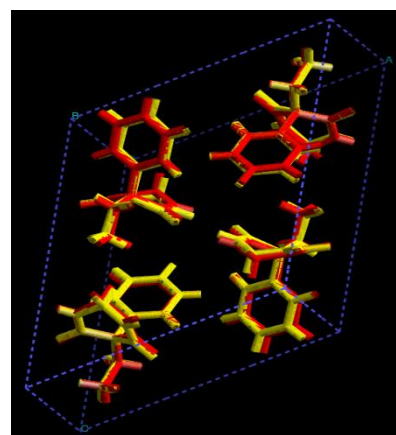
3.43 (a)   **3.43 (b)**

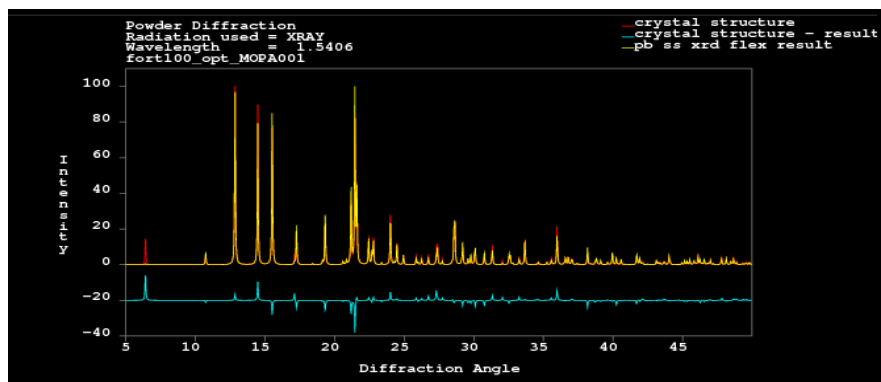**Figure 3.43:** conformationally flexible searches of primidone A. The crystal structure (red) compared with the top lattice energy result (yellow) from (a) the systematic search (b) simulated annealing.

Primidone B: The following two figures show an overlay of the crystal structure and the top optimized structures found by the two different search methods.

3.44 (a) the systematic search

3.44 (b) simulated annealing

Figure 3.44; conformationally flexible searches of primidone b. The crystal structure (red) compared with the top lattice energy result (yellow).

3.45 (a)

3.45 (b)

Figure 3.45; conformationally flexible searches of primidone A. The crystal structure (red) compared with the top xrd result (yellow) from (a) the systematic search (b) simulated annealing, the difference plots for both sets are in blue

3.46 (a)
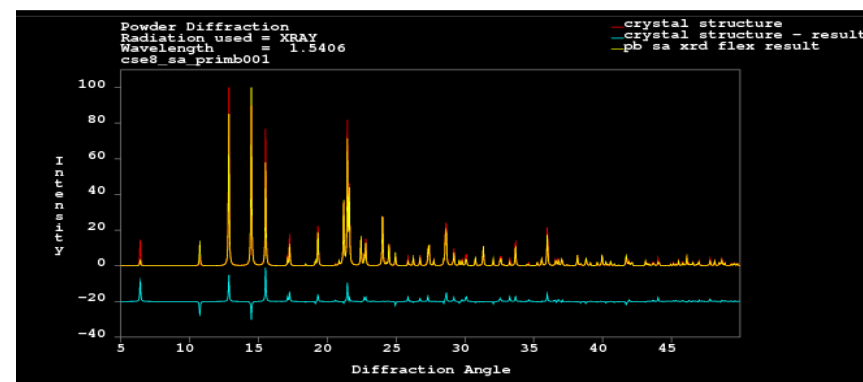
3.46 (b)

Figure 3.46; conformationally flexible searches of primidone B. The crystal structure (red) compared with the top xrd result (yellow) from (a) the systematic search (b) simulated annealing, the difference plots for both sets are in blue.

3.8   Discussion of Results

The work included in this chapter aimed to validate the new simulated annealing code and test it against the systematic searching method. The systematic search had already produced excellent results for structures with π-π stacking interactions and hydrogen bonding features such as indigo, 6,13-dichlorodiphendioxazine, phenanthrene, paracetamol and benzophenone [3.61].

At the time of this work, the key steps of search methods for crystal structure determination [3.1] were;

- Generation of the molecular structure and unit cell parameters.
- Selection of a theoretical set to describe the intramolecular energy and intermolecular forces.
- Global search of an experimentally calculated unit cell.
- Optimization of generated structures.

For the purposes of validating the simulated annealing search method and comparing results between this and the grid based search, the first step was carried out using as much published information as possible. The starting molecular conformations were downloaded from the CSD. Cerius^2 was used to 'add-in' or build the hydrogen atoms necessary and to implement charges on each atom. The unit cell was built in Cerius^2 from the published data and for both systematic and random search methods the molecule was placed in a random orientation at the origin of the unit cell.

In the rigid body searches, the aim was to show whether the step sizes for translations and orientations were set to cover a suitable amount of space without taking too much time. If smaller step sizes were required, how would this affect the time taken to reach the global minimum? Initially, molecular rotation steps were set at 5 degrees but this was soon changed to 10 degrees when the same results were found in a shorter amount of search time.

For both search methods, in the case of flexible searches, the molecule was again placed at the origin of the unit cell in a random orientation. In this instance however, any flexible groups identified were also placed in a random starting orientation. The flexible searches carried out following the rigid body searches used the assumption that the translational and molecular rotational step sizes were optimized for this example. In the flexible

searches, the aim was to show whether the step sizes for torsional rotations were set to cover a suitable amount of geometries without taking too much additional time. The torsional step sizes were set to 5 degrees initially which was increased to 10 degrees for the same reasons as the molecular rotational step sizes.

Understandably, if some knowledge of the molecular configuration of the sample is already available then these searches are very useful tools. When no such information is available however, the rigid searches would require multiple iterations of the global search using different, reasonable starting conformations, such as in the examples investigated within the first blind test set by the CSD [3.63]. Increasing the number of rigid body grid based searches required to cover all reasonable starting structures within the unit cell would seem to be too time-consuming a problem for anything other than very small molecules. Even in this study, with the benefit of prior knowledge of molecular configuration, the systematic method required days of processor time to carry out fine grid searches especially with unit cells of the proportions seen in the trigonal carbamazepine example.

The force field Dreiding [3.62] was used throughout this study as an example of a good quality generic force field. It has already been used in many other such cases of predicting the structure of hydrogen bonding systems and has been shown both in those cases and in this study to be able to model intra and inter molecular interactions to a sufficiently high standard for the optimization steps.

Lattice Energy calculations were the most popular method of ranking generated structures from crystal structure searches [3.63]. The calculations included terms for bonded and non-bonded interactions and effectively calculated the energy of a motionless system. The lowest energy results would be considered as the most thermodynamically stable structure and therefore the most likely structure to be observed or the most stable polymorph at all temperatures. The treatment of vibrational effects was ignored. However, more modern methods have started to take these effects into account and this has shown to have an impact on the ranking of generated structures in the later CSD blind tests [3.64, 3.65].

This validation investigation also considered x-ray powder diffraction pattern comparisons as a ranking method and so highlighted the need for good quality powder data. Samples of carbamazepine were prepared in the lab and analysed using capillary x-ray powder diffraction. Simulated data was used for the primidone examples. Although peak positions were generally good for carbamazepine, the intensities of the peaks did seem to differ significantly and this was reflected in the values for the powder pattern fits.

Optimization of generated structures in Cerius^2 was carried out in two steps [3.66]. The first being an energy evaluation step and the second being a conformational adjustment step. The sensitivity of the energy evaluation has been mentioned above and the same issues hold for this evaluation. This stage of the structure determination process relies on the fact that the structures generated from the searches will be of high quality for refinement. As tables 11 and 12 show, the differences between the observed molecular structures of primidone and carbamazepine with calculated geometry optimized structures generated in Cerius^2 in the quantum mechanics module can be significant.

	Carbamazepine Monoclinic		Carbamazepine Trigonal	
	Geometry Optimized	Crystal Structure	Geometry Optimized	Crystal Structure
Torsion Angle of the Amide Group [º]	-3.40	-9.60	-3.40	-3.90
Angle made between central ring and Amide Group [º]	42.80	60.10	70.66	42.80
Bond Length of N-CONH$_2$ [Å]	1.42	1.38	1.42	1.38

Table 3.12; molecular geometry measurements for carbamazepine.

	Primidone A		Primidone B	
	Geometry Optimized	Crystal Structure	Geometry Optimized	Crystal Structure
Angle made between pyrimidine ring and phenyl group [º]	100.80	90.70	101.00	107.00
Torsion angle of pyrimidine ring with ethyl group [º]	-60.60	-61.50	-60.70	-62.50
Torsion angle of pyrimidine ring with phenyl group [º]	-38.40	-40.10	-41.70	-68.40

Table 3.13; molecular geometry measurements for primidone.

Although it was possible to introduce a degree of flexibility in the molecular structure, thus diminishing the need for these groups to be positioned accurately at the start, it was features such as the pyrimidine ring formation and the angle it made with the phenyl ring in the case of primidone that were surprisingly different.
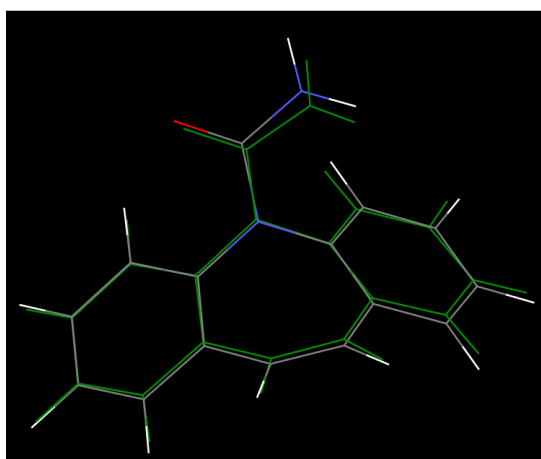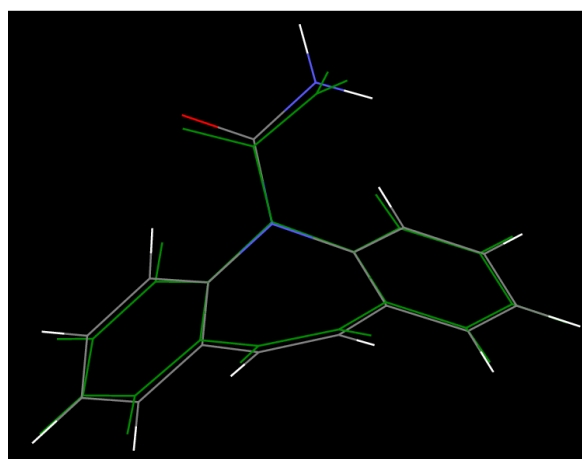
3.47 (a)   **3.47 (b)**

Figure 3.47; comparison of the crystal structures of (a) monoclinic and (b) trigonal carbamazepine (green) against the structure built and optimized in Cerius^2 (default) using MOPAC and AM1.
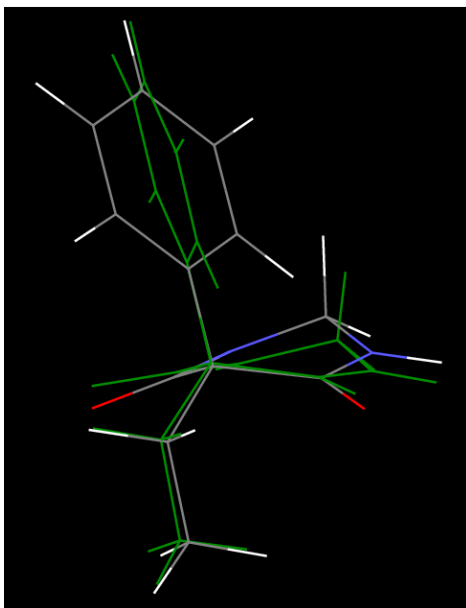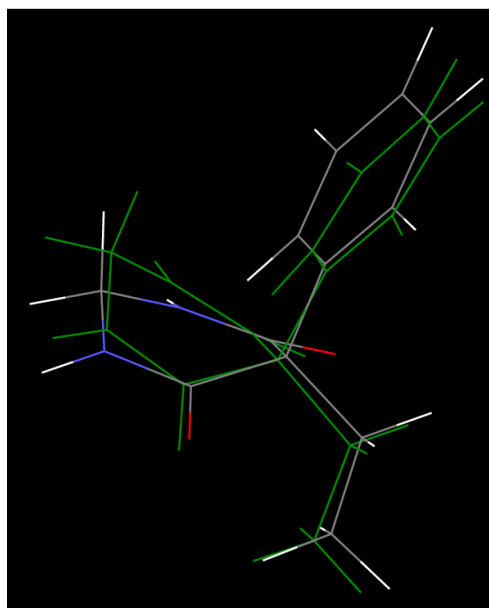
3.48 (a)   **3.48 (b)**

Figure 3.48; comparison of the crystal structures of (a) monoclinic and (b) orthorhombic primidone (green) against the structure built and optimized in Cerius^2 (default) using MOPAC and AM1.

Given the differences in molecular conformation between the polymorphs it was clearly important that the appropriate torsions were treated explicitly in the search space during trial structure determination.

As the results show, there is a significant change in the structure for primidone B highlighting the fact that it is the less stable structure. The reason for the difference in molecular conformation lies with the effect that the lattice has on the molecule. Therefore, in primidone B, the lattice must exert more of an effect on molecular conformation than in primidone A [3.57].

Considering grid step size in the systematic search methods; if the grid size is not small enough, the generated structure may be too far away from a minimum for the optimization method to find. This was often the case for the initial systematic rigid searches of trigonal carbamazepine. It was observed that large displacements of the molecule were made during the optimization step. However, making the grid search any finer would result in unreasonable calculation times even before taking any optimization steps into account. Here, the simulated annealing search did appear to be more appropriate although due to the random nature of the method, the compromise of more than one global search of the unit cell space was required.

3.9 References

[3.1]   R. B. Hammond, *Determining the Crystal Structures of Organic Solids using X-Ray Powder Diffraction Together with Molecular and Solid State Modeling Techniques,* Mol. Cryst. and Liq. Cryst., **356,** 389-405 (2001)

[3.2]   P. G. Fagan, *An Ab Initio Approach to Crystal Structure Determination Using High Resolution Powder Diffraction and Computational Chemistry Techniques,* Conference Proceedings Series Crystal Growth of Organic Materials, **August 27-31,** 22-27 (1995)

[3.3]   R. B. Hammond, *Crystal Structure Determination for Pharmaceutical Materials From Powder X-Ray Diffraction Combined with Molecular Modelling,* Industrial Crystallization, 1-10 (1999)

[3.4]   M. J. Brodie, *Established Antiepileptic Drugs,* Seizure, **6,** 159-174 (1997)

[3.5]   M. A. Dichter, *Old and New Mechanisms of Antiepileptic Drug Actions,* New Antiepileptic Drug Development: Preclinical and Clinical Aspects (Epilepsy Res. Suppl.), **10,** 9-17 (1993)

[3.6]   R. L. MacDonald, *Antiepileptic Drug Mechanisms of Action,* Epilepsia, **36**(Suppl. 2), S2-S12 1995

[3.7]   P. J. P. Reboul, *5H-Dibenz[b,f]azepinecarboxamide-5 (Carbamazepine),* Acta Cryst., **B37,** 1844-1848 (1981)

[3.8]   V. L. Himes, *Structure of Carbamazepine: 5H-Dibenz[b,f]azepine-5-carboxamide,* Acta Cryst., **B37,** 2242-2245 (1981)

[3.9]   M. M. J. Lowes, *Physicochemical Properties and X-Ray Structural Studies of the Trigonal Polymorph of Carbamazepine,* J. Pharm. Sci., **76**(9), 744-752 (1987)

[3.10] R. Ceolin, *X-Ray Characterization of the Triclinic Polymorph of Carbamazepine,* J. Pharm. Sci., **86**(9), 1062-1065 (1997)

[3.11] A. L. Grzesiak, *Comparison of the Four Anhydrous Polymorphs of Carbamazepine and the Crystal Structure of Form I,* J. Pharm. Sci., **92**(11), 2260-2271 (2003)

[3.12] M. Lang, *Form IV of Carbamazepine,* J. Pharm. Sci., **91**(4), 1186-1190 (2002)

[3.13] C. Chang, *The Crystal Structures of (S) and (R) Baclofen and Carbamazepine,* Crystallography in Biochemistry and Pharmacology, Acta Cryst., **A37,** 71 (1981)

[3.14] R. K. Harris, *Structural Studies of the Polymorphs of Carbamazepine, its Dihydrate, and Two Solvates,* Organic Process Research & Development, **9,** 902-910 (2005)

[3.15] G. Reck, *The Order-Disorder Structure of Carbamazepine Dihydrate: 5 H-Dibenz[b,f]azepine-5-carboxamide Dihydrate, $C_{15}H_{12}N_2O \cdot 2H_2O$,* Cryst. Res. Technol., **21**(11), 1463-1468 (1986)

[3.16] E. Laine, *Formation of Dihydrate from Carbamazepine Anhydrate in Aqueous Conditions,* Int. J. Pharm., **20,** 307-314 (1984)

[3.17] L. E. McMahon, *Characterization of Dihydrates Prepared from Carbamazepine Polymorphs,* J. Pharm. Sci.,**85**(10), 1064-1069 (1996)

[3.18] P. J. P. Reboul, *Structure de la 5H-Dibenzo[b,f]azépine (Iminostilbène), Support Tricyclique d'Analogues Structuraux des Antidépresseurs Imipraminiques,* Acta Cryst., **B36,** 2683-2688 (1980)

[3.19] P. Kahela, *Pharmakokinetics and dissolution of Two Crystalline Forms of Carbamazepine,* Int. J. Pharm., **14,** 103-112 (1983)

[3.20] F. U. Krahn, *Effect of Type and Extent of Crystalline Order on Chemical and Physical Stability of Carbamazepine,* Int. J. Pharm., **53,** 25-34 (1989)

[3.21] R. J. Behme, *Heat of Fusion Measurement of a Low Melting Polymorph of Carbamazepine that Undergoes Multiple-Phase Changes During Differential Scanning Calorimetry Analysis,* J. Pharm. Sci., **80**(10), 986-990 (1991)

[3.22] J. Han, *Applications of Pressure Differential Scanning Calorimetry in the Study of Pharmaceutical Hydrates I. Carbamazepine Dihydrate,* Int. J. Pharm., **157,** 209-218 (1997)

[3.23] Y. Li, *In Situ Dehydration of Carbamazepine Dihydrate: A Novel Technique to Prepare Amorphous Anhydrous Carbamazepine,* Pharmaceutical Development and Technology, **5**(2), 257-266 (2000)

[3.24] R. J. Roberts, *Mechanical Property Predictions for Polymorphs of Sulphathiazole and Carbamazepine,* European Journal of Pharmaceutical Sciences, **9,** 277-283 (2000)

[3.25] C. Rustichelli, *Solid State Study of Polymorphic Drugs: Carbamazepine,* Journal of Pharmaceutical and Biomedical Analysis, **23**(1), 41-54 (2000)

[3.26] D. Murphy, *Solution-Mediated Phase transformation of Anhydrous to Dihydrate Carbamazepine and the Effect of Lattice Disorder,* Int. J. Pharm., **246,** 121-134 (2002)

[3.27] S. G. Fleischman, *Crystal Engineering of the Composition of Pharmaceutical Phases: Multiple-Component Crystalline Solids Involving Carbamazepine,* Crystal Growth & Design, **3**(6), 909-919 (2003)

[3.28] J. Y. Bogue, *The Evaluation of "Mysoline" – A New Anticonvulsant Drug,* Brit. J. Pharmacol., **8,** 230-236 (1953)

[3.29] R. Handley, *Mysoline: A New Drug in the Treatment of Epilepsy,* The Lancet, **259**(6711), 742-744 (1952)

[3.30] M. A. Dichter, *New Antiepileptic Drugs,* Drug Therapy, **334**(24), 1583-1590 (1996)

[3.31] D. G. R. Yeates, *The Crystal Structure of Primidone,* Acta Cryst., **B31,** 1077-1082 (1975)

[3.32] R. D. Daley, *Primidone,* Analytical Profiles of Drug Substances, **2,** 409-437 (1973)

[3.33] M. P. Summers, *Preparation and Properties of Solid Dispersion System Containing Citric Acid and Primidone,* J. Pharm. Sci., **65**(11), 1613-1617 (1976)

[3.34] R. S. Payne, *The Mechanical Properties of Two Forms of Primidone Predicted from their Crystal Structures,* Int. J. Pharm., **145,** 165-173 (1996)

[3.35] R. S. Payne, *Examples of Successful Crystal Structure Prediction: Polymorphs of Primidone and Progesterone,* Int. J. Pharm., **177,** 231-245 (1999)

[3.36] L. Pauling, *The Shared-Electron Chemical Bond,* Proc. Natl. Acad. Sci., **14**(4), 359-362 (1928)

[3.37] E. D. Isaacs, *Covalency of the Hydrogen Bond in Ice: A Direct X-Ray Measurement,* Phys. Rev. Lett., **82**(3), 600-603 (1999)

[3.38] S. Shaik, *A Chemist's Guide to Valence Bond Theory,* Wiley & Sons, Inc. (2008)

[3.39] F. Cordier, *Observation of Through-Hydrogen-Bond $^{2h}J_{HC'}$ in a Perduterated Protein,* J. Mag. Res., **140,** 510-512 (1999)

[3.40] L. Pauling, *The Structure and Entropy of Ice and Other Crystals with Some Randomness of Atomic Arrangement,* J. Am. Chem. Soc., **57,** 2680-2684 (1935)

[3.41] E. Arunan, *Definition of the Hydrogen Bond,* IUPAC Provisional Recommendation, (2004-026-2-100)

[3.42] P. A. Kollman, *The Theory of the Hydrogen Bond,* Chem. Rev., **72**(3), 283-303 (1972)

[3.43] B. W. Gung, *Enhanced Intramolecular Amide-Amide Hydrogen Bonding Through Cooperativity,* Tetrahedron Letters, **37**(13), 2189-2192 (1996)

[3.44] A. C. Legon, *Directional Character, Strength, and Nature of the Hydrogen Bond in Gas-Phase Dimers,* Acc. Chem. Res., **20,** 39-46 (1987)

[3.45] A. R. P. Jagarlapudi, *The Role of Cl···Cl and C-H···O Interactions in the Crystal Engineering of 4-Å Short-Axis Structures,* Acc. Chem. Res., **19,** 222-228 (1986)

[3.46] T. Steiner, *The Hydrogen Bond in the Solid State,* Ang. Chem. Int. Ed., **41**(1), 48-76 (2002)

[3.47] G. R. Desiraju, *The C-H···O Hydrogen Bond in Crystals: What is it?,* Acc. Chem. Res., **24,** 290-296 (1991)

[3.48] R. Taylor, *Crystallographic Evidence for the Existence of C-H···O, C-H···N, and C-H···Cl Hydrogen Bonds,* J. Am. Chem. Soc., **104,** 5063-5070 (1982)

[3.49] M. D. Joesten, *Hydrogen Bonding and Proton Transfer,* Journal of Chemical Education, **59**(5), 362-366 (1982)

[3.50] J. Donohue, *The Hydrogen Bond in Organic Crystals,* J. Phys. Chem., **56**(4), 502-510 (1952)

[3.51] P. L. Wash, *Acid-Amide Intermolecular Hydrogen Bonding,* J. Am. Chem. Soc., **119,** 3802-3806 (1997)

[3.52] F. H. Allen, *Systematic Analysis of the Probabilities of Formation of Biomolecular Hydrogen-Bonded Ring Motifs in Organic Crystal Structures,* New. J. Chem., 25-34, (1999)

[3.53] G. R. Desiraju, *Hydrogen Bridges in Crystal Engineering: Interactions without Borders,* Acc. Chem. Res., **35**, 565-573 (2002)

[3.54] L. Leiserowitz, *Molecular Packing Modes. Part III. Primary Amides,* J. Chem. Soc. A: Inorganic, Physical, Theoretical, 2372-2382 (1969)

[3.55] (a) H. Szatyłowicz, *Structural Aspects of the Intermolecular Hydrogen Bond Strength: H-Bonded Complexes of Aniline, Phenol and Pyridine Derivatives,* J. Phys. Org. Chem., **21,** 897-914 (2008); (b) G. H. Stout, *X-ray Structure Determination 2nd Edition A Practical Guide,* Wiley and Sons Inc. (1989)

[3.56] R. A. Sykes, *New Software for Statistical Analysis of Cambridge Structural Database Data,* J. Appl. Cryst., **44,** 882 – 886 (2011)

[3.57] (a) M. C. Etter, *Hydrogen Bonds as Design Elements in Organic Chemistry,* J. Phys. Chem., **95,** 4601-4610 (1991); (b) M. C. Etter, *Graph-Set Analysis of Hydrogen-Bond Patterns in Organic Crystals,* Acta Cryst., **B46,** 256-262 (1990); (c) J. Bernstein, *Decoding Hydrogen –bond Patterns. The Case of Iminodiacetic Acid,* J. Ch*em. Soc. Perkin Trans. 2.,* **Issue 5,** 695-698, (1990); (d) M. C. Etter, *Encoding and Decoding Hydrogen-bond Patters of Organic Compounds,* Acc. Chem. Res., **23,** 120-126 (1990)

[3.58] R. Taylor, *Comparison of X-ray and Neutron Diffraction Results for the N-H···O=C Hydrogen Bond,* Acta Cryst., **B39,** 133-138 (1983)

[3.59] A.J. Cruz-Cabeza, *Amide Pyramidalization in Carbamazepine: A Flexibility Problem in Crystal Structure Prediction?*, Cryst. Growth and Des., **6**(8), 1858-1866 (2006)

[3.60] (a) P. Dauber, *Crystal Packing, Hydrogen Bonding, and the Effect of Crystal Forces on Molecul*ar Con*formation,* Acc. Chem. Res., **13,** 105-112 (1980); (b) M. A. Buntine, *Influence of Crystal Packing on Molecular Geometry: A Crystallographic and Theoretical Investigation of Selected Diorganotin Systems,* J. Phys. Chem., **A102,** 2472-2482 (1998); (c) J. Bernstein, *Crystal Growth, Polymorphism and Structure-Property Relationships in Organic Crystals,* J. Phys. D., **B26,** 66-76 (1993); (d) J. Bernstein, *Conformational Polymorphism. The Influence of Crystal Structure on Molecular Conformation,* J. Am. Chem. Soc., **100(3)**, 674-681 (1978)

[3.61] R. B. Hammond, *Computationally Assisted Structure Determination for Molecular Materials from X-ray Powder Diffraction Data*, J. Phys. Chem. B., **101**, 6532-6536 (1997)

[3.62] S. L. Mayo, *DREIDING: A Generic Force Field For Molecular Simulations,* J. Phys. Chem., **94**(26), 8897-8909 (1990)

[3.63] J. P. M. Lommerse, *A Test of Crystal Structure Prediction of Small Organic Molecules,* Acta Cryst., **B56**, 697-714 (2000)

[3.64] (a) K. Berland, *van der Waals Forces in Density Functional Theory: A Review of the vdW-DF Method,* Rep. Prog. Phys., **78**, 66501-1 – 66501-41 (2015); (b) J. Klimeš, *Perspective: 50 Years of Density-functional Theory in Physics,* J. Chem. Phys., **140**, 18A301 (2014); (c) J. Gräfenstein, *An Efficient Algorithm for the Density-Functional Theory Treatment of Dispersion Interactions,* J. Chem. Phys., **130**, 124105-1 – 124105-16 (2009); (d) J. van de Streek, *Validation of Experimental Molecular Crystal Structures with Dispersion-Corrected Density Functional Theory Calculations,* Acta Cryst., **B66**, 544-558 (2010); (e) J. Nyman, *Accurate Force Fields and Methods for Modelling Organic Crystals at Finite Temperatures,* Phys. Chem. Chem. Phys., **18**, 15828-15837 (2016)

[3.65] G. M. Day, *Significant Progress in Predicting the Crystal Structures of Small Organic Molecules – A Report on the Fourth Blind Test,* Acta Cryst., **B65**, 107-125 (2009)

[3.66] http://openmopac.net/

CHAPTER 4 – CRYSTAL STRUCTURE DETERMINATION OF THE COCRYSTAL CATECHOL UREA

4.1 Introduction

In the stable crystalline state, molecules will tend to arrange themselves in a way to minimize the overall potential energy of the system. The atoms in the molecules move towards each other to maximize the attractive forces, yet remain far enough away to minimize the repulsive forces. The actual value lies at around the minimum of the atom-atom potential energy curve for the respective atoms involved. This was the close packing theory proposed by Kitaigorodsky [4.1].

When molecular crystals do undergo reactions they start from an extremely ordered state, therefore the resulting structure is generally predictable as there are only a set number of ways molecules can approach each other. In other words the reactions occur under topochemical control, where the nature of packing is more important than molecular reactivity [4.2]. Here, orientation and molecular separation are the key factors which define how molecules may approach and therefore react.

However, this is not always the case and deviations to expected results do occur [4.3]. There are two main reasons why this should happen;

1. Polymorphism – molecules and crystals can adopt different spatial arrangements [4.4]
2. Crystal defects – such as dislocations where there are pockets of space for new routes of approach.

The type of external energy applied to a system will classify the field of chemistry a reaction belongs to. Mechanochemistry [4.5] is the term applied when a reaction is caused by mechanical energy. A subclass of mechanochemistry is tribochemistry, which has been defined as 'a reaction generated by friction during the milling of solid reagents' [4.6].

When selecting the starting materials for solid state reactions such as milling, two key points were noted from the research published by Etter et al. [4.8], also referenced as [3.54(a)]. These are;

"at least one of the components should have some volatility at the temperature of the grinding experiments, and the product dimers should have a stronger intermolecular hydrogen bond than any of the hydrogen bonds in the structures of the two starting materials".

Solid state reactions have a number of characteristics making them a preferable synthesis route over solvent crystallization [4.7].

These advantages include;

- The starting materials are safer and easier to keep in their un-solvated forms.

- The starting materials have poor solubility, therefore designing solid phase experiments would be easier.

- The possibility of solvent inclusion in the final product is eliminated.

- In general, smaller quantities of starting materials are required for solid phase reactions. Consequently, solid state reactions can be environmentally favourable as well as economically preferable.

- Single crystals are not always required for structure determination and having a reliable, repeatable alternative eliminates the need for this step.

In the interests of balance however, solvent crystallization does allow for a purification step and there is a chance of increased crystalline disorder from grinding. Also, solvent crystallization can produce high quality large single crystals, which are better to evaluate surface features and crystal habits [4.6].

4.2   Crystal Engineering

By combining two different starting materials, without breaking any covalent bonds, the overall properties of the resulting material may be altered. Interest in this kind of 'crystal engineering' [4.9] has developed rapidly over the last couple of decades especially for effect chemicals such as pharmaceuticals. Properties of active ingredients such as dissolution rates, mechanical properties and stability have been investigated and 'designed' for many products [4.10]. Recently the area of computer-assisted drug design has also been a great boon in developing *de novo* design approaches [4.11].

The goal of Structure-based De Novo Design (SBDND) is to develop new molecules by stitching molecular fragments together [4.12]. The fragments will be selected based on rules set by the practitioner so that the final molecule will fit into the receptor-binding

site. There are many methods to choose from in order to build a molecule and there is also the consideration of whether to use atom or fragment based methods. As the name implies, atom based methods start building with an atom or a very small fragment and are also known as 'growing' methods. Fragment methods start with larger fragments and are also known as linking methods. Linking methods tend to reduce the search space available due to their larger starting sizes. Evolutionary Algorithm methods, such as the Genetic Algorithm, are used when fragments have been selected. The linking of the fragments is carried out with these stochastic searches within the binding site search space. Initially, *de novo* programs used one objective as an acceptance criteria such as binding strength within the binding site. The more modern *de novo* approaches work with a combination of objectives such as good synthetic feasibility, higher potency and lower toxicity. These constraints are important to define since they effectively guide the design of the molecule. Possible results are assessed using scoring methods. These include; force-field based scoring functions, Empirical scoring functions, Knowledge-based scoring functions and Consensus scoring.

The terminology for combined or engineered crystal structures has come under some debate in the past. The three main terms which have been used are;
1. Inclusion compounds, which can be broken down into two groups [4.13]:
   - The host molecule contains voids into which suitable guest species can fit.
   - The crystalline structure of the host creates cavities into which suitable guest molecules can fit.
2. Complexes. In a chemical sense, the term complex describes a central structure connected to surrounding atoms or molecules. Dissatisfaction with this term has arisen from it not defining the physical state of the constituents adequately.
3. The term pharmaceutical cocrystals is a more modern description which has been used to describe any crystalline structure comprising two or more components with at least one component being an active pharmaceutical ingredient [4.15]. Although some are dissatisfied with this definition, it has been widely accepted for use.

Extensive studies of the crystalline host inclusion compounds have been carried out for example by Palin et al with their work on β-quinol [4.14].
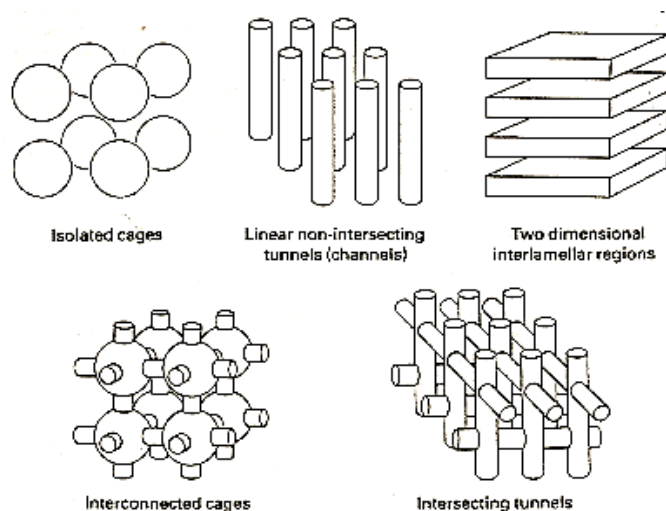
Examples of some topologies of inclusion compound cavities are shown in figure 1 when the host is a crystalline solid. However, the ambiguity of whether the compound is in the crystalline form has caused some unease in using this term for crystalline structures.

Figure 4.1; topologies of some inclusion cavities [4.13].

In many instances of inclusion compounds, the host molecule forms extended hydrogen bonds which depending on the backbone, leaves large 'holes' in the structure. This means that the most efficient packing arrangement has not been reached, but the hydrogen bonds, which can sometimes be only a weak interaction, do seem to compensate for this feature. The guest molecule can then position itself inside the holes formed by the host, this structure is called a clathrate. In other cases the host-guest relationship produces a much larger network of links with the possibility of proton transfer along linear chains.

It is well known that the physical properties of polymorphic compounds vary and are difficult to control, however pharmaceutical cocrystals provide an opportunity to engineer new solid forms with desired physical properties [4.16 - 4.20]. For example, a pharmaceutical cocrystal which is an 'easy to handle', stable crystalline solid may result from combining an amorphous API, which is difficult to crystallize, with a cocrystal former (CCF). Cocrystals can be designed to provide desired properties, such as solubility [4.21] and therefore bioavailability, to a much greater extent and with much more control [4.22]. The desired use of the cocrystal will also influence the choice of CCF and with pharmaceutical cocrystals, the selection of CCF will be restricted to the GRAS list (Generally Regarded As Safe) [4.23].

Understanding how a crystal structure affects mechanical properties enhances the quality of product formulation [4.24]. For example, the inclusion of a coformer with an API which creates slip planes in the cocrystal structure can increase the plasticity and therefore tableting ability [4.25]. Or increasing the shelf life of a drug by creating a cocrystal with

increased stability to humidity [4.26a and 4.26d]. Poor water solubility of various pharmaceuticals has led to them being used in case studies to either improve solubility as a cocrystal and/or to improve their performance as an API [4.26b]. At the other end of the scale, slow release formulations have also been investigated using cocrystal formation in order to reduce the solubility of an API [4.26c]. In summary, by exploiting concepts in manufacturing sciences to reduce production costs whilst streamlining drug development, the pharmaceutical manufacturing industry has embraced the philosophy of "Quality by Design" [4.9(b)].

4.3   Polymorphism in Cocrystals

With that said regarding the ease of controlling the design and production of cocrystals, polymorphism in cocrystals has been also reported [4.27 – 4.34]. Therefore, due to the increasing interest in cocrystals as pharmaceuticals, it becomes as important to investigate the polymorphic behaviour of cocrystals as it is for the component parts [4.27, 4.35]. Using the CSD as a data source, information on the structural stability of known hydrogen-bonded systems can be used to predict the stability of newly constructed systems [4.36, 4.37]. It is reasonable to assume that confidence in a proposed structure will be higher if the prevalence of similar examples of intramolecular geometry and intermolecular interactions is high. Applying a statistical assessment to hydrogen bond behaviour to complement experimental protocols could assist in identifying potential structural modifications and the likelihood of 'hidden' polymorphs being formed. Therefore, combining computational assessments with experimental processes could highlight whether sufficient screening has been carried out.

There are different types of polymorphism in cocrystals some of which are described here;

4.3.1 Synthon Polymorphism

When the hydrogen bond motifs of cocrystals differ, they are considered to be synthon polymorphs. An example of this feature is shown for the co-crystals of 4-hydroxybenzoic acid/4,4'-bipyridine
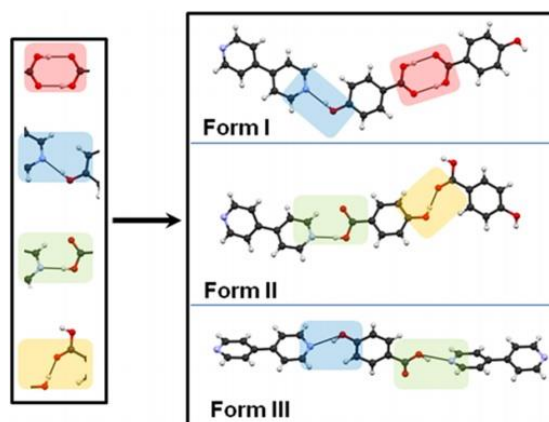
Figure 4.2; an example of synthon polymorphism, from [4.30b]

4.3.2 Conformational Polymorphism

If the conformation of one or more of the molecules within the cocrystals differ, this is classed as conformational polymorphism. This feature is commonly exhibited with molecules having multiple degrees of conformational flexibility.

This has been rationalised due to the fact that energies required to rotate bonds are similar in magnitude to the lattice energy differences between polymorphs. Conformational polymorphism appears more frequently than synthon polymorphism [4.27].

4.3.3 Packing Polymorphism

As the name suggests and as is the case with single component crystals, polymorphs are identified as packing polymorphs when their packing arrangements differ. This is most commonly seen with when the molecule is a rigid structure.

4.4 Cocrystal Polymorph Screening

Design methods for co-crystals are multistage processes. In order to select a suitable API/CCF system, there are many hurdles to overcome. Some strategies for design include trial and error approaches, supramolecular synthon searches (using the CSD), Solubility Predictions [4.43, 4.45], computational prediction [4.46] and intermolecular interaction characterization [4.44].
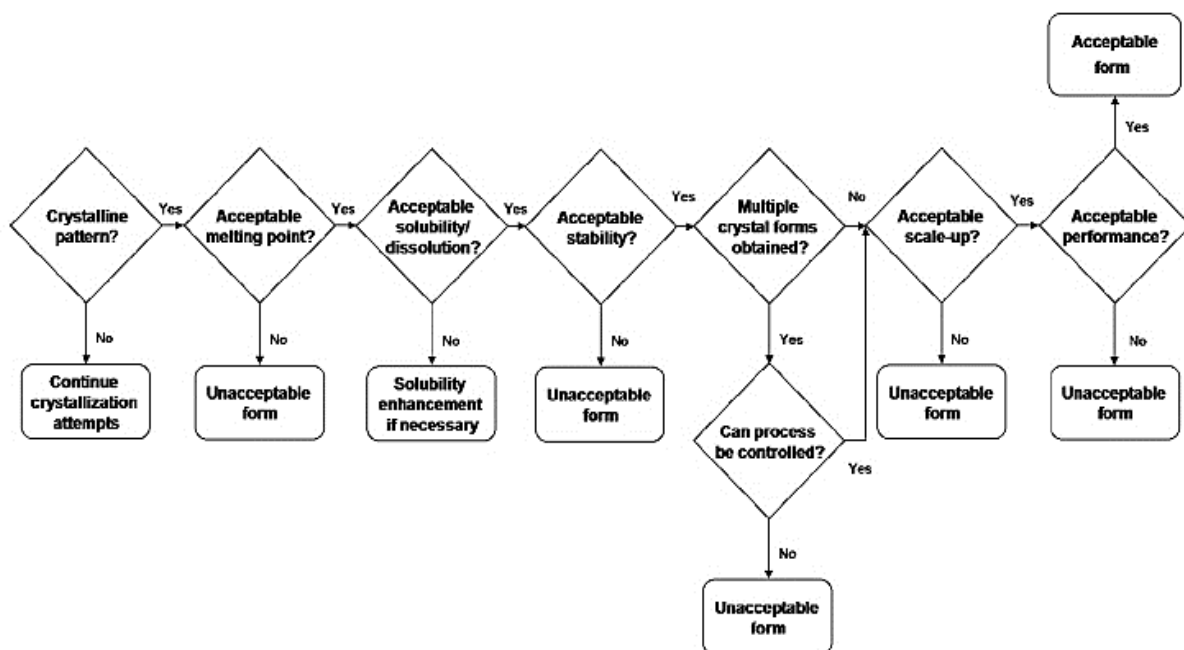
Figure 4.3; cocrystal design strategies. From [4.20b].

4.4.1   Knowledge-based Screening Methods

This technique involves assessing the predictability of interactions between proton acceptors and donors in neutral organic compounds [4.39, 4.40, 4.53 and 4.54]. This methodology is based on the empirical hydrogen-bonding rules laid out by Etter *et al.* [4.41].

Figure 4.4; identification of possible synthons is the first step. Here is shown a homo-synthon (left) and a hetero-synthon (right) [4.18].

In a similar approach [4.42] a two tier screening method was employed. In tier 1, CCFs with the highest potential for hydrogen-bonding with the API were selected. To add a level of diversity to this selection process, molecules with a range of molecular volumes were considered. Successful combinations (cocrystal 'hits') provided the knowledge for the second tier of coformer selection. In tier 2, further diversity is introduced by choosing new sets of coformers which have similar molecular structures to those which had proved successful in tier 1.

4.4.2 Hydrogen Bond Motif Searches

H-bond motif searches, which are very similar to the idea expressed in 4.4.1, however, this approach uses Motif search technology within Mercury CSD [4.47] to search for competitive hydrogen-bonding groups.

4.4.3 Multi Component Hydrogen-Bond Propensity

In effect, this is also a knowledge based methodology which assesses and predicts the likelihood of hydrogen bond formation [4.40, 4.48]. The Hydrogen Bond Propensity (HBP) method is automated to calculate a quantitative interaction likelihood term, the "Multi-Component Score" whilst also considering other variables such as donor/acceptor competition, steric-hindrance and the extent of aromaticity. The quantitative result is expressed as a probability (between 0 and 1) and the search can be tailored to the specific set of functional groups [4.22, 4.49].

$$MC\ Score = \begin{pmatrix} Propensity\ of\ best \\ hetero-interaction \end{pmatrix} - \begin{pmatrix} Propensity\ of\ best \\ homo-interaction \end{pmatrix} \quad (4.1)$$

4.4.4 Molecular complementarity [4.50]

The likelihood of cocrystals forming based on correlations between molecular properties of the molecules is evaluated. Molecular properties such as shape and polarity are considered to be strong factors and co-crystal formation is thought to be likely if there is a high correlation. Molecular complementarity appears to be a helpful tool when dealing with compounds which have weak hydrogen bonds or where hydrogen bonding is in competition with other intermolecular interactions such as strong stacking forces. During the tests, molecular complementarity appeared to reduce the number of proposed co-formers by over 40% implying it would be useful to insert as a filtering step before experimental screening.

4.4.5 Cohesive Energy Density (CED) Measurements [4.51]

Cohesive energy is simply described as the sum of forces holding a material intact or as the amount of energy required to break all of these interactions. Therefore, the cohesive energy density (CED) is the cohesive energy per unit volume and in non-polar systems, the CED is used to calculate the solubility parameter (Equation 4.2).

$$\delta = (CED)^{0.5} \quad (4.2)$$

In polar systems, it is necessary to include descriptions of the interatomic/intermolecular forces. In this instance, the total solubility parameter is broken down into the following partial solubility parameters (Equation 4.3), also termed Hansen Solubility Parameters (HSPs); hydrogen bonds (δ_h), dispersion forces (δ_d) and polar interactions (δ_p).

$$\delta_t = \left(\delta_d^2 + \delta_h^2 + \delta_p^2\right)^{0.5} \tag{4.3}$$

The total solubility parameter is also known as the three-dimensional solubility parameter, and various methods have been developed to calculate this value [4.51, 4.52]. A commonly used method to calculate the partial solubility parameters only requires knowledge of the materials chemical structure, the 'Group Contribution Method'.

HSPs are used to predict the miscibility of two components based on the principle that 'like dissolves like' – compounds with similar δ values will likely be miscible. As well as being a useful tool to select coformers in cocrystal screening processes, other applications of this method include predicting intestinal drug absorption properties and calculating permeation rates through human skin.

4.5   Cocrystal Formation

Numerous methods have been reported for generating cocrystals [4.42, 4.43b, 4.17, 4.37, 4.55, 4.56, 4.26, 4.57, 4.58 and 4.68- 4.74], and a few of the main processes are listed here;

- Dry grinding [4.59]/kneading [4.60, 4.38, 4.63]
- Liquid-assisted or solvent drop grinding [4.59, 4.61 and 4.66]
- Melt assisted grinding [4.62]
- Solution methods including anti-solvents [4.64, 4.65, 4.67, 4.71]

Cocrystallization from solution may be difficult if the solubility of one component differs significantly from the other. Solvent/solute interactions may also hinder the cocrystallization process. In these instances, it may prove advantageous to employ mechanochemical methods such as kneading or grinding. In cases where catalytic amounts of liquid are added to the grinding mixture, improvements in yields and in crystallinity have been noted [4.75]. Polymorphism of the reactants and the cocrystal may be controlled by careful selection of this liquid. Grinding experiments have been carried out to speculate on mechanochemical mechanisms of cocrystallization and the relative stabilities of different stoichiometric variations, i.e. when stoichiometric

variations interconvert, the relative stabilities are similar. Therefore, when the difference in relative stabilities is large, the conversion should proceed in one direction only [4.59].

4.11   Definition of Cocrystals
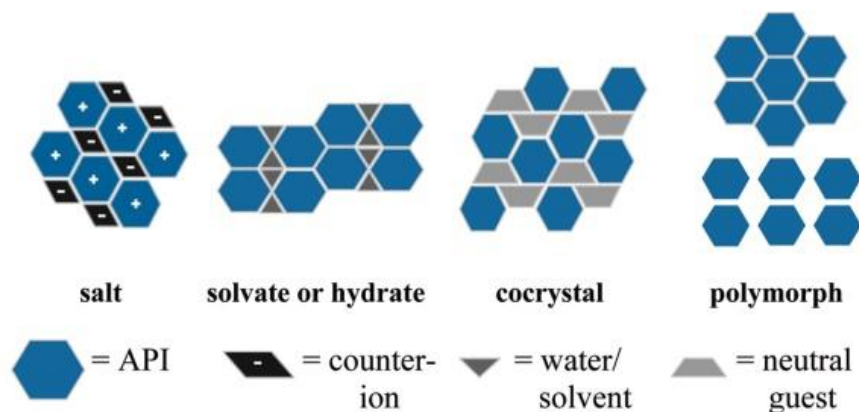
Figure 4.5; simple depiction of multicomponent systems, from [4.20b].

There have been numerous discussions regarding the term cocrystal and what that term means [4.16, 4.17, 4.22, 4.42 and 4.76]. In 2011 the FDA proposed guidance for industry on how to define a cocrystal. Following this guidance, a paper was published from the discussions arising at the Indo-U.S. Bilateral Meeting on the Evolving Role of Solid State Chemistry in Pharmaceutical Science in 2012. This discussion concluded that the FDA guidance to the definition of a co-crystal was too restrictive. While it would be convenient for guidance to have three mutually exclusive classification systems for Polymorphs, Salts and Co-crystals, in reality, such distinctions do not exist. In practical terms, the three classifications are found to overlap as shown in figure 6.

An example used to illustrate this issue was the cocrystallization of Prozac (fluoxetine hydrochloride – a salt) and carboxylic acid cocrystal formers. Multicomponent forms of the antifungal drug fluconazole with dibasic acids were also quoted in this text [4.77]. The authors argued that although it would be possible to describe these compounds scientifically, complications would arise with the regulatory classifications.

In 2013, the FDA then published a non-binding recommendation which states a cocrystal to be "Crystalline materials composed of two or more molecules within the same crystal lattice" [4.78]. In this document, co-crystals are treated as a "drug product intermediate" (DPI). Applicants for regulation would need to provide two sets of data for API-excipient molecular complexes to show that no ionic transfer is taking place.

Figure 4.6; combinations of multicomponent systems from [4.20b].

The applicant is also required to show that the API is dissociated from the coformer prior to it being pharmacologically active [4.16]. The pharmaceutical industry would obviously prefer cocrystals be classified as APIs as opposed to DPIs, which has led to discussion as to the risks involved with the practical development of cocrystals [4.79].

The European Medicines Agency published similar guidance in 2014 which similarly states;

"Although the detailed definition of cocrystals is still debated in the scientific literature, they are in general defined as homogenous crystalline structures made up of two or more components in a definite stoichiometric ratio where the arrangement in the crystal lattice is not based on ion pairing (as with salts)." [4.80]
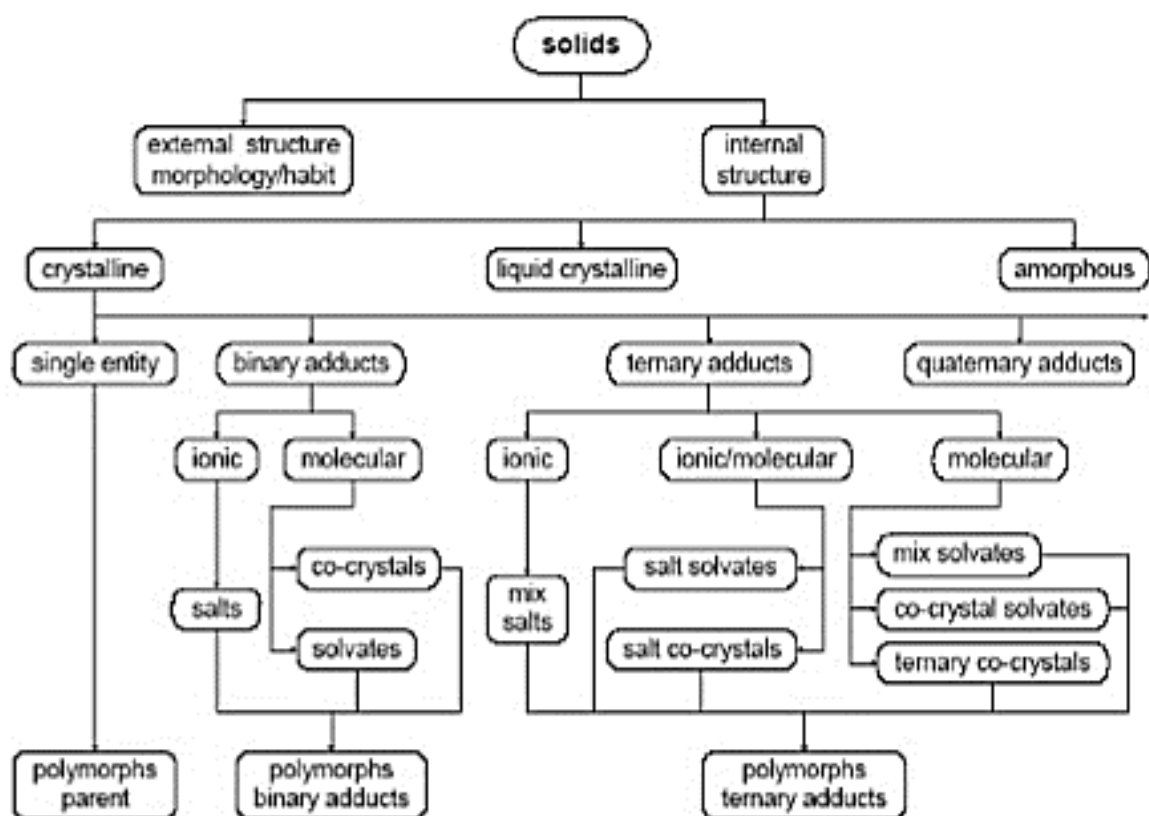
Figure 4.7; the EMA defining a cocrystal, from [4.80]

Despite the increased interest in designing pharmaceutical cocrystals [4.81], currently there are still very few approved pharmaceuticals as cocrystals being produced for commercial use. The first example approved by the US-FDA is Entresto, launched by Novartis, prescribed for the treatment of chronic heart failure [4.82] and Ipragliflozin L-proline, approved in Japan, has recently been approved worldwide [4.83].

'Drug-drug cocrystals' and 'ternary cocrystals' are sub-sets of cocrystals. As the name suggest in drug-drug cocrystals, the coformer is also a bioactive molecule rather than a benign compound. This arrangement may be desirable in combination therapies, or in order to counteract drug side effects [4.84 and 4.85]. Ternary cocrystals are formed when three neutral, solid state compounds in a fixed stoichiometry are ordered by hydrogen bonds. While hydrogen bonding is most common, halogen bonds have been used and charge transfer interactions may also play a part in the ternary system [4.25].

4.7   Urea and Hydroxybenzenes – a Review

4.7.1   Urea

Since its discovery in 1773, urea has become not only an extremely versatile but a most important chemical in all areas of industry, from agriculture to medicine. It forms interactions with many different 'guests' and in the case of organic substances it develops strong hydrogen bonded networks. Urea will readily form tunnels for suitable guest molecules to fit into but will recrystallize into its 'pure crystalline structure' if the guest is removed [4.86].

Urea crystallizes in a highly symmetrical model - Figure 8. The unit cell parameters are a = 5.58 Å, c = 4.68 Å, the space group is P421m [4.100]. The urea molecules are seen to form columns along the c axis of the unit cell which are connected by hydrogen bonds. The connection uses one hydrogen atom from each NH_2 group of the neighbouring molecule donating to the O atom - Figure 4.6 (bond type A = 2.06Å). The structure becomes further connected by the columns being hydrogen bonded together to form a sheet in the ac plane - Figure 4.7 (bond type B = 1.99Å). In this structure, all possible hydrogen bonds are formed and an unusual feature is that the oxygen atom accepts four hydrogen bonds.
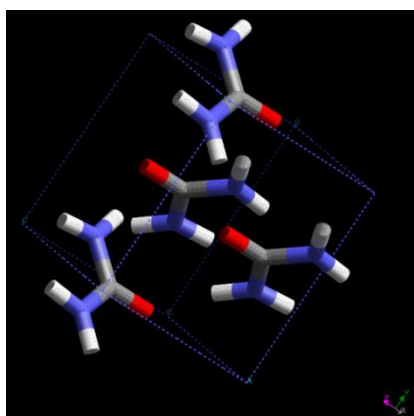
Figure 4.8; urea crystal structure.
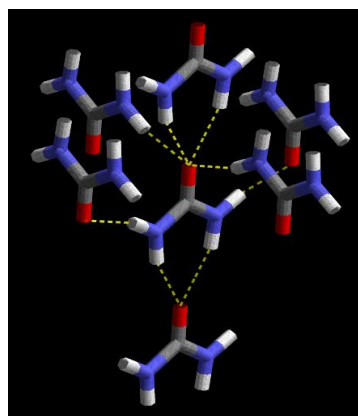
Figure 4.9; columns parallel to the c axis of urea connected by hydrogen bonds.
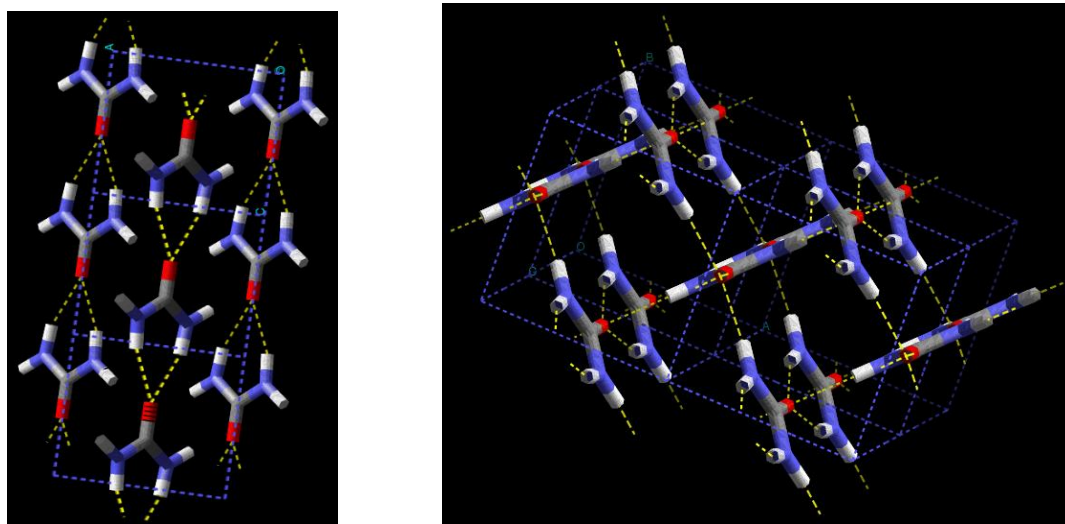
Figure 4.10; sheets of urea in the ac plane formed from the columns along the c axis being connected by hydrogen bonds.

Prior to this study, the family of hydroxybenzene:urea cocrystals had already been investigated well and these results have been used for comparative purposes. The only unsolved dihydroxybenzene:urea co-crystal structure was that of catechol urea, the focus of this study. Resorcinol and the gamma form of quinol both compared well with the properties found for catechol. The structures of the solved dihydroxybenzene:urea cocrystals appeared to have similar unit cell volumes and ratios of starting materials so it seemed a reasonable assumption that catechol urea would also behave in a similar nature. In order to attain the lowest possible potential energy it seems reasonable to also assume that the fullest complement of hydrogen bonds would be achieved. However, when the main body of the molecule under investigation is bulky or a rigid structure, this might not be possible [4.98]. This was shown to be the case for some of the hydroxybenzene molecules. In this instance, the phenyl ring would not distort easily in order to aid the formation of the hydrogen bonding system. Unusual or unexpected features may be observed as a result. A summary of physical properties is compiled in Table 1

4.7.2   Phenol

There have been two monoclinic crystal structures of phenol reported in the Cambridge Structural Database [4.87]. The unit cell parameters of the first monoclinic form are a = 6.020 Å, b = 9.040 Å, c = 15.180 Å, γ = 90.36°, space group P21. There are seven phenol molecules within the unit cell. There appears to be three different types of phenol molecules and these are shown in Figure 12, type A has a C1-O1 bond length of 1.374Å and there are three of these within the unit cell. Type B has a C7-O2 bond length of

1.376Å, type C has a C13-O3 bond length of 1.365Å, and there are two of each type of molecule within the unit cell. Hydrogen bonds form chains of molecules along the a axis. There are three lengths of hydrogen bond, repeating along the chain in the order 1.806Å, 1.976Å, 1.887Å.
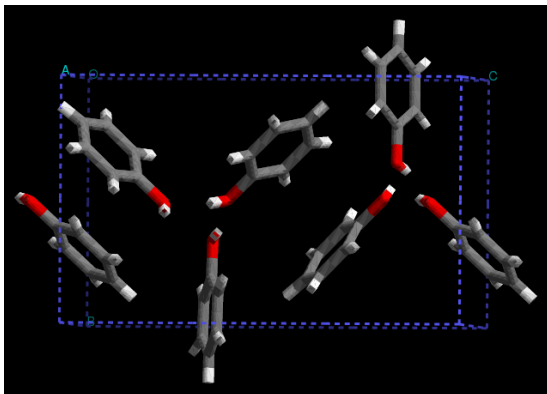
Figure 4.11; crystal structure of phenol CSD reference PHENOL03.
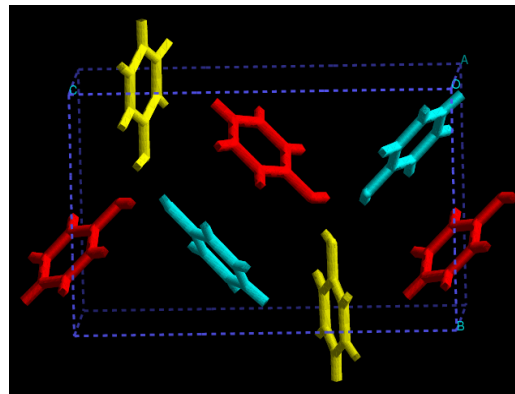
Figure 4.12; type A in red, type B in blue, type C in yellow.
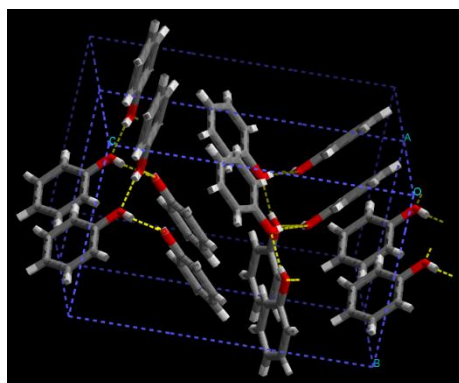
Figure 4.13; hydrogen bonds forming chains of phenol molecules along the a axis.
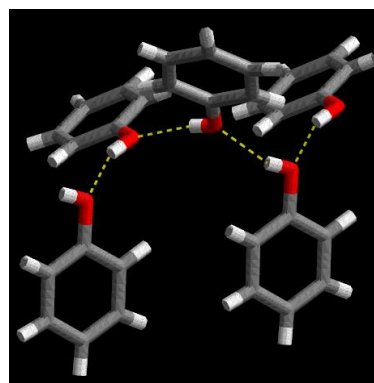
The second monoclinic structure has unit cell parameters of a = 11.61Å, b = 5.442 Å, c = 12.217Å, β = 101.47°, space group P21. There are six phenol molecules within the unit cell. Again there appear to be three different types of phenol molecules, type A has a C1-O1 bond length of 1.351Å, there are two of these within the unit cell. Type B has a C7-O2 bond length of 1.353Å, type C has a C13-O3 bond length of 1.338Å, also two of each type of molecule within the unit cell. Hydrogen bonds form chains of molecules along the b axis. There are two lengths of hydrogen bond, alternating along the chain in the order 2.232 Å, 2.2.44Å.
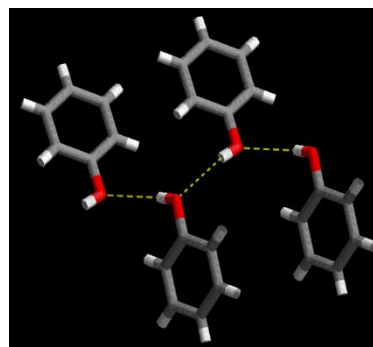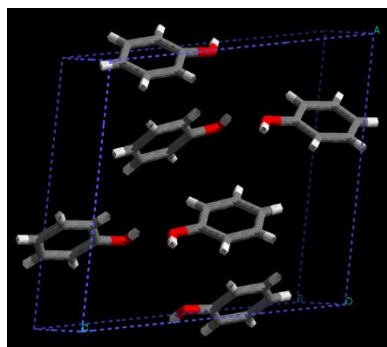
Figure 4.14; crystal structure of phenol CSD reference PHENOL11.
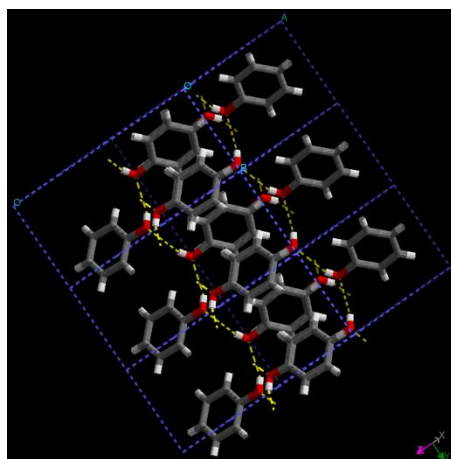
Figure 4.15; hydrogen bonded chains
of phenol molecules along the a axis.
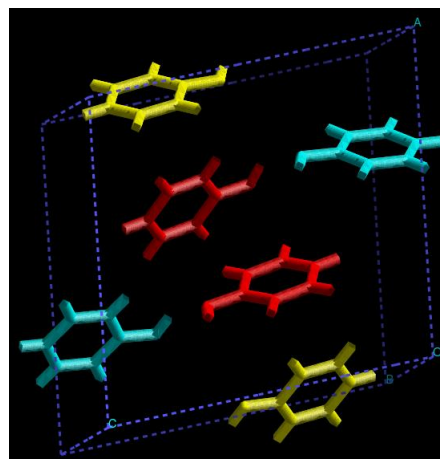
Figure 4.16; type A in red, type B
in blue, type C in yellow.

4.7.3   Diphenol:Urea

The crystal structure of this co-crystal was reported in 1987, the unit cell parameters are
a = 26.93 Å, b = 6.65Å, c = 7.43 and β = 92º, space group Cc [4.88].   There are three
molecules in the asymmetric unit (one urea two phenol) and two asymmetric units in the
unit cell - Figure 4.14.   As Figure 4.14 shows there is no possible bonding between the
urea molecules and there are two columns of phenol molecules in between each column
of urea molecules.
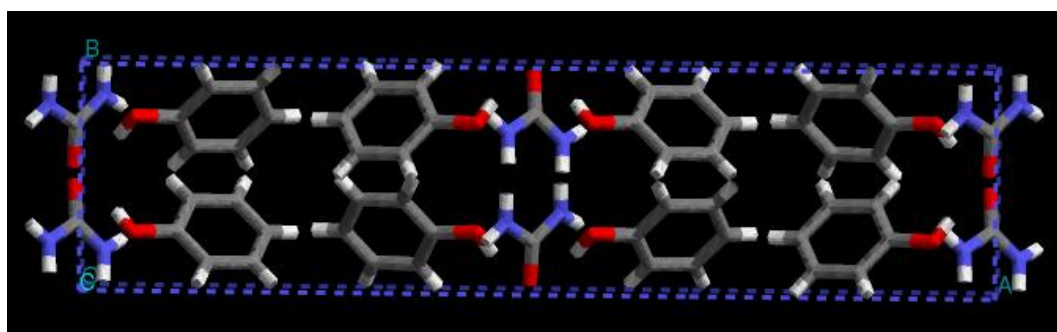
Figure 4.17; crystal structure of diphenol.urea.

A urea molecule uses its oxygen lone pairs to hydrogen bond to two phenol molecules forming a $O_u \cdots H_p$-O interactions (1.822Å, 1.849 Å) and secondly by donating one hydrogen atom from both NH_2 groups to an oxygen on two more phenol molecules to form $NH_u \cdots O_p$ (2.108 Å, 2.251 Å). O_u and H_u are the urea oxygen and urea hydrogen respectively, H_p is the phenol donating hydrogen and O_p is the phenol oxygen. This interaction of one urea molecule to four phenol molecules forms chains along the b axis Figure 4.15. The structure is further expanded when the urea molecule creates its fifth hydrogen bond with another sheet forming $NH_u \cdots O_p$ bond (2.292 Å).
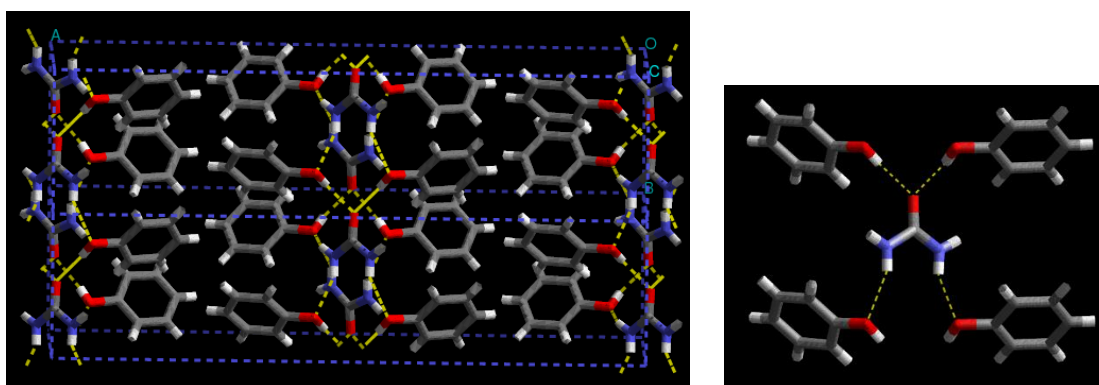
Figure 4.18; hydrogen bonding chain network in diphenol urea.
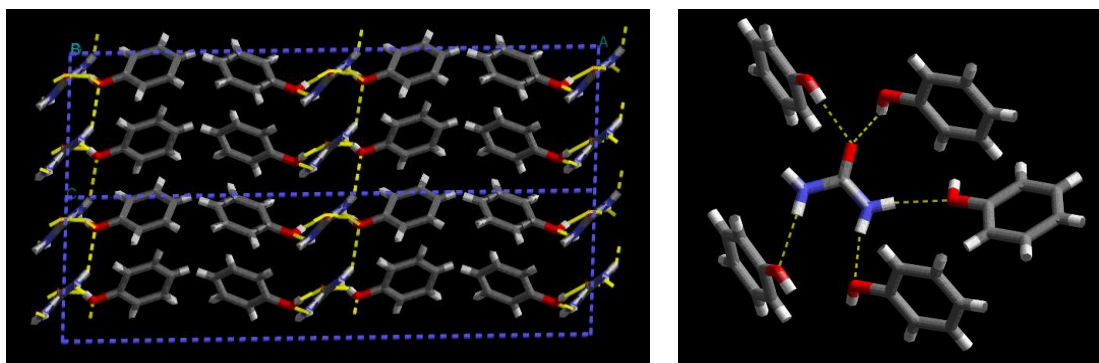
Figure 4.19; hydrogen bonding sheets in diphenol urea.

4.7.4   Resorcinol

Two orthorhombic forms of resorcinol have been reported in the Cambridge Structural Database [4.89]

The first, α-form was reported in 1936 with the co-ordinates published in 1973. The unit cell dimensions are a = 10.53Å, b = 9.53Å, c = 5.66Å, space group Pna. There are four molecules within the unit cell with the hydroxyl groups on each molecule directed away from each other. The Oxygen atoms in the hydrogen bonding hydroxyl groups are

approximately 2.70Å apart and each resorcinol molecule is attached to four other molecules. The four molecules do not create a closed unit. The molecules create hydrogen bonds with other molecules in adjoining unit cells to create a 'spiral' effect along the c-axis. These spirals are then connected through further hydrogen bonding.
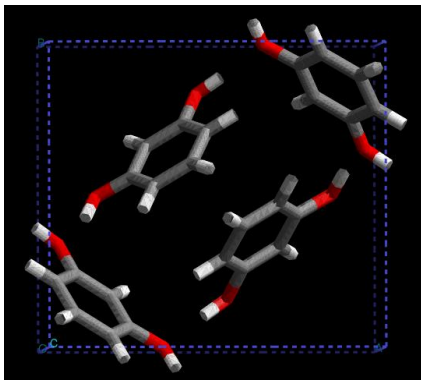
Figure 4.20; crystal structure of resorcinol (α-form).
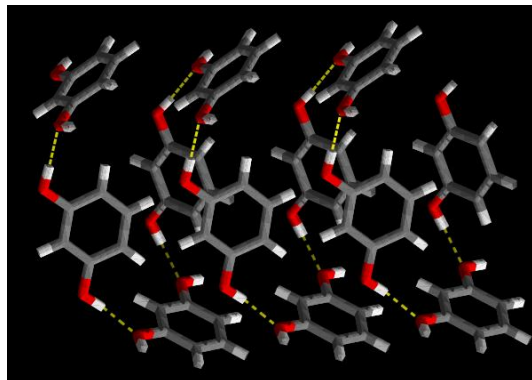
Figure 4.21; spirals of resorcinol molecules along the c axis.
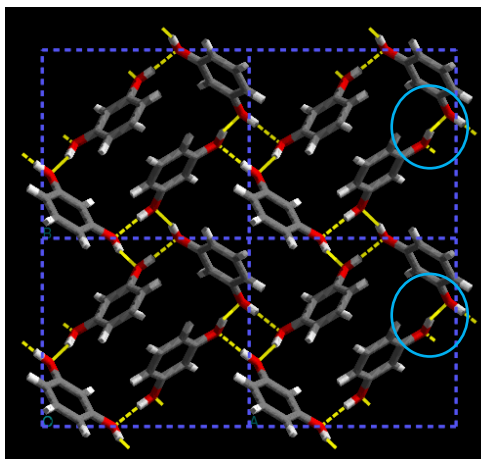
Figure 4.22; spirals along c connected by hydrogen bonds along the a-axis.

The second, β-form was reported in 1938. The unit cell dimensions are a = 7.81Å, b = 12.61Å, c = 5.42Å, space group Pna. There are four molecules within the unit cell with the hydroxyl groups on each molecule pointing in the same direction, deviating slightly from the plane of the ring structure. The resorcinol molecules form hydrogen boneded chains along the b-axis (1.702Å) with each molecule being attached to two other resorcinol molecules. These chains are further hydrogen bonded to other resorcinol chains along the a axis (1.78Å).
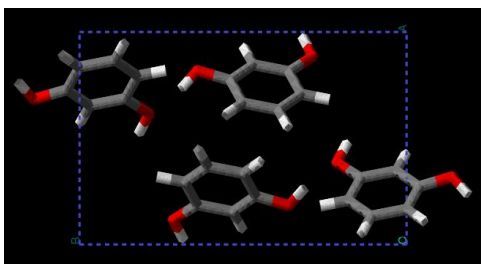
Figure 4.23; crystal structure of resorcinol (β-form).
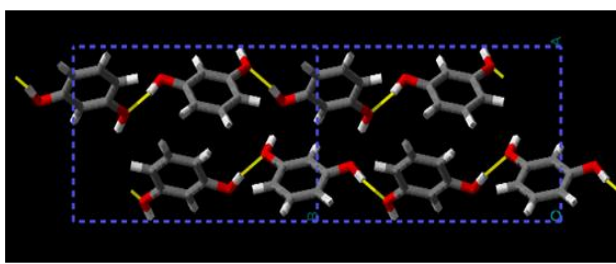
Figure 4.24; hydrogen bonded chains along the b-axis.
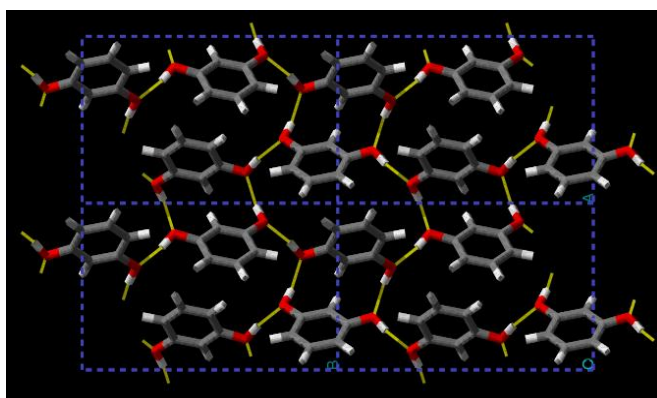
Figure 4.25; hydrogen bonded chains connect along the a-axis.

4.7.5 *Resorcinol.Urea*

The crystal structure of resorcinol:urea is shown in figure 26, the cell is orthorhombic belonging to the space group P212121 with parameters of a = 7.142Å, b= 7.798Å and c = 15.428Å [4.90]. There are two molecules in the asymmetric unit (1 resorcinol 1 urea) and there are four asymmetric units in the unit cell. The structure is built up from columns of resorcinol and columns of urea running along the b axis with two columns of resorcinol between each urea column. The hydrogen bonding network is constructed by a urea molecule using its oxygen atom to accept the hydrogens from two resorcinol molecules forming an $O_u \cdots H_r$-O bond (1.84 Å), which makes the isolated columns form chains along the b axis Figure 4.18. These chains are interconnected along the c axis by more hydrogen bonds forming between a resorcinol oxygen atom and a urea hydrogen $O_r \cdots H_u N$, (2.33Å).
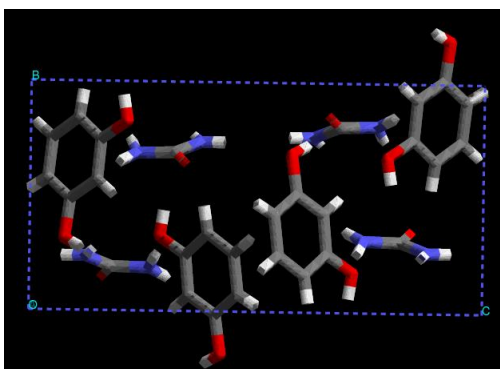
Figure 4.26; crystal structure of
resorcinol:urea.
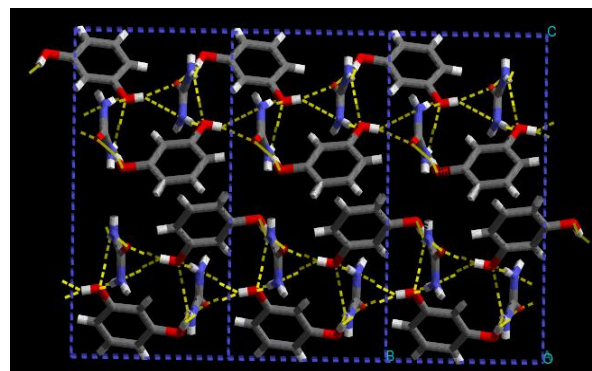
Figure 4.27; chains of resorcinol and urea
molecules along the b axis
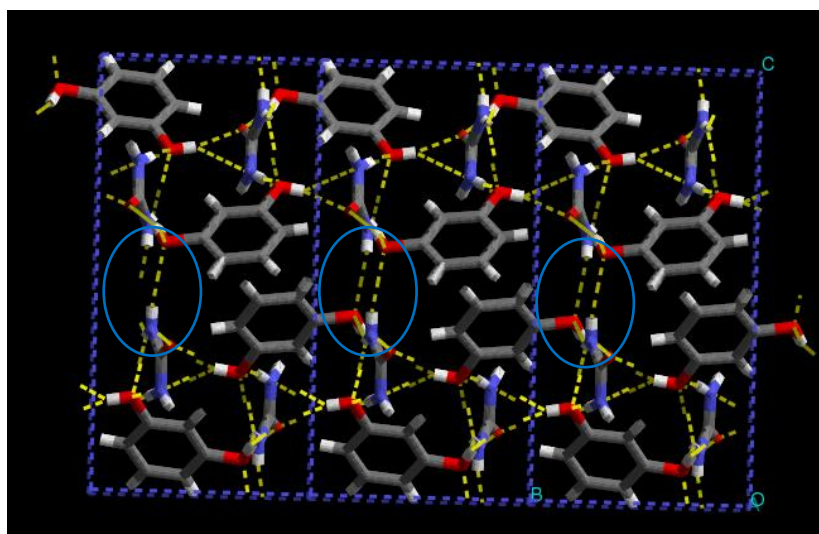
Figure 4.28; sheets forming along the c axis.

4.7.6 Quinol

There are three published forms of hydroquinone (quinol) [4.91], these are the trigonal
alpha, the trigonal beta and the monoclinic gamma forms.

The alpha form of quinol has unit cell parameters a = b = 38.46Å, c = 5.65Å, γ = 120.00°,
space group R-3 with three molecules in the asymmetric unit. Two of the molecules are
involved in making 'cages' resulting in a rather open structure, while the third forms a
double helix. The cages (1.73Å) and helices (1.83Å) are hydrogen bonded together
(1.77Å) along the c-axis resulting in the stable polymorph which is found at room
temperature.
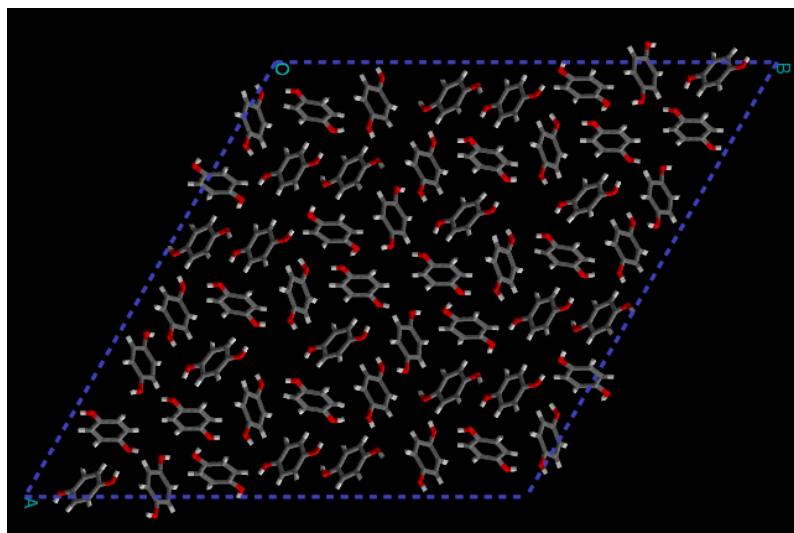
Figure 4.29; crystal structure of alpha quinol.
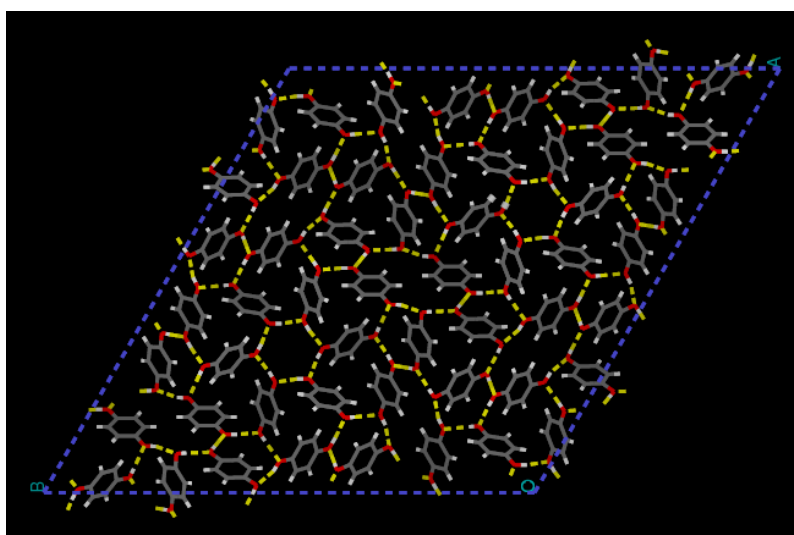
Figure 4.30; hydrogen bonding within alpha quinol.
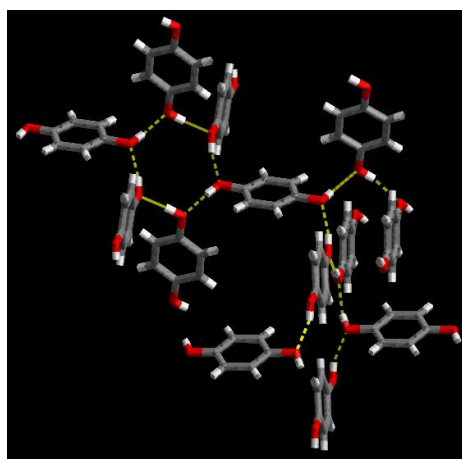
Figure 4.31; hydrogen bonded 'cage' and helix within alpha quinol.

The beta modification of quinol has unit cell parameters a = b = 16.61Å, c = 5.47Å, γ = 120.00°, space group R-3, with three molecules in the asymmetric unit. Cyclic 'cage' structures each comprising six molecules form within the unit cell via hydrogen bonds (1.90 Å) creating tunnels running along the c axis, similar to those seen in the alpha modification.
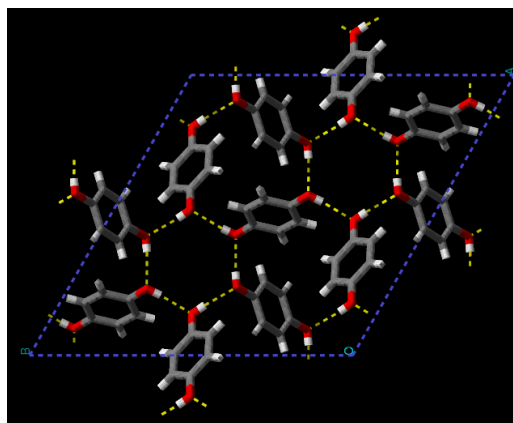
Figure 4.32; crystal structure of beta quinol showing the hydrogen bond network.

The gamma form of quinol has unit cell parameters a =8.07Å, b = 5.20Å, c = 13.20Å, β = 107.00°, space group P21/c. There are sheets of molecules formed through hydrogen bonding along the b axis (1.83Å) but these sheets are not further connected.
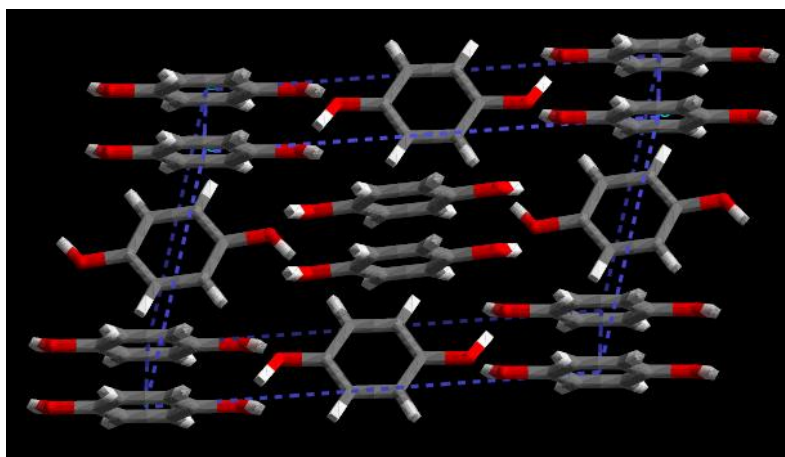
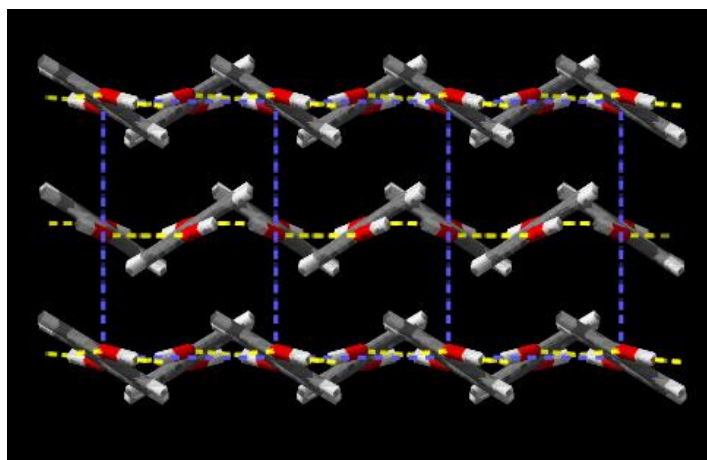**Figure 4.33;** Crystal structure of gamma quinol.

Figure 4.34; Hydrogen bonded sheets within gamma quinol.

4.7.7 Quinol Urea

Quinol:urea forms a 1:1 complex with urea in a monoclinic unit cell of dimensions a = 17.18Å, b = 6.601Å, c = 7.341Å and β = 94.4º [4.92]. The space group is P21/c with two molecules in the asymmetric unit and four asymmetric units in the unit cell. The crystal structure is shown in figure 32.
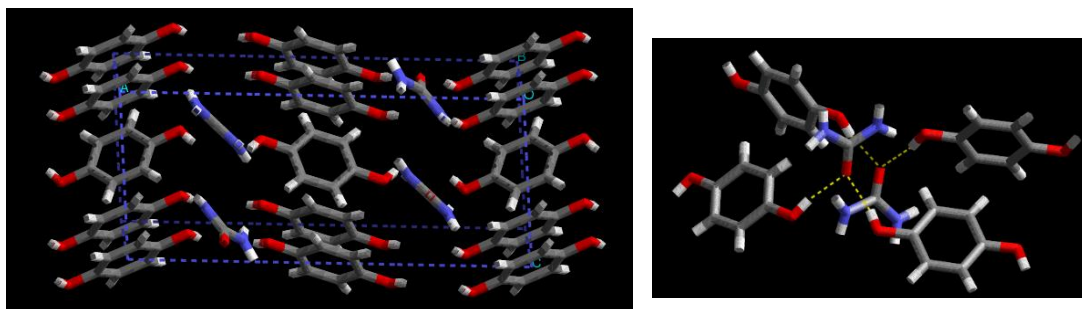

Figure 4.35; crystal structure of quinol:urea and **Figure 4.36;** hydrogen bonded chains in quinol:urea.
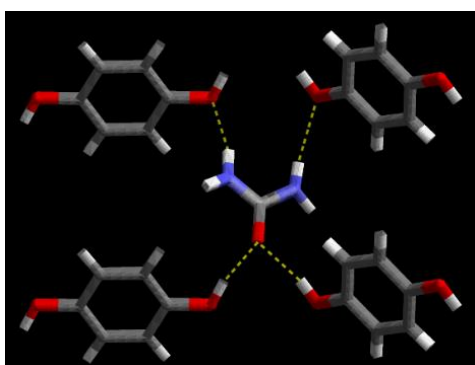

The network of hydrogen bonds is built up from the O-H$_q$···O$_u$ interaction (1.76Å and 1.80Å) along the a axis where two quinol molecules donate a hydrogen atom to the urea oxygen atom. Sheets in the ab plane are then formed from the hydrogen bond between N-H$_u$···O$_q$ (2.08Å and 2.20Å) where two quinol molecules are bonded by one hydrogen atom from both NH$_2$ groups in a urea molecule.

Figure 4.37; hydrogen bonded sheets in quinol.urea.

4.7.8 Catechol

Catechol has been used extensively in numerous industrial applications, including photography, pesticides, perfumes and medicines [4.93], as an intermediate or a precursor. Catechol occurs naturally in small amounts and is often synthesized for industrial purposes.

The structure of catechol (o-dihydroxybenzene) was published in 1971 [4.94] with unit cell dimensions of a = 10.082Å, b = 5.518Å, c = 10.943Å, β = 119.00° with space group P 21/c. The molecules form hydrogen bonded chains along the c axis (2.306Å between

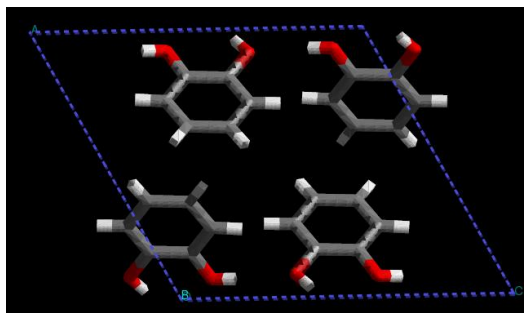HC1—OC1', and 1.991Å between HC2—OC1). Two chains are then hydrogen bonded via OC2—HC3 and OC3'—HC2'(2.124Å).

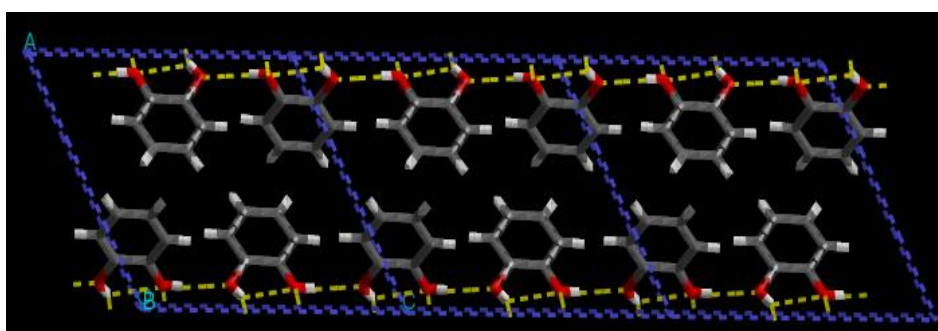

Figure 4.38; crystal structure of catechol.

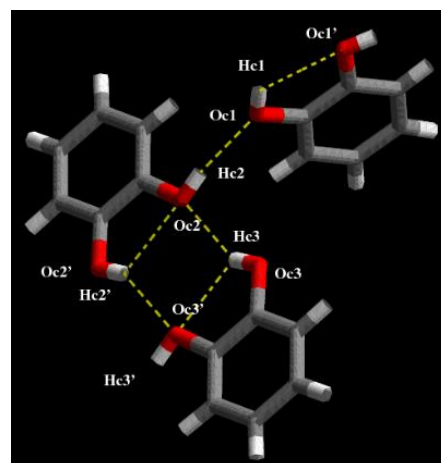**Figure 4.39;** hydrogen bonding along the c-axis in catechol.
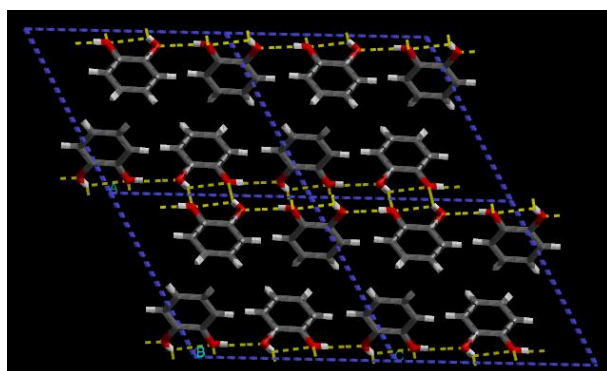
Figure 4.40; hydrogen bonding along the bc plane in catechol.

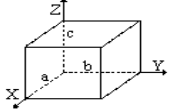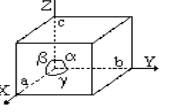| Parameter	Compound							
	Phenol	Diphenol Urea	Resorcinol	Resorcinol Urea	Quinol	Quinol Urea	Catechol	Urea
Cell Edges a, b, c [Å]	6.02, 9.04, 15.18	26.93, 6.64, 7.42	**α form** 10.53, 9.53, 5.66 **β form** 7.91, 12.57, 5.50	7.14, 7.79, 15.42	**α form** 38.46, 38.46, 5.65 **β form** 16.61, 16.61, 5.47 **γ form** 8.07, 5.20, 13.20	17.18, 6.60, 7.34	10.94, 5.51, 10.07	5.66, 5.66, 4.71
Cell Angles, α, β, γ [°]	90.0, 90.0, 90.0	90.0, 92.3, 90.0	**both forms** 90.0, 90.0, 90.0	90.0, 90.0, 90.0	**α form** 90.0, 90.0, 120.0 **β form** 90.0, 90.0, 120.0 **γ form** 90.0, 107.0, 90.0	90.0, 94.4, 90.0	90.0, 119.0, 90.0	90.0, 90.0, 90.0
Cell Volume [Å^3]	826	1328	**α form** 568 **β form** 547	859	**α form** 7238 **β form** 1308.5 **γ form** 529.7	830	530.81	151.01
Space Group	P 2221	Cc	**both forms** Pna	P 212121	**α form** R-3 **β form** R-3 **γ form** P 21/c	P 21/c	P 21/a	P -421m
Melting Point [°C]	43	60.8	**α form** 116 **β form** 109−110	110	**α form** 172-175 **β form** 166 **γ form** 169	120−130	105	133-135
Hydrogen Bond Length [Å]	2.20	2.61, 2.70, 2.95, 3.05	**α form** 2.70 **β form** 1.70, 1.78	2.696, 2.679, 2.944, 3.128	**α form** 1.73, 1.83, 1.77 **β form** 1.90 **γ form** 1.83	1.76, 1.80, 2.08, 2.20	2.12, 2.31, 1.20	2.14, 2.33

Table 4.1; physical properties of published data for urea, phenol, dihydroxybenzenes and their co-crystals.

154

4.7.9 Examination of hydroxybenzene molecular geometry

Since the backbone of a hydroxybenzene is a rigid structure, the distortion of the flexible O-H bond away from the plane of the benzene ring was assessed. This flexibility would be the best way to optimize the position of the molecule in order to achieve the most hydrogen bonds which would stabilize the overall structure.
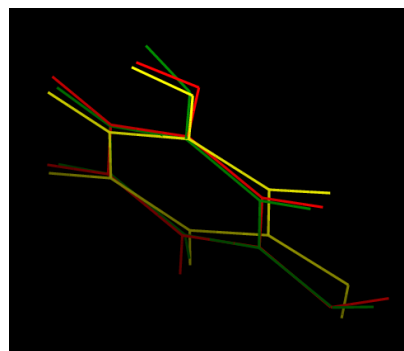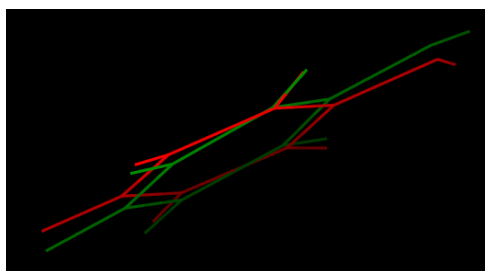

Figure 4.41; Comparison of the molecular structure of phenol in the crystal structure (red) and in the cocrystal with urea (green). **Figure 4.42;** Comparison of the molecular structures of phenol in the crystal structures (red, yellow) and in the cocrystal with urea (green).
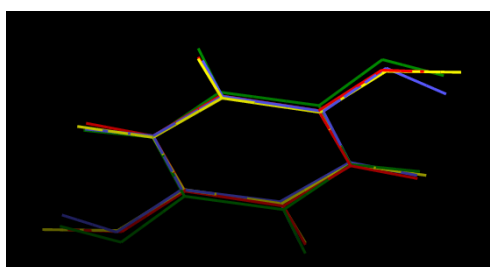

Figure 4.43; comparison of the molecular structures of quinol in the crystal structures (red, yellow, blue) and in the cocrystal with urea (green).

Hydroxyl group id	Model	C-O Bond Length	O-H Bond Length	C-O-H Bond Angle	C-C-O-H Torsion
phen03	red	1.365	0.737	102.5	-179.9
fesraw	green	1.378	0.795	128.3	-175.6
resora09	red	1.341, 1.372	1.003, 1.000	111.1, 110.2	161.7, 13.0
resora13	yellow	1.347, 1.363	0.984, 0.981	113.3, 112.3	173.5, -177.0
bodsao	green	1.372, 1.375	0.899, 0.887	109.1, 106.0	-167.7, -0.2
hyquin	yellow	1.394, 1.394	1.072, 1.072	116.5, 116.5	169.9, 10.0
hyquin02	red	1.377, 1.377	0.783, 0.783	111.6, 111.6	167.9, 12.9
hyquin05	blue	1.375, 1.303	0.791, 0.909	114.4, 108.5	-168.3, -21.2
quolur	green	1.379, 1.379	0.912, 0.912	106.6, 106.6	-180.0, 1.0
catcol12	red	1.369, 1.373	0.796, 0.814	111.1, 106.4	176.9, -170.9

**Table 4.2;** Bond lengths in Å, bond angles in ° and torsions in °.

Torsion energies of phenol, resorcinol, quinol and catechol were calculated when the hydrogen atom(s) were rotated in steps of 10° and these results are shown in Graphs 1A to 1D. Combinations of methods AM1, PM3 and EF, BFGS were used in all examples. AM1 (Austin Model 1) and PM3 (Modified Neglect of Diatomic Overlap Parametric Model Number 3) are methods used to determine the electrostatic distribution of a model. While both methods have their own limitations, they are both improvements on the MINDO and NDDO methods. EF (Eigenvector Following) and BFGS (Broyden-Fletcher-Goldfarb-Shanno) are optimization techniques based on variations of the Newton-Raphson Gradient methods.
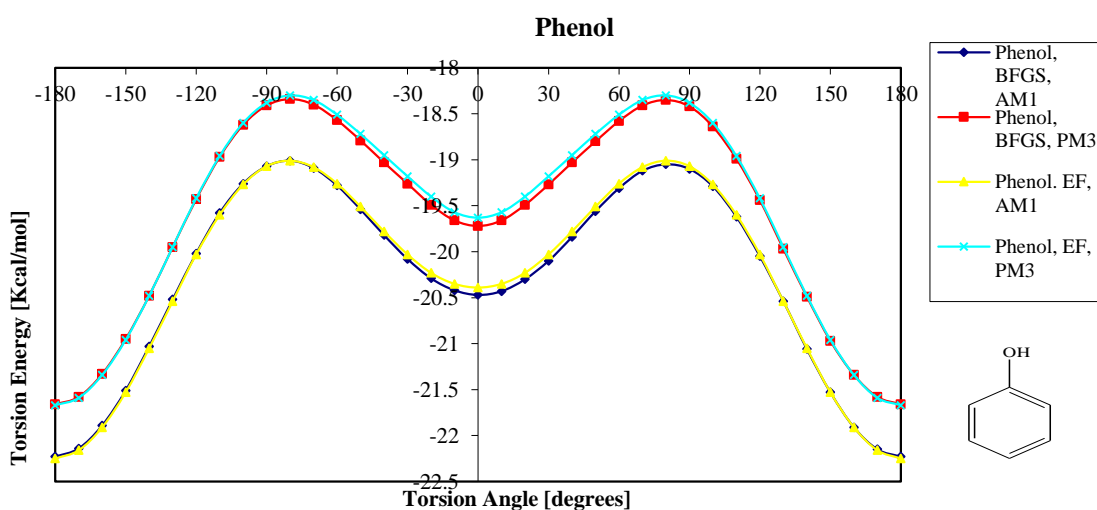
Graph 4.1A; calculating the torsion energy of phenol using various combinations of optimization and electrostatic distribution descriptors.
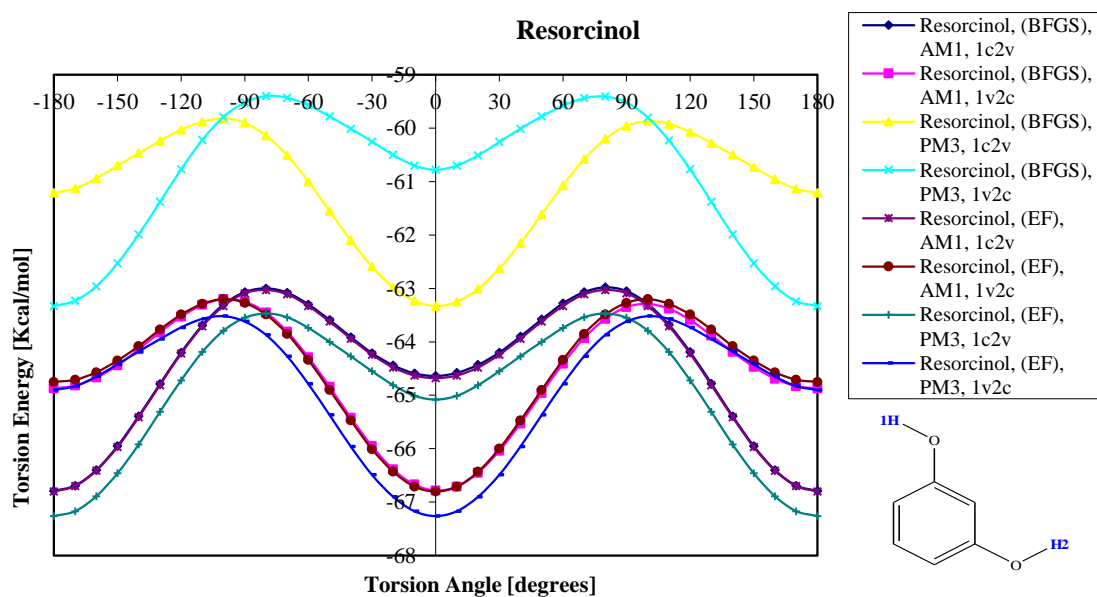
Graph 4.1B; calculating the torsion energy of resorcinol using various combinations of optimization and electrostatic distribution descriptors.
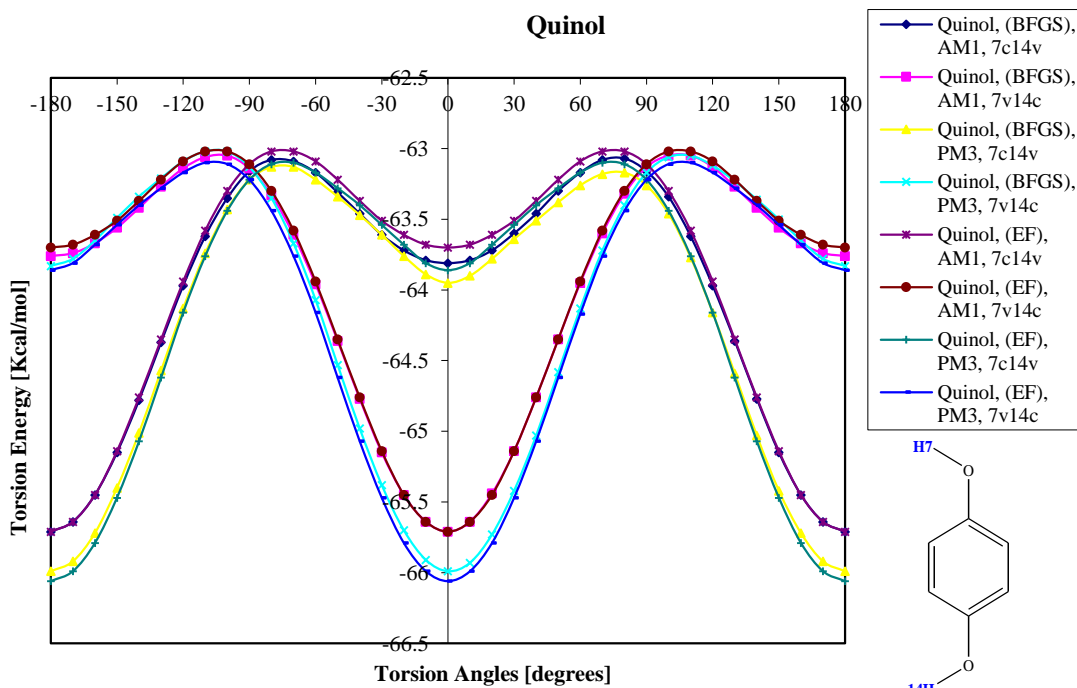
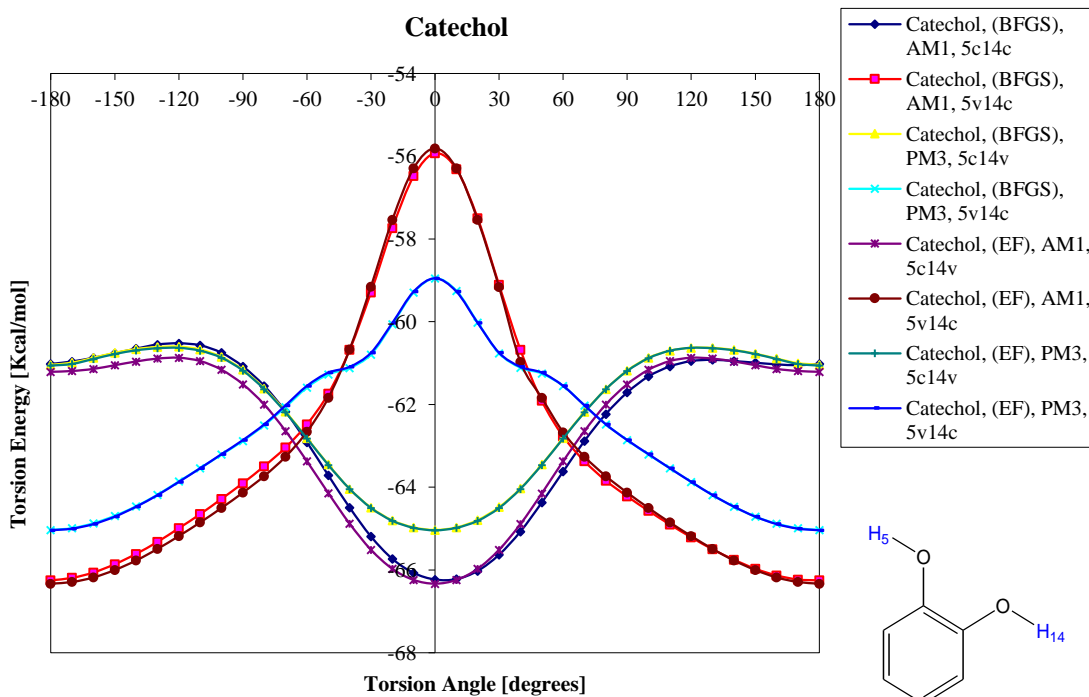**4.1C;** calculating the torsion energy of quinol using various combinations of optimization and electrostatic distribution descriptors.

4.1D; calculating the torsion energy of catechol using various combinations of optimization and electrostatic distribution descriptors

A simple piece of FORTRAN code was compiled to calculate the lattice energy of phenol, resorcinol, catechol, diphenol urea and resorcinol urea when the rigid molecules were rotated around the published co-ordinates. This was also carried out for the generated structure for catechol urea to assess the located energy minima. These results are shown in Graphs 2A to 2F.

Phenol - Calculated Gradients
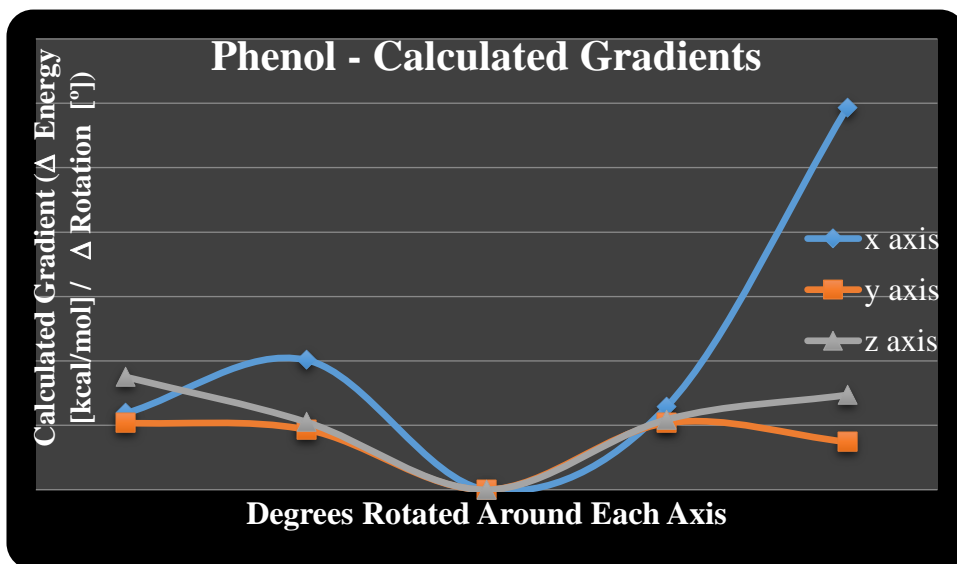
4.2A; calculating the energy gradients of phenol when the asymmetric unit is rotated around the x, y and z axes.

Diphenol Urea Calculated Gradients
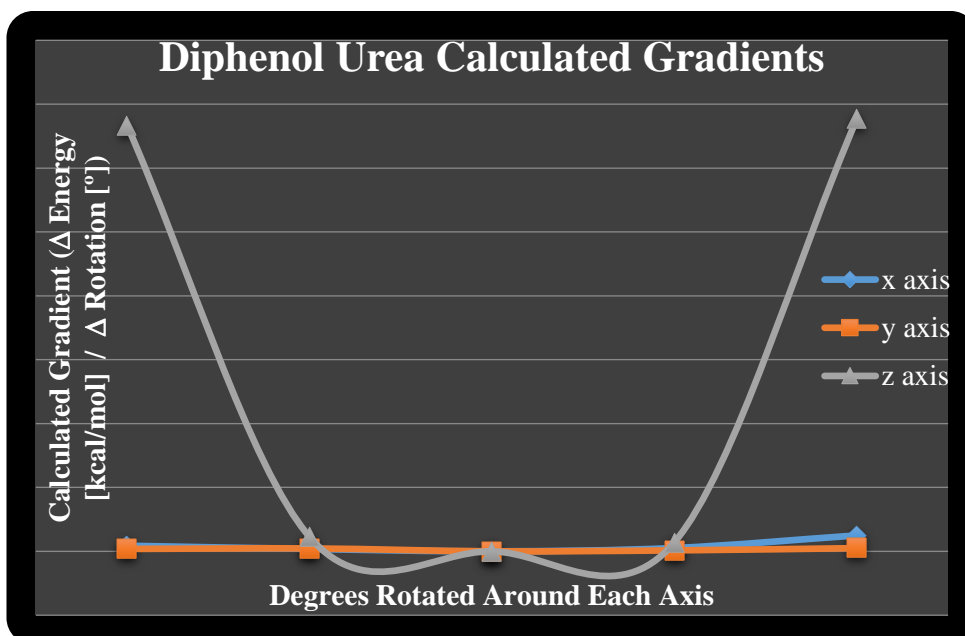
4.2B; calculating the energy gradients of diphenol urea when the asymmetric unit is rotated around the x, y and z axes.
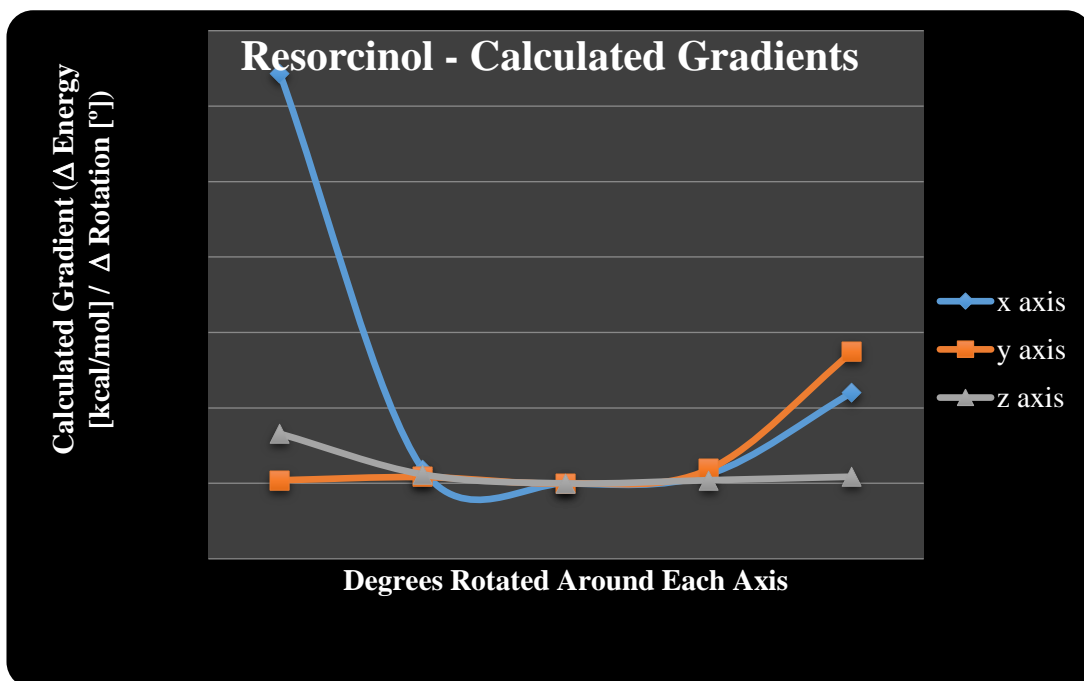
4.2C; calculating the energy gradients of resorcinol when the asymmetric unit is rotated around the x, y and z axes.

4.2D; calculating the energy gradients of resorcinol urea when the asymmetric unit is rotated around the x, y and z axes.

4.2E; calculating the energy gradients of catechol when the asymmetric unit is rotated around the x, y and z axes.

4.2F; calculating the energy gradients of catechol urea when the asymmetric unit is rotated around the x, y and z axes.

4.8 Experimental Considerations – Catechol Urea
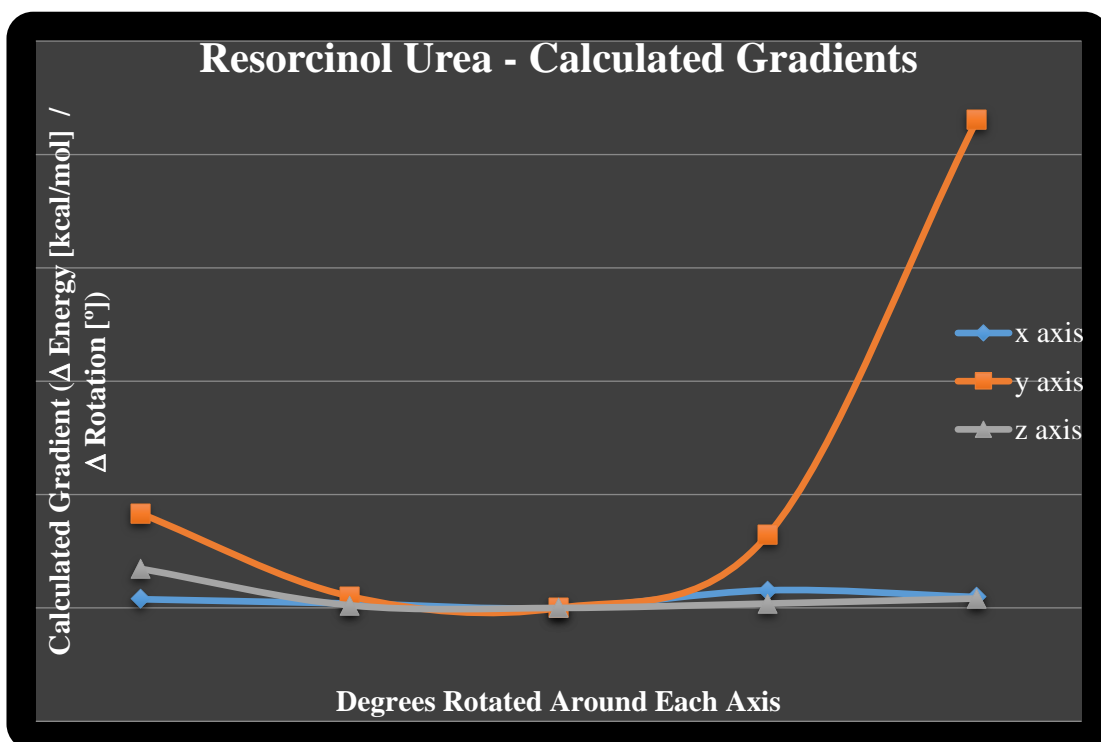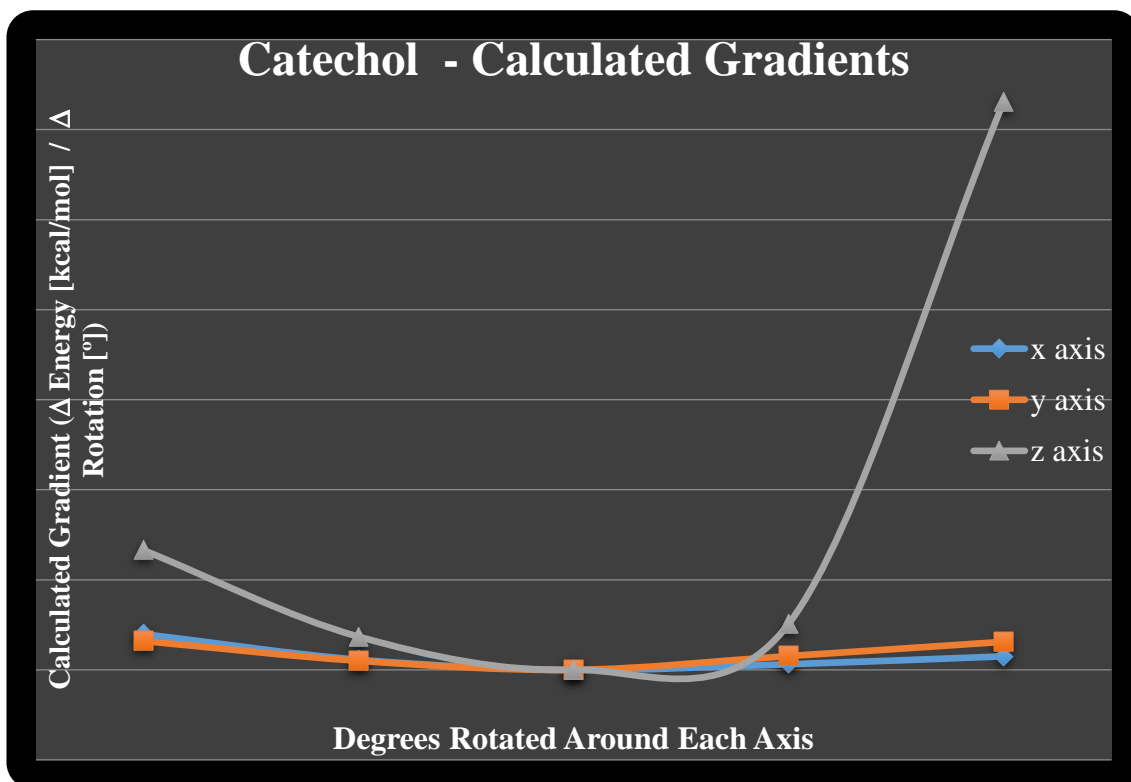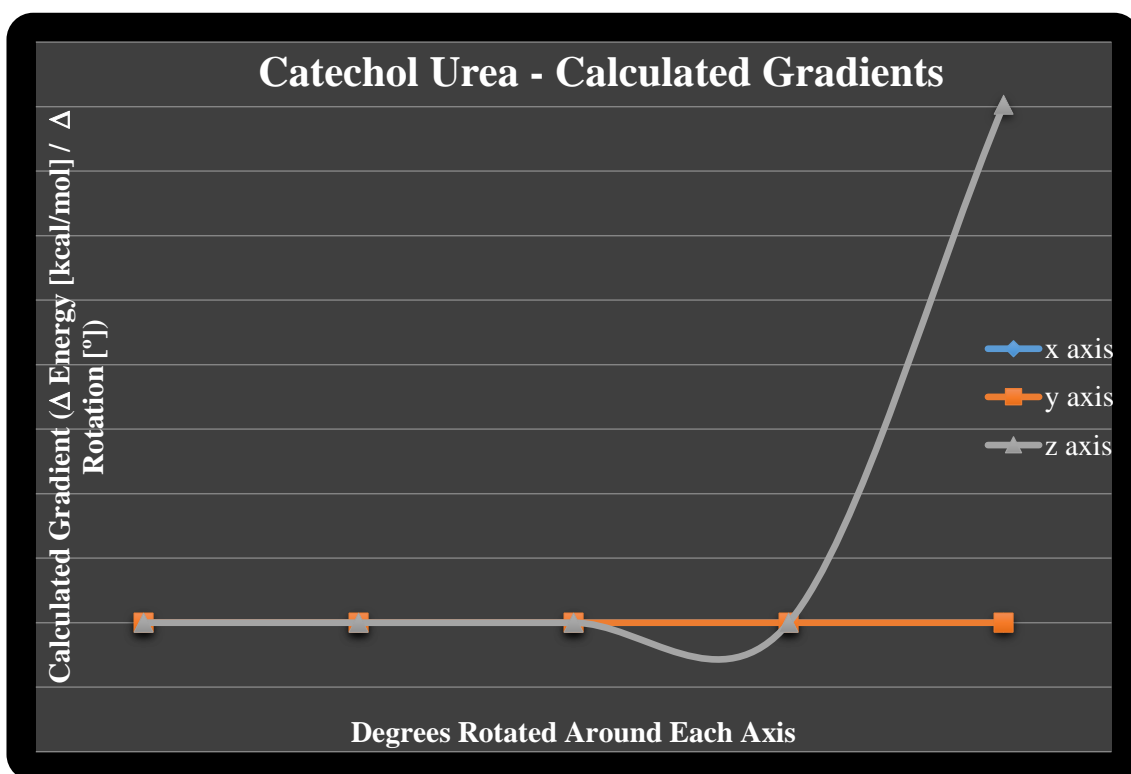
4.8.1 Solvent Crystal Growth

In the first instance, catechol:urea was crystallized from a solution of distilled water. Equal ratios of the starting materials were mixed together in water and left to slowly evaporate. Since oxidation during crystallization may arise, oxygen free nitrogen was bubbled through the solvent before use and during the experiment. The sample was also warmed gently to promote evapouration. The resulting crystals were purple in colour and were analysed using differential scanning calorimetry and x-ray powder diffraction.

4.8.2 Solid State Reactions

Following the set of solvent experiments, a solid state reaction was carried out where samples of the starting materials catechol and urea were accurately weighed and ground together in a mortar and pestle from one minute for up to forty minutes. Samples were collected at 5 minute intervals and analysed using differential scanning calorimetry and x-ray powder diffraction.

4.9 Theoretical Considerations – Generating a Structure of Catechol/urea

4.9.1 Trial Structure Generation

The experimental data collected for the catechol:urea complex was indexed using TREOR and DICVOL to define the unit cell parameters and a suitable space group. The published data for resorcinol:urea was also indexed using these programs for comparison of a known structure. Since there were two independent molecules in the asymmetric unit, trial structures were generated using the 'in-house' simulated annealing program previously validated with primidone and carbamazepine.

At this time, the systematic search method could only cope with one independent molecule in the asymmetric unit. If this method was to be used for trial structure generation, the arrangement of the catechol molecule with respect to the urea molecule would have needed to be determined. The search could then proceed with the two molecules being considered as a single rigid unit. Since the arrangement of the catechol and urea molecules wasn't previously known, it was preferable to search the unit cell keeping the molecules independent from each other. The published data for resorcinol/urea and quinol/urea were also searched for comparative purposes.

4.9.2 Post Search analysis

Once a structure for catechol/urea had been generated using the simulated annealing search, it was optimized using the crystal packer module in Cerius2. Any significant changes were noted, starting and final atomic positions were compared to see how far the optimization had moved the molecules which would indicate how well the search was performing.

4.9.3 Refinements

The refinement package GSAS was used throughout this study and is discussed in Chapter 2. It allowed for the molecular positions, unit cell parameters and peak profile fits to be optimized. The diffraction pattern using these new parameters was simulated and compared with the experimental pattern. It is hoped that the refinement does not produce significant changes and that the trial structure generated is close to the optimized structure.

4.10   Results

4.10.1 Differential Scanning Calorimetry

A series of experiments using the differential scanning calorimeter were carried out alongside the x-ray powder diffraction analysis. Ground samples of catechol urea were extracted at the beginning of grinding (after 1minute) and thereafter at 5 minute intervals. These samples were compared with the crystals grown from crystallizing catechol and urea using water as a solvent (shown in the series of Figure 45). A gradual change in the content of the samples was seen through a shift in the melting point of the samples.
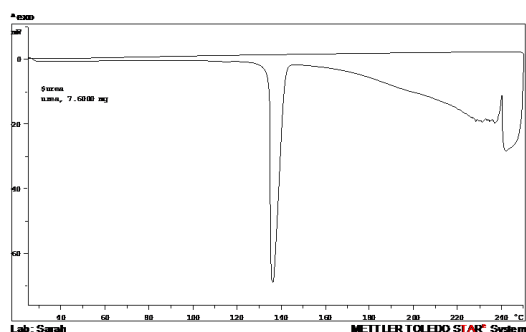
Figure 4.44A; sample of urea powder. **Figure 4.44B;** sample of catechol powder.
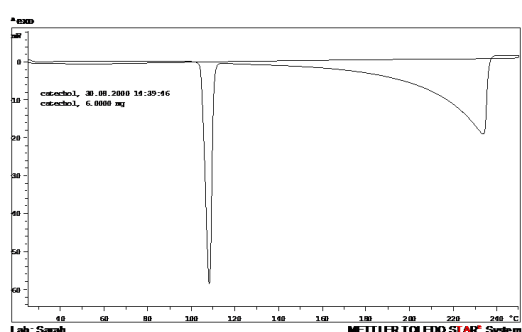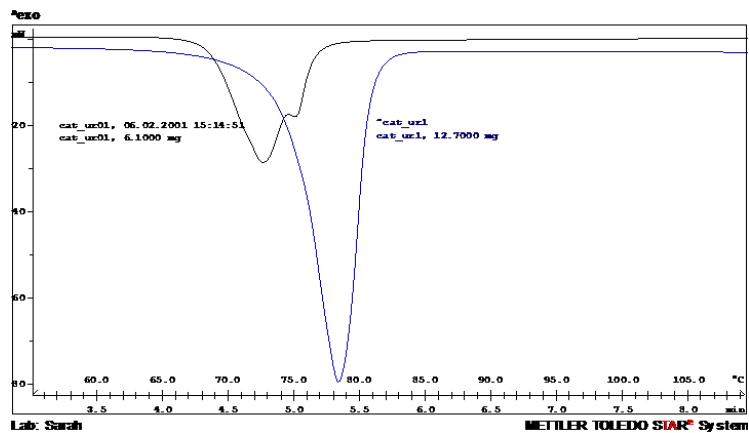
Figure 4.45A; sample of ground catechol:urea after 1 minute.
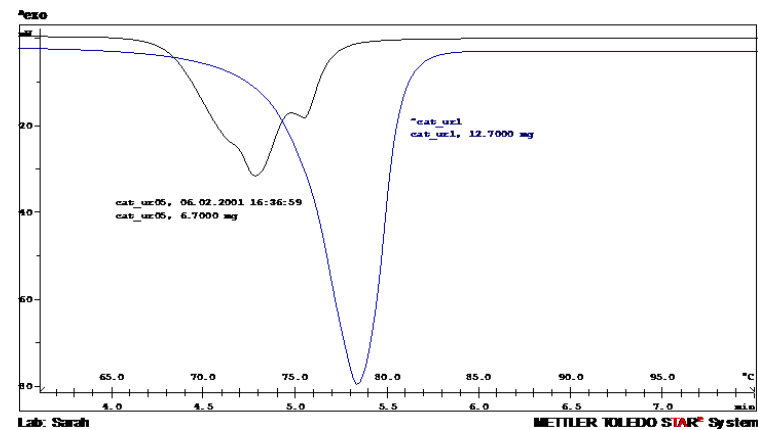
Figure 4.45B; sample of ground catechol urea after 5 minutes.
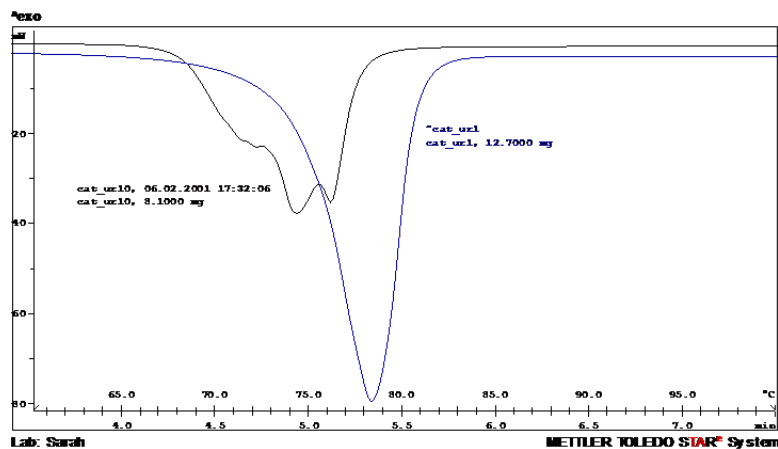
Figure 4.45C; sample of ground catechol urea after 10 minutes.
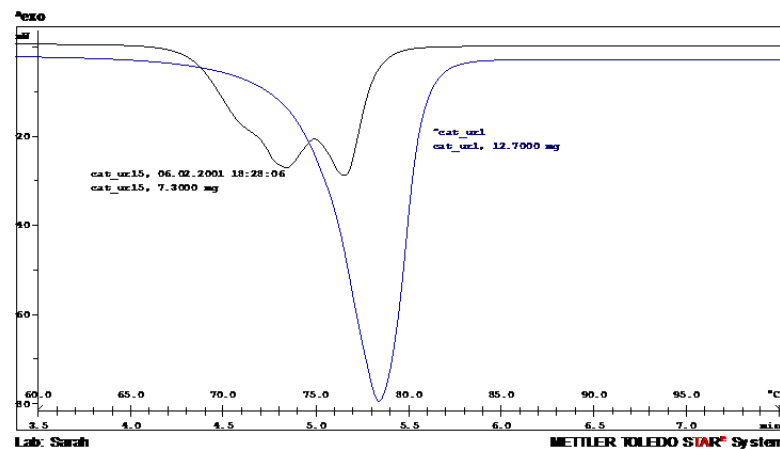
Figure 4.45D; sample of ground catechol urea after 15 minutes.

163
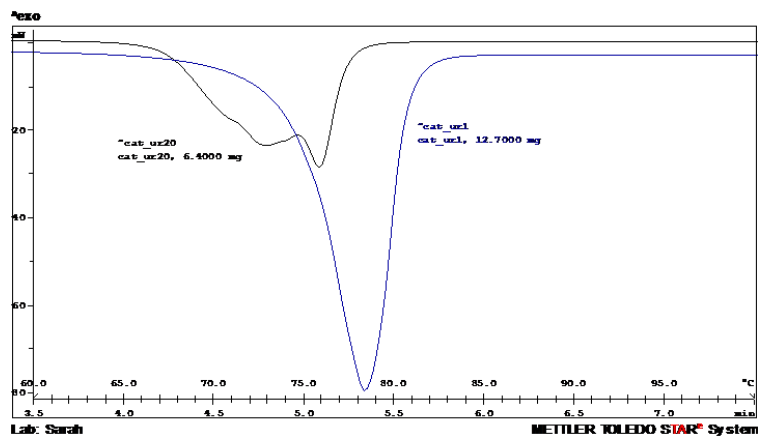
Figure 4.45E; sample of ground catechol urea after 20 minutes.
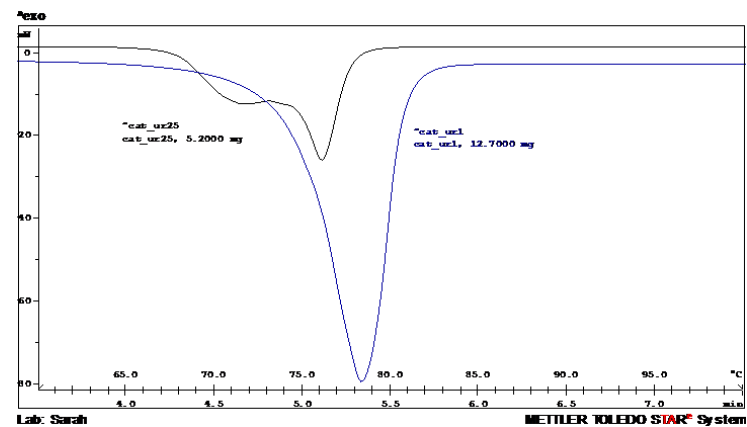
Figure 4.45F; sample of ground catechol urea after 25 minutes.
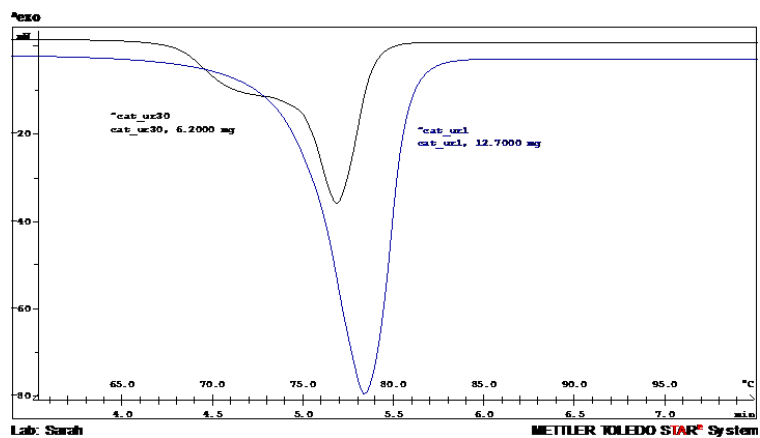
Figure 4.45G; sample of ground catechol urea after 30 minutes.
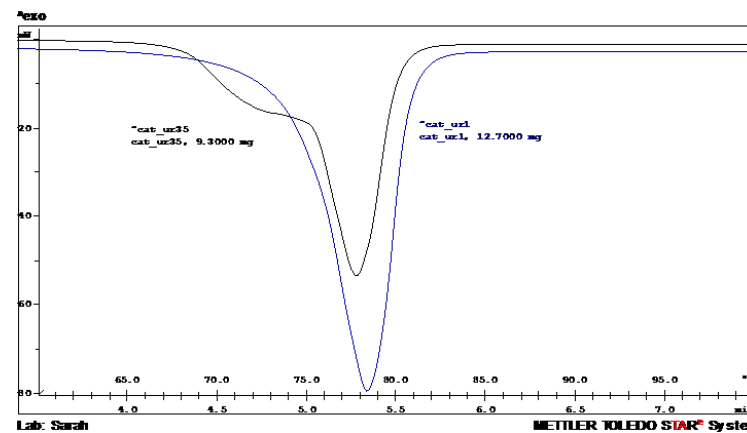
Figure 4.45H; sample of ground catechol urea after 35 minutes.

164

4.10.2 Indexing Results

The x-ray powder diffraction pattern was calculated in Cerius2 from the crystal structure of resorcinol urea obtained from the crystal structure database. The results are compared in table 3.

Resorcinol Urea			
Parameter	Published Data	*DICVOL*	*TREOR*
a (Å)	7.14	15.39	15.39
b (Å)	7.80	7.79	7.79
c (Å)	15.25	7.14	7.14
α (°)	90.00	90.00	90.00
ß (°)	90.00	90.00	90.00
γ (°)	90.00	90.00	90.00
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Figure of Merit	N/a	50.90	47.00
V ($Å^3$)	849.21	855.46	855.64

Table 4.3; Indexed data for simulated powder diffraction pattern of resorcinol:urea.
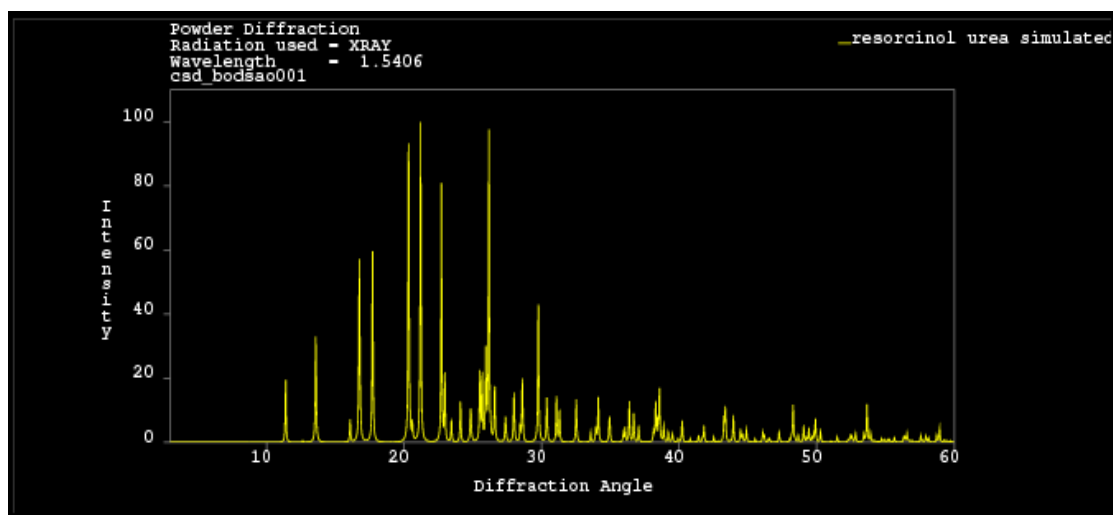
Figure 4.46; simulated x-ray powder diffraction pattern for resorcinol urea.

The powder diffraction data collected from both the solvent and ground forms of catechol urea were indexed and the results shown in table 4. Compared to the data set calculated for resorcinol urea, catechol urea seems to have similar bond lengths and cell volume.

Catechol Urea		
Parameter	*DICVOL*	*TREOR*
a (Å)	19.21	19.22
b (Å)	6.26	6.25
c (Å)	7.18	7.18
α (°)	90.00	90.00
ß (°)	97.96	97.90
γ (°)	90.00	90.00
Space Group	$P2_1/a$	$P2_1/a$
Figure of Merit	16.40	33.00
V (Å³)	854.82	854.26

Table 4.4; unit cell parameters generated from experimental data for catechol urea.
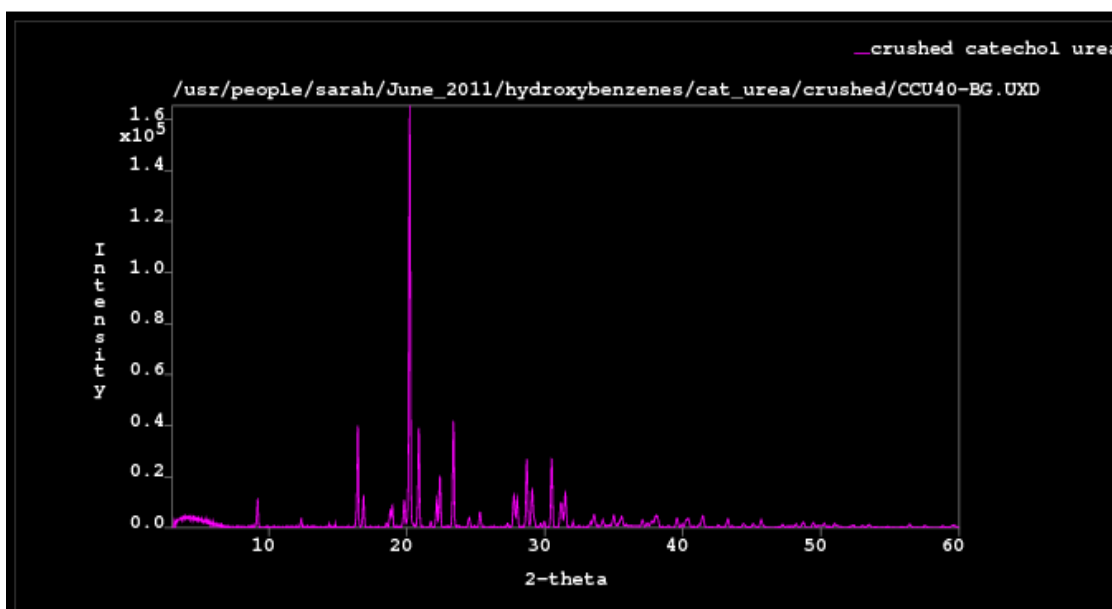
Figure 4.47A; X-ray powder diffraction pattern of a sample of catechol urea 1:1, ground together for 40minutes.
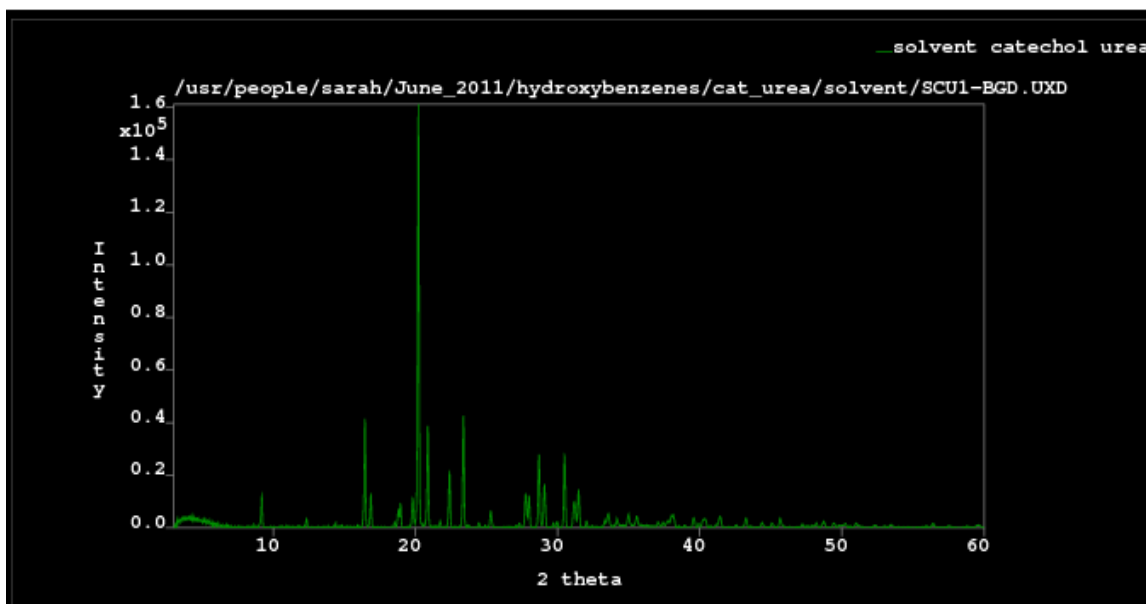
Figure 4.47B; X-ray powder diffraction pattern of a sample of catechol urea 1:1 crystallized using water as a solvent.

4.10.3   Trial Structure Generation

Cocrystal	Maximum step sizes			Cut-off Parameter
	Translations [fractional co-ordinates]	Rotations [º]	Torsions [º]	R_{wp}
Catechol/urea	0.5	3	0.25	0.97

Table 4.5; search parameters used in the simulated annealing searches for catechol/urea.

Top Trial Structure	R_{wp}
Catechol/urea (solvent)	0.995
Catechol/urea (ground)	0.990

Table 4.6; top structures generated from simulated annealing searches for catechol urea.

Examining the geometry of the molecular structures of dihydroxybenzenes revealed that the individual molecular geometry did not vary significantly from the cocrystal geometry. Based on this, the hydrogen atoms in the simulated structure of catechol urea were positioned in the same conformation as found in catechol.
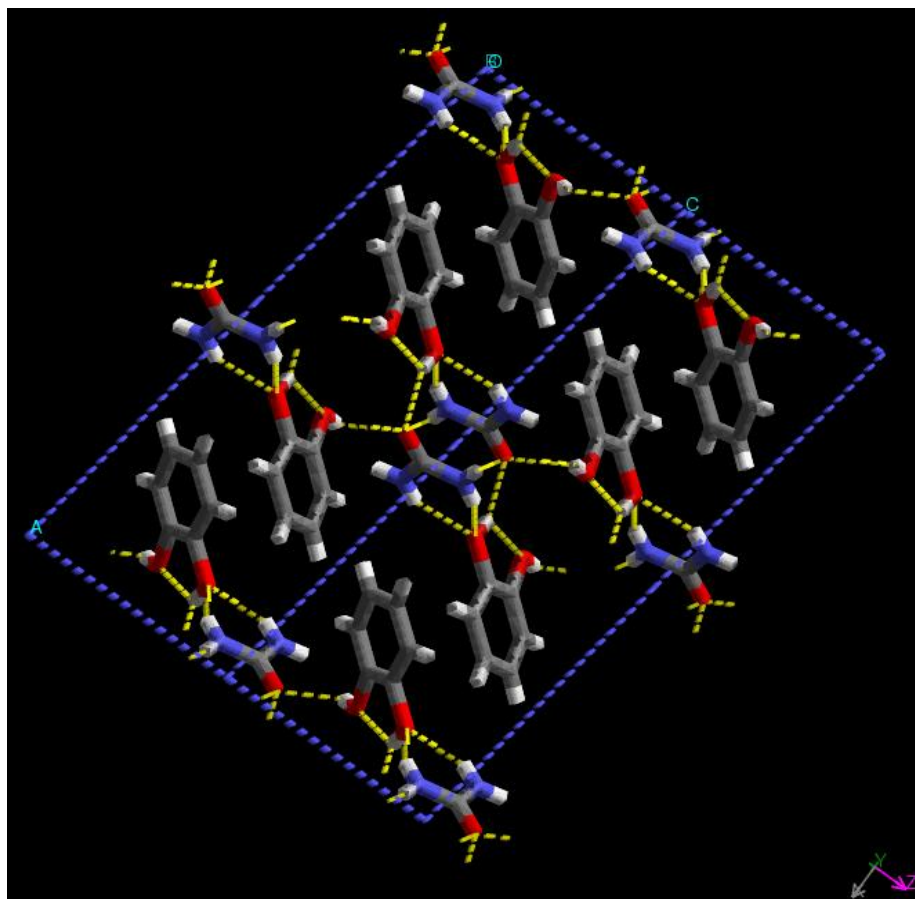
Figure 4.48A; the top simulated structure based on powder pattern fit for catechol urea from solvent crystallization.

Two catechol molecules connect via hydrogen bonds with two urea molecules in the basic building unit within the unit cell. $O_C \cdots H_U = 2.099$Å, $O\text{-}H_C \cdots O_U = 2.005$Å, $O\text{-}H_C \cdots O_C = 2.271$Å. This unit is further hydrogen bonded ($O\text{-}H_C' \cdots O_U = 2.010$Å, $O_U \cdots H_U'$ and $H_U \cdots O_U' = 2.081$Å) to similar units along the c-axis. In this structure only one hydrogen atom on each urea molecule cannot form a hydrogen bond.
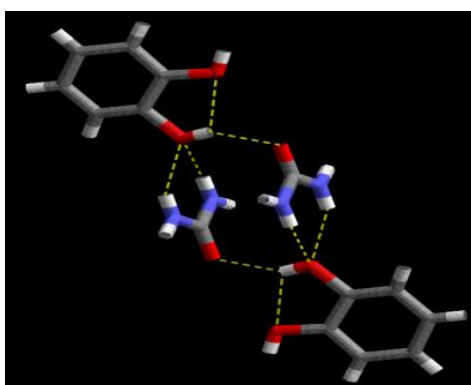
Figure 4.48B; basic hydrogen bonded unit of catechol urea from solvent crystallization.
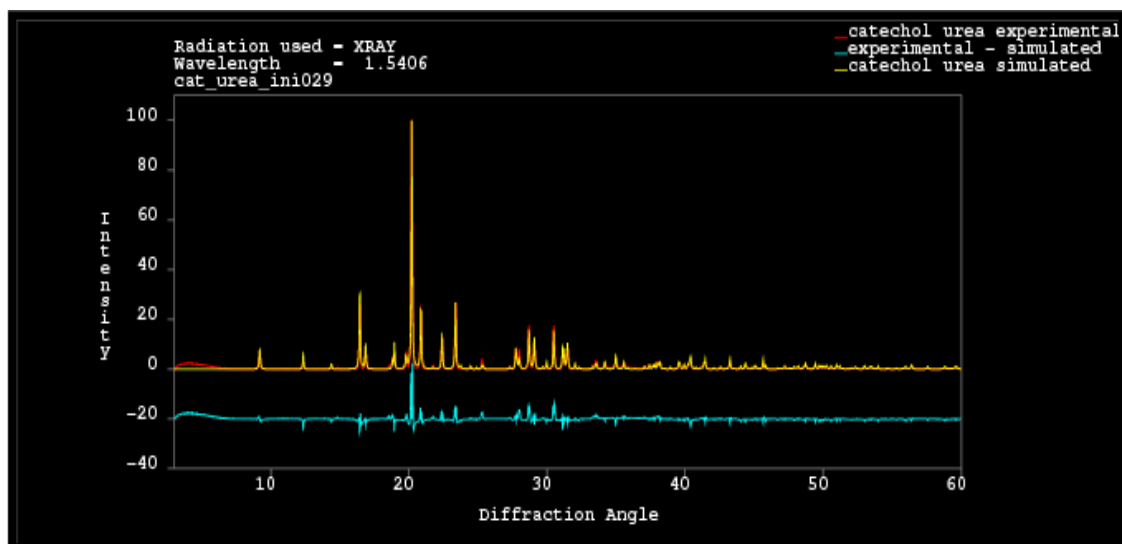
Figure 4.48C; comparison (blue) of experimental (red) and simulated (yellow) powder diffraction patterns for catechol urea (via solvent crystallization).
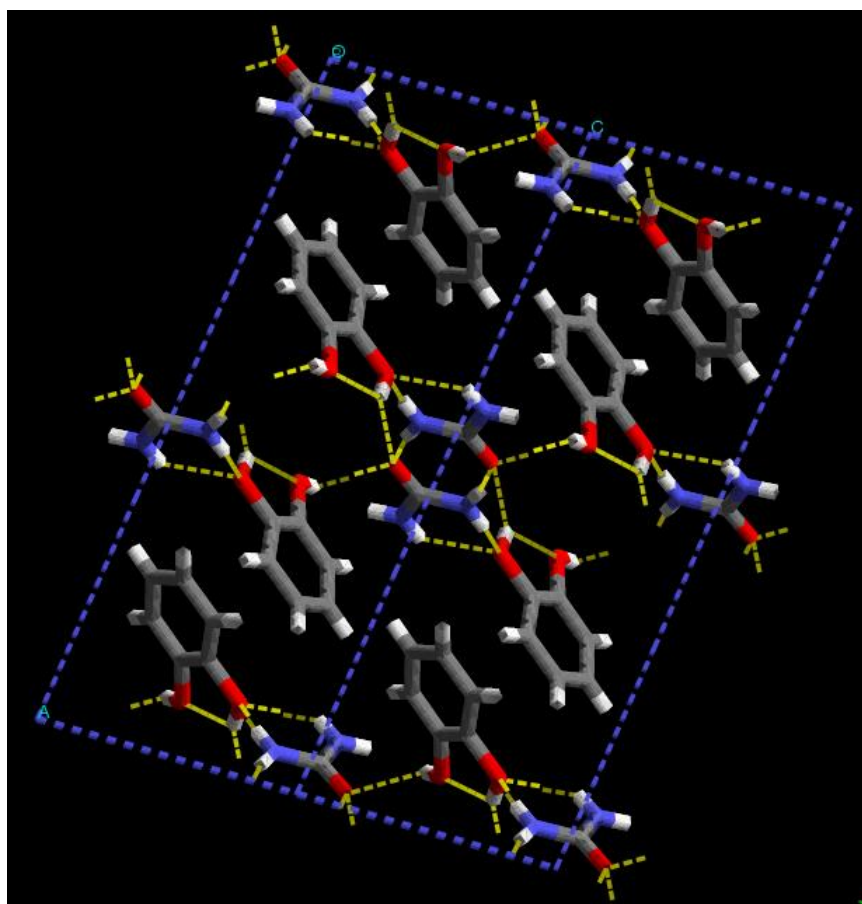
Figure 4.49A; the top simulated structure based on powder pattern fit for catechol urea from grinding.
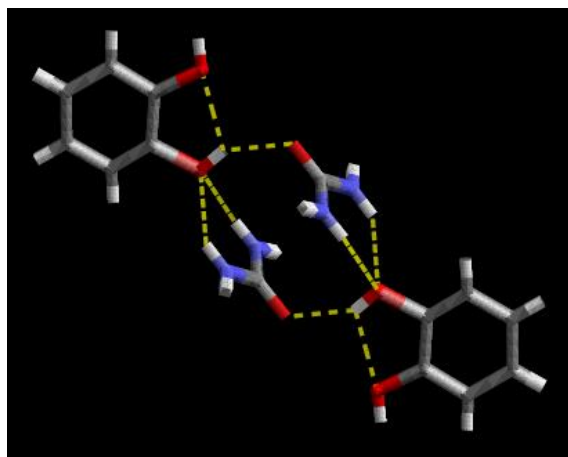
Figure 4.49B; Basic hydrogen bonded unit of catechol urea from grinding.

Two catechol molecules hydrogen bond with two urea molecules in the basic building unit within the unit cell. $O_C \cdots H_U = 2.205$ and 2.121Å, $O\text{-}H_C \cdots O_U = 1.904\text{Å}$, $O\text{-}H_C \cdots O_C = 2.271\text{Å}$. This unit is further hydrogen bonded ($O\text{-}H_C' \cdots O_U = 2.007\text{Å}$, $O_U \cdots H_U'$ and $H_U \cdots O_U' = 2.079\text{Å}$) to similar units along the c-axis. Again, there is only one hydrogen atom on the urea molecule unable to form a hydrogen bond in this structure.
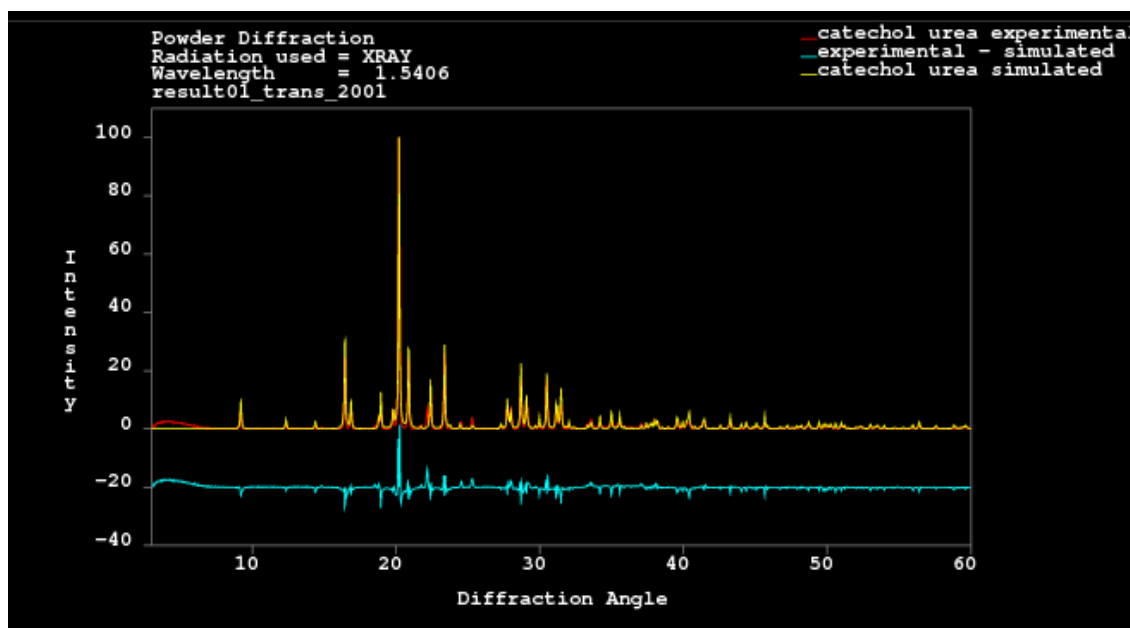
Figure 4.49C; comparison (blue) of experimental (red) and simulated (yellow) powder diffraction patterns for catechol urea ground sample.

There was a good correlation between the experimental and simulated x-ray powder diffraction patterns. The structure generated from the simulated annealing searches was put forward for GSAS refinement.
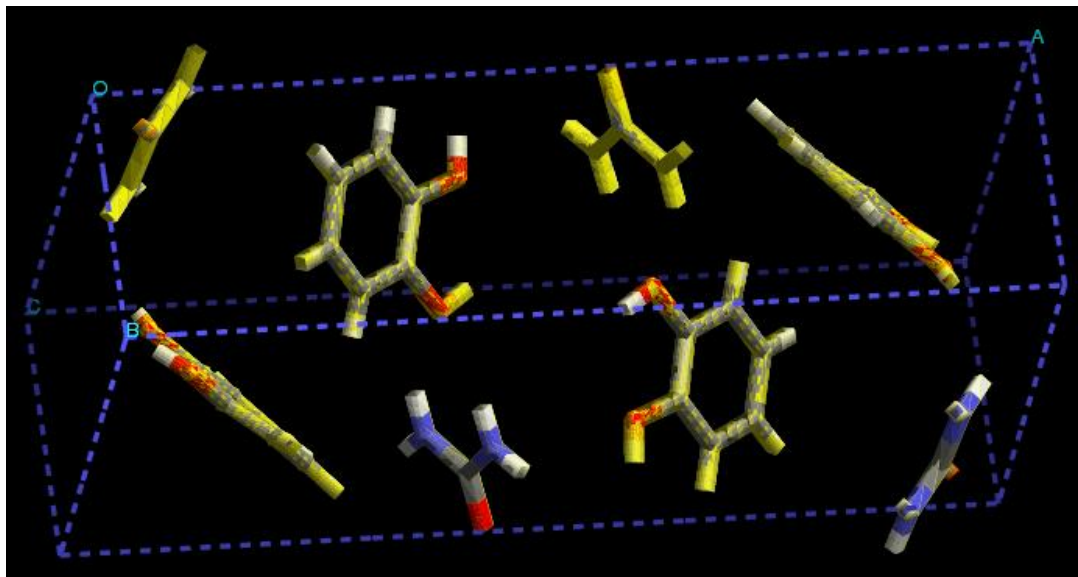
Figure 4.50; comparison of the crystal structures generated for catechol urea from experimental powder pattern data. Crushed sample (yellow), crystallized sample (default).

For comparison the MOPAC charges were calculated for crystallized catechol urea. The lattice energy was calculated using the Dreiding potential. The initial energy was calculated at -91kcal/mol (22.6kcal/mol/cell) and the resulting 'optimized' structure was -128kcal/mol (32kcal/mol/cell). However, when the structure determined from powder pattern data was compared to the optimized structure based on lattice energy, the differences were significant. The urea molecule is translated across the c axis and there is a rotation in the catechol position. This difference was also shown in the simulated powder diffraction data for the optimized structure when compared to the experimental data.
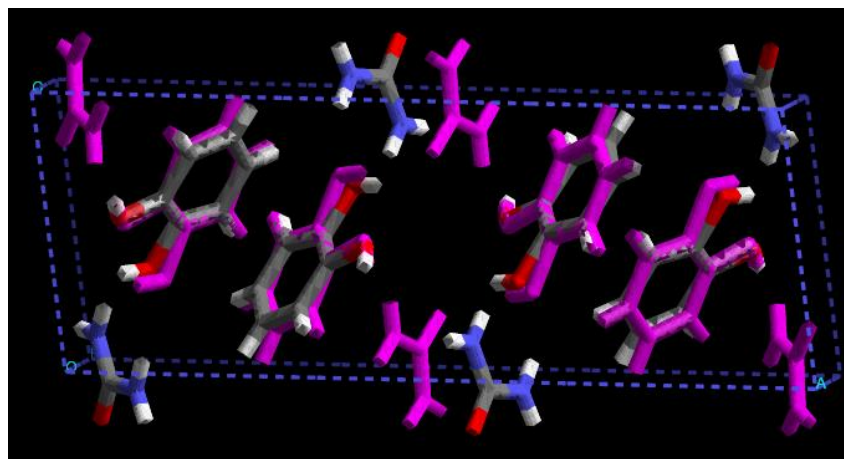
Figure 4.51A; Comparison of the structure generated for catechol urea based on powder pattern fit (default) and the optimized structure calculated based on lattice energy (magenta)
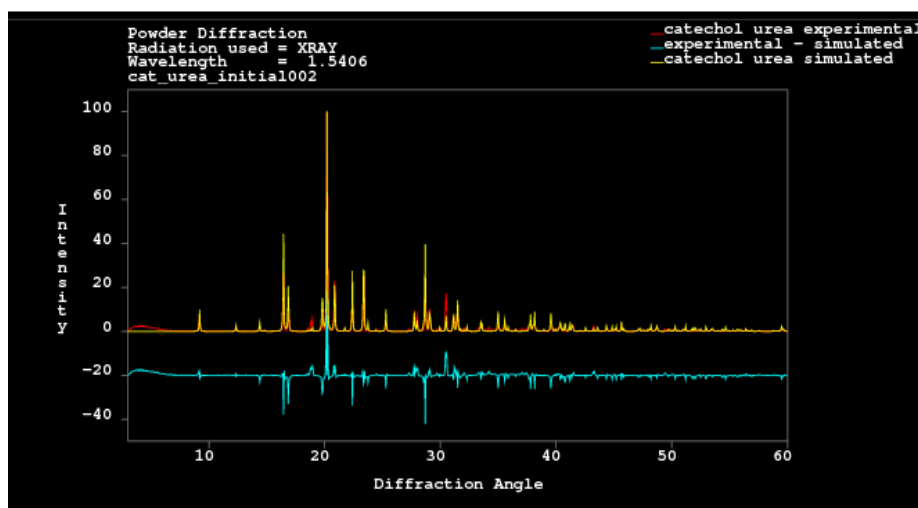
Figure 4.51B; Comparison of the x-ray powder data for catechol urea based on powder pattern fit (experimental) and the optimized structure calculated based on lattice energy (simulated).

Hydrogen positions for the generated cocrystal structure from powder diffraction data were kept the same as for the catechol crystal structure. However, a brief study was completed into how the structure would differ if the hydroxyl groups were facing in opposite directions instead. Limited information would be available from the powder diffraction patterns in this instance since only a hydrogen atom was being moved.
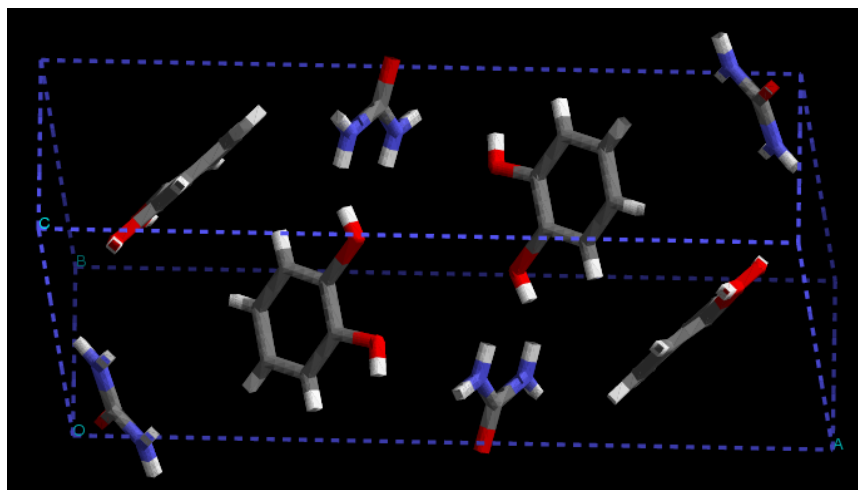
Figure 4.52; structure of catechol urea with a rotated hydroxyl group.

Two catechol molecules form hydrogen bonds with two urea molecules in the basic building unit within the unit cell. $O_C \cdots H_U$ = 2.214 and 2.118Å, $N_U \cdots H_C\text{-}O_C$ = 2.343Å and 2.435Å $O_U \cdots H_U$ = 2.086Å. This unit is further hydrogen bonded (O-$H_C' \cdots O_U$ = 2.002Å) to similar units along the c-axis. The interactions are generally over slightly longer distances in this structure and the interaction between O-$H_C \cdots O_U$ and O-$H_C \cdots O_C$ have disappeared. The urea molecule has one hydrogen atom unable to participate in any hydrogen bonding. The hydroxyl group on catechol involved in linking the chains of the basic unit together only utilizes the hydrogen atom, the oxygen atom remains unused. In the structure originally proposed, this oxygen atom is hydrogen bonded to the catechol hydrogen in the adjacent hydroxyl group. The calculated lattice energy of this structure is -53kcal/mol reducing to -97kcal/mol after optimization.

4.10.4 GSAS Refinements of Generated Trial Structures

Finally, the refined profile simulation resulted in an R_{wp}*(–back)* value of 4.74%. The profile difference plot is illustrated in figure 53, and the reduced χ^2 value is 5.327.
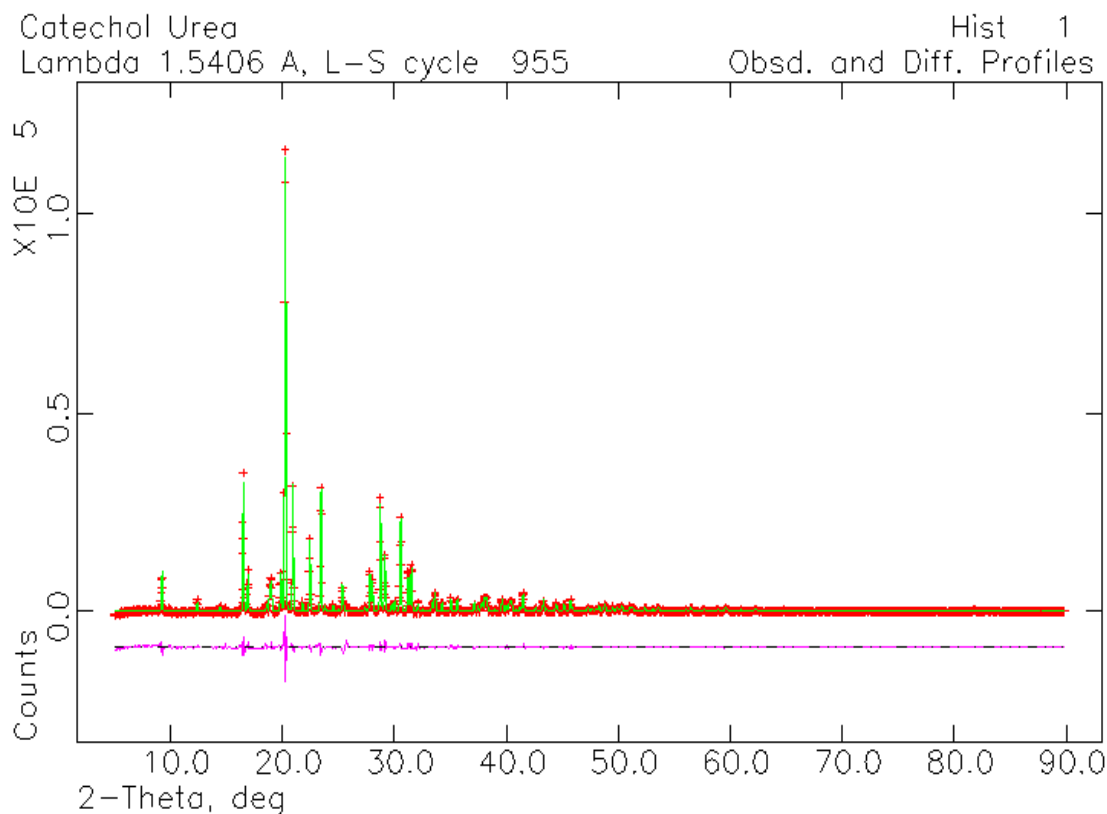
Figure 4.53; profile difference plot of catechol.urea.

Figure 4.54; screen results from GENLES.

4.11   Discussion of Results

The aim of this chapter was to determine the structure of the cocrystal catechol urea using x-ray powder diffraction coupled with molecular modelling. The steps involved were;

- Generation of material for x-ray powder diffraction.
- Collection of good quality powder diffraction data for indexing and comparison with simulated results.
- Generating a molecular model, selection of a force field and assignment of atomic charges for this model.
- Random searching of the unit cell determined from the indexing step.
- Optimization and refinement of results.
- Comparison of simulated data with experimental data.

Two methods were used to generate cocrystals and these were dry grinding and crystallization from solution. These methods are well established and are both widely used in industry [4.7].

There are advantages and disadvantages to both methods [4.7, 4.37]. Cocrystallization in solvents may become problematic if the solutes differ significantly in solubility. However, since urea and catechol are both readily soluble in water, this factor did not seem to be a concern in this instance. For future experiments, it may be of interest to generate cocrystals from a variety of solvents at different temperatures. Fast-cooling and slow-cooling crystallization from hot saturated solutions might reveal different polymorphs of this cocrystal. All solvent experiments in this study were carried out at just above room temperature, gently warming the solution to promote dissolution, without any attempt to regulate temperature stability. Sample grinding was carried out without any addition of solvent and it has been noted that there are many examples where liquid assisted grinding methods have shown more success than dry grinding [4.59, 4.61 and 4.66].

The series of spectra in figure 45 show a transition of the mixture from sampling after short grinding times. Gradually, this transition develops over a period of 35 minutes so that the results from DSC experiments eventually compare better with the crystals grown from evaporation. Grinding over more time and sampling the mixture for x-ray powder diffraction experiments could also be of interest in any further work. There is a risk in mechanochemical crystallization that friction can generate enough heat to allow form

transformations but the sample taken after 35 minutes seemed to be in good comparison with the cocrystal from solvent evaporation.

The generation of the unit cell parameters from indexing multiple samples showed consistent results for both crystallization methods.

The molecular model for catechol urea was developed using the individual crystal structures stored in the CSD for the backbone of catechol and the urea molecule. The reasoning behind this was that there are two sets of forces under consideration during crystal structure determination; the intramolecular forces and the intermolecular forces. The intramolecular forces are usually on a much greater scale than the weaker intermolecular forces. Covalent bond lengths and angles will rarely move significantly between polymorphic forms as too much energy is required than can be gained by a change of intermolecular bonding. In this study the flexible search would adjust any torsion angles and so the starting geometry for the OH groups was initially set to be in the same plane as the phenyl ring. For larger, more complex molecules where there are more degrees of freedom than in this study, it is common to optimize the covalent bonds and keep the flexible groups in a 'neutral' position as a starting point.

The MOPAC charge set and Dreiding force field [4.96] were used in energy calculations as these have been used throughout the previous validation of the simulated annealing work. Also, Dreiding had shown to be parameterized well for hydrogen bonded systems. For smaller molecules, these methods work well however, currently there are more sophisticated quantum mechanical methods available to evaluate the three-dimensional molecular structure. DFT has also reduced the time taken to carry out such evaluations, pushing the boundaries of the size of molecules for which quantum mechanics can be used [4.97 – 4.102].

During crystal structure searches, the computer program is moving the molecules around the unit cell and evaluating whether the structure could be plausible based on predefined selection criteria. In this study the unit cell parameters were maintained throughout all searches but these are also parameters which may be altered throughout certain searches. The degrees of freedom added to the searches in this study were translations along the three axes, rotations around three directions and rotations of the hydrogens around the O-H bonds. Only the space group defined by indexing was used in the searches although it

is agreed that this is a restrictive approach. Often multiple searches are carried out using the same molecular structure but for the most commonly encountered space groups. This is information which is gathered from the CSD. The drawbacks to this approach too is that less common space groups will be missed and there are now methods which are capable of searching all 230 space groups for one molecular starting structure [4.100]. A further approach is to search only in P1 and apply any symmetry considerations post search. These two methods still require multiple searches if the number of molecules in the asymmetric unit is unknown however.

There are drawbacks to using random search methods. While they are capable of covering large areas there is the danger that the same area is searched multiple times while other areas are overlooked. There is no way of knowing when a complete search of the unit cell has been achieved and so multiple searches are usually employed to increase this likelihood. Multiple random searches were still found to be faster than a fine grid search in this study however with the advent of parallel computing and more powerful processors, this may well have changed. The mid ground between these two searching methods is the 'quasi-random' or 'sobol' searches [4.103]. These give a better distribution than a truly random search but are still more random than the grid search.

The most common selection criterion for crystal structure generation is usually the calculated lattice energy [4.97 – 4.102]. In this study the lattice energy was calculated assuming that the system was static and all atoms were spherical. This is a very simplistic approach and over time, more elaborate and elegant methods of modelling these interactions have been developed and are now routine [4.104]. For each system under investigation the ease of implementation and use of a parameter set should be optimized along with the computational expense in order to balance the needs of accuracy and simplicity. The success of a force field usually lies with choosing the best empirical data (parameter set) to represent the molecule under investigation. The drawback to using empirical data sets is that they have usually been collected at room temperature. Crystal structure searches are assumed to be at absolute zero and so inaccuracies due to poor data parameterization will inevitably occur. Newer methods use quantum mechanical approaches to create model potentials. The advantage to this approach is that a model can be derived for each specific system. Therefore, many more molecular orientations may also be considered accurately as the model does not rely only on known empirical

data. It can also be added that if the stationary model is described accurately, there is also a better starting point from which to investigate lattice dynamics [4.105].

Atomic point charges were used throughout this study and these create a spherical cloud of charge around each atom. This is a useful tool to assign separations of charges but it is known that such clear cut examples are not the norm. An alternative approach has been to assign off-nuclear charges usually at the center of a bond or at likely lone pair positions. A more advanced technique is to calculate electrostatic multipoles which are centered at each atomic site and are derived from the calculated molecular wavefunction. Methods of modelling anistropic interatomic forces have shown great improvements in energy calculations for small organic molecules. It would be an interesting piece of work to use such methods used for this study. While multipoles don't lend themselves to large flexible molecules as they are dependent on molecular conformation, the example of catechol could be simple enough to attempt. A more recent development has been to use periodic DFT to evaluate the lattice energy [4.104]. With this method, the electron density and nuclear positions are optimized together. The benefits of this is that the molecular conformation adjusts naturally to the crystalline environment which means the electron distribution does not need to be fixed. This is particularly beneficial when assessing hydrogen bonded systems as the hydrogen bond motif in one example may be stabilized in preference over another example. There are still problems with periodic DFT however, and these are the time taken for the processes to complete as well as the inaccuracies in the long range contributions [4.105]. This has been examined further and by combining periodic DFT with an empirically derived term there was significant successes in the fourth blind tests of the CSD. This approach was used successfully for a large flexible molecule and for the example of a cocrystal. It would again be of interest to see how this method could improve or change the suggested structure for catechol urea.

Optimization and refinement stages in this study were briefly attempted using minimizer in Cerius^2 and a Rietveld refinement. The powder diffraction pattern for the resulting structure was compared with the experimental powder diffraction pattern and a promising match was seen. This comparison may well be improved upon by using the alternative methods mentioned above. Since this study, a very similar structure has been proposed using a systematic grid search method [4.95]. This approach also used point charges and a force field with empirical data. Following any future work with improved charge models, a comparison with this structure would also be of interest.

4.12   References

[4.1]   A.I. Kitaigorodsky, *The Principle of Close Packing and the Condition of Thermodynamic Stability of Organic Crystals,* Acta Cryst., 18(4), 585-590 (1965)

[4.2]   J. Bernstein, *Conformational Polymorphism. The Influence of Crystal Structure on Molecular Conformation,* J. Am. Chem. Soc., 100(3), 673-681 (1978)

[4.3]   A. T. Hagler, *Conformational Polymorphism. 2. Crystal Energetics by Computational Substitution. Further Evidence for the Sensitivity of the Method,* J. Am. Chem. Soc., **100**(20), 6349-6354 (1978)

[4.4]   G. R. Desiraju, *Polymorphism: The Same and Not Quite the Same,* Cryst. Growth and Des., **8**(1), 3-5 (2008)

[4.5]   J. F. Fernández-Bertran, *Mechanochemistry: An Overview,* Pure Appl. Chem., 71(4), 581-586 (1999)

[4.6]   T.E. Fischer, *Tribochemistry,* Ann. Rev. Mater. Sci., 18, 303-323 (1988)

[4.7]   (a) D. Braga, *Applications of Crystal Engineering Strategies in Solvent-Free Reactions: Towards a Supramolecular Green Chemistry,* Chapter 1 - Frontiers in Crystal Engineering, John Wiley & Sons, Edited by E.R.T. Tiekink and J.J. Vittal (2006); (b) K. Chadwick, *How does Grinding Produce Co-crystals? Insights from the Case of Benzophenone and Diphenylamine,* Cryst. Eng. Comm., 9, 732-734 (2007); (c) D. R. Weyna, *Synthesis and Structural Characterization of Cocrystals and Pharmaceutical Cocrystals: Mechanochemistry vs Slow Evaporation,* Cryst. Growth and Des., **9**(2), 1106-1123 (2009); (d) T. Friščić, *Recent Advances in Understanding the Mechanism of Cocrystal Formation via Grinding,* Cryst. Growth and Des., **9**(3), 1621-1637 (2009)

[4.8]   M.C. Etter, *Hydrogen Bonds as Design Elements in Organic Chemistry,* J. Phys. Chem., 95, 4601-4610 (1999)

[4.9]   (a) G. R. Desiraju, *Crystal Engineering: Structure, Design and Function,* Curr. Opin. Solid State Mat. Sci., **13**, 35 (2009); (b) J. Rantanen, *The Future of Pharmaceutical Manufacturing Sciences,* J. Pharm. Sci., **104**, 3612-3638 (2015); (c) S. Datta, *Crystal Structures of Drugs: Advances in Determination, Prediction and Engineering,* Nat. Rev. Drug Discovery, **3**, 42-57 (2004); (d) R. D. Bailey Walsh, *Crystal Engineering of the Composition of Pharmaceutical Phases,* Chem. Commun., 186-187 (2003); (e) G. R. Desiraju, *Designer Crystals: Intermolecular Interactions, Network Structures and Supramolecular Synthons,* Chem. Commun, 1475-1481 (1997); (f) J. Elliott, *Pharmaceutical Materials Science: An Active New Frontier in Materials Research,* MRS Bull., **31**, 869-873 (2006); (g) C. B. Aakeröy, *Recent Advances in Crystal Engineering,* Cryst. Eng. Comm., **12**, 22-43 (2010); (h) N. Feeder, *The Integration of Solid-Form*

179

Informatics into Solid-Form Selection, J. Pharm. Pharmacol., **67**, 857-868 (2015); (i) M. D. Ticehurst, *Integration of Active Pharmaceutical Ingredient Solid Form Selection and Particle Engineering into Drug Product Design,* J. Pharm. Pharmacol., **67**, 782-802 (2015); (j) C.B. Aakeröy, *Building Co-crystals with Molecular Sense and Supramolecular Sensibility,* Cryst. Eng. Comm., 7(72), 439-448 (2005); (k) J. D. Wuest, *Engineering Crystals by the Strategy of Molecular Tectonics,* Chem Commun., 5830-5837 (2005); (l) G. R Desiraju, *Crystal Engineering: A Holistic View,* Angew Chemie Int. Ed., **46**(44) 8342-8356 (2007)

[4.10] (a) M. M. Parmar, *Polymorph Selection with Morphology Control Using Solvents,* Cryst. Growth and Des., **7**(9), 1635-1642 (2007); (b) A. T. M. Serajuddin, *Salt Formation to Improve Drug Solubility,* Adv. Drug Del. Rev., **59**, 603-616 (2007)

[4.11] (a) M. Hartenfeller, *Enabling Future Drug Discovery by de novo Design,* WIREs Computational Molecular Science, **1**, 742-759 (2011); (b) G. Schneider, *De Novo Design: From Models to Molecules,* De Novo Molecular Design, Chapter 1, G. Schneider (Ed.), 1-55 (2014)

[4.12] (a) G. Schneider, *De Novo Design of Molecular Architectures by Evolutionary Assembly of Drug-Derived Building Blocks,* J. Comput. Aid. Mol. Des., **14**, 487-494 (2000); (b) K. Loving, *Computational Approaches for Fragment-Based and De Novo Design,* Curr. Top. Med. Chem., **10**, 14-32 (2010); (c) A. S. Reddy, *Structure-Based De Novo Drug Design,* De Novo Molecular Design, Chapter 4, G. Schneider (Ed.), 97-123 (2014); (d) V. J. Gillet, *De Novo Molecular Design,* Evolutionary Algorithms in Molecular Design, Chapter 4, D. E. Clark (Ed.), 49-69 (2000)

[4.13] K.D.M. Harris, *Molecular Confinement,* Chemistry in Britain, February, 132-136 (1993)

[4.14] D.E. Palin, *The Structure of Molecular Compounds. Part III. Crystal Structure Addition Complexes of Quinol with Certain Volatile Compounds.,* J. Am. Chem. Soc., 208-221 (1947)

[4.15] (a) G.R. Desiraju, *Crystal and Co-Crystal,* Cryst. Eng. Comm., 5(82), 466-467 (2003); (b) J. D. Dunitz, *Crystal and Co-Crystal: A Second Opinion,* Cryst. Eng. Comm., 5(91), 506 (2003); (c) A.D. Bond, *What is A Co-Crystal?* Cryst. Eng. Comm., 9, 833-834 (2007); (d) C.B. Aakeröy, *Co-crystal or Salt: Does it Really Matter?,* Molecular Pharmaceutics, 4(3), 317-322 (2007); (e) E. Grothe, *Solvates, Salts, and Cocrystals: A Proposal for a Feasible Classification System,* Cryst. Growth Des., **16**, 3237-3243 (2016)

[4.16] H. G. Brittain, *Pharmaceutical Cocrystals: The Coming Wave of New Drug Substances,* J. Pharm. Sci., **102**(2), 311-317 (2013)

[4.17] J. W. Steed, *The Role of Co-crystals in Pharmaceutical Design,* Trends in Pharmacological Sciences, **34**(3), 185-193 (2013)

[4.18] (a) Ö, Almarsson, *Crystal Engineering of the Composition of Pharmaceutical Phases. Do Pharmaceutical Co-crystals Represent a New Path to Improved Medicines?,* Chem. Commun., 1889-1896 (2004); (b) N. K. Duggirala, *Pharmaceutical Cocrystals: Along the Path to Improved Medicines,* Chem. Commun., **52**, 640-655 (2016)

[4.19] R. Thakuria, *Pharmaceutical Cocrystals and Poorly Soluble Drugs,* Int. J. Pharm., **453**, 101-125 (2013)

[4.20] (a) G. Kuminek, *Cocrystals to Facilitate Delivery of Poorly Soluble Compouds Beyond-rule-of-5,* Adv. Drug Del. Rev., **101**, 143-166 (2016); (b) N. Schultheiss, *Pharmaceutical Cocrystals and Their Physicochemical Properties,* Cryst. Growth Des., **9**(6), 2950-2967 (2009); (c) A. V. Yadav, *Co-crystals: A Novel Approach to Modify Physicochemical Properties of Active Pharmaceutical Ingredients,* Ind. J. Pharm. Sci., **July-August**, 359-370 (2009)

[4.21] (a) B. Sarma, *Hydrogen Bond Synthon Competition in the Stabilization of Theophylline Cocrystals*, Cryst. Eng. Comm., **16**, 4753-4765 (2016); (b) T. Patole, *Co-crystallization – a Technique for Solubility Enhancement,* Int. J. Pharm. Sci. Res., **5**(9), 3566-3576 (2014); (c) S. G. Fleischman, *Crystal Engineering of the Composition of Pharmaceutical Phases: Multiple-Component Crystalline Solids Involving Carbamazepine,* Cryst. Growth Des., **3**(6), 909-919 (2003); (d) M. B. Hickey, *Performance Comparison of a Co-crystal of Carbamazepine with Marketed Product,* Eur. J. Pharm. Biopharm., **67**, 112-119 (2007)

[4.22] (a) P. A. Wood, *Knowledge-Based Approaches to Co-crystal Design,* Cryst. Eng. Comm., **16**, 5839-5848 (2014); (b) D. H. Dressler, *Controlling Polymorphism by Crystallization on Self-Assembled Multilayers,* Cryst. Growth Des., **7**(5), 847-850 (2007)

[4.23] U.S. Department of Health and Human Services Food and Drug Administration, http://www.accessdata.fda.gov/scripts/fdcc/?set=SCOGS

[4.24] M. R. Shimpi, *New Cocrystals of Ezetimibe with L-Proline and Imidazole,* Cryst. Eng. Comm., **16**, 8984-8993 (2014)

[4.25] A. Nangia, *Pharmaceutical Cocrystals: Walking the Talk,* Chem. Commun., **52**, 8342-8360 (2016)

[4.26] (a) X. Shi, *Concomittant Cocrystallization on Engineered Surfaces,* Cryst. Eng. Comm., **15**, 7450-7455 (2013); (b) S. Domingos, *New Forms of Old Drugs: Improving Without Changing,* J. Pharm. Pharma., **67**, 830-846 (2015); (c) N. R. Goud, *Modulating the Solubility of Sulfacetamide by means of Cocrystals,* Cryst. Eng. Comm., **16**, 5859-

5869 (2014); (d) A. V. Trask, *Pharmaceutical Cocrystallization: Engineering a Remedy for Caffeine Hydration,* Cryst. Growth Des., **5**(3), 1013-1021 (2005); (e) C. C. Sun, *Improving Mechanical Properties of Caffeine and Methyl Gallate Crystals by Cocrystallization,* Cryst. Growth Des., **8**(5), 1575-1579 (2008); (f) P. M. Bhatt, *Cocrystals of the Anti-HIV Drugs Lamivudine and Zidovudine,* Cryst. Growth Des., **9**(2), 951-957 (2009)

[4.27] S. Aitipamula, *Polymorphism in Cocrystals: A Review and Assessment of its Significance,* Cryst. Eng. Comm., **16**, 3451-3465 (2014)

[4.28] G. R. Desiraju, *Pharmaceutical Salts and Co-crystals: Retrospect and Prospects*, Chapter 1, RSC Drug Discovery Series No 16, Pharmaceutical Salts and Co-crystals, J. Wouters and L. Quėrė (Ed.), 1-8 (2012)

[4.29] P. Vishweshwar, *Crystal Engineering of Pharmaceutical Co-crystals from Polymorphic Active Pharmaceutical Ingredients,* Chem. Commun., 4601-4603 (2005)

[4.30] B. R. Seekanth, *Supramolecular Synthon Polymorphism in 2:1 Co-crystal of 4-Hydroxybenzoic Acid and 2,3,5,6-Tetramethylpyrazine,* Chem. Commun., 2375-2377 (2007); (b) A. Mukherjee, *Synthon Identification* in Co-Crystals and Polymorphs with IR Spectroscopy. Primary Amides as a Case Study, Cryst. Eng. Comm., **15,** 4640-4654 (2013)

[4.31] P. M. Bhatt, *Tautomeric Polymorphism in Omeprazole,* Chem. Commun., 2057-2059 (2007)

[4.32] S. Aitipamula, *Conformational and Enantiotropic Polymorphsim of a 1:1 Cocrystal Involving Ethenzamide and Ethylmalonic Acid,* Cryst. Eng. Comm., **12**, 3691-3697 (2010)

[4.33] C.B. Aakeröy, *Do Polymorphic Compounds Make Good Cocrystallizing Agents? A Structural Case Study that Demonstrates the Importance of Synthon Flexibility,* Crystal Growth & Design, 3(2), 159-165 (2003)

[4.34] A. Lemmerer, *Polymorphic Co-crystals from Polymorphic Co-crystal Formers: Competition between Carboxylic Acid⋯Pyridine and Phenol⋯Pyridine Hydrogen Bonds,* Cryst. Growth and Des., **13** 3935-3952 (2013)

[4.35] S. Qiu, *Role of Polymers in solution and Tablet-Based Carbamazepine Cocrystal Formulations,* Cryst. Eng. Comm., **18**, 2664-2678 (2016)

[4.36] P. T. A. Galek, *Hydrogen-bond Coordination in organic crystal structures: Statistics, Predictions and Applications,* Acta Cryst., **B70**, 91-105 (2014)

[4.37] K. Fucke, *How Good are the Crystallization Methods for Cocrystals? A Comparative Study of Piroxicam,* New J. Chem., **36**, 1969-1977 (2012)

[4.38] S. R. Bysouth, *Cocrystallization via Planetary Milling: Enhancing Throughput of Solid-State Screening Methods,* Int. J. Pharm., **411**, 169-171, (2011)

[4.39] C. B. Aakeröy, *Ten Years of Co-crystal Synthesis; The Good, The Bad, and the Ugly,* Cryst. Eng. Comm., **10**, 1816-1821 (2008)

[4.40] P. T. A. Galek, *Knowledge-based H-bond Prediction to Aid Experimental Polymorph Screening,* Cryst. Eng. Comm., **11**, 2634-2639 (2009)

[4.41] (a) M. C. Etter, *Graph-Set Analysis of Hydrogen Bond Patterns in Organic Crystals,* Acta Cryst., **B46**, 256-262 (1990); (b) M. C. Etter, *Encoding and Decoding Hydrogen-Bond Patterns of Organic Compounds,* Acc. Chem. Res., **23**, 120-126 (1990); (c) M. C. Etter, *Hydrogen Bond Directed Cocrystallization and Molecular Recognition Properties of Acyclic Imides,* J. Am. Chem. Soc., **113**, 2586-2598 (1991); (d) J. H. Loehlin, *Hydrogen-Bond Patterns in Several 2:1 Amine-Phenol Cocrystals,* Chem. Mater., **6**, 1218-1221 (1994); (e) M. C. Etter, *Hydrogen-Bond Directed Cocrystallization as a Tool for Designing Acentric Organic Solids,* Chem. Mater., **1**, 10-12 (1989)

[4.42] L. Zhou, *Co-crystal Formation Based on Structural Matching,* European Journal of Pharmaceutical Sciences, **88**, 191-201 (2016)

[4.43] (a) S. R. Fukte, *Coformer Selection: An Important Tool in Cocrystal Formation,* Int. J. Pharmacy and Pharm Sci., **6**(7), 9-14 (2014); (b) G. P. Stahly, *Diversity in Single- and Multiple-Component Crystals. The Search for and Prevalence of Polymorphs and Cocrystals, Cryst.* Growth and Des., **7**(6), 1007-1026 (2007)

[4.44] G. He, *Screening for Cocrystallization Tendency: The Role of Intermolecular Interactions,* J. Phys. Chem., **B112**, 9890-9895 (2008)

[4.45] C. Loschen, *Solubility Prediction, Solvate and Cocrystal Screening as Tools for Rational Crystal Engineering,* J. Pharm. Pharmacol., **67**, 803-811 (2015)

[4.46] (a) N. Issa, *Can the Formation of Pharmaceutical Cocrystals be Computationally Predicted? 1. Comparison of Lattice Energies,* Cryst. Growth and Des., **9**(1), 442-453 (2009); (b) P. G. Karamertzanis, *Can the Formation of Pharmaceutical Cocrystals be Computationally Predicted? 2. Crystal Structure Prediction,* J. Chem. Theory Comput., **5**, 1432-1448 (2009)

[4.47] C. F. Macrae, *Mercury CSD 2.0 – New Features for the Visualization and Investigation of Crystal Structures,* J. Appl. Cryst., **41**, 466-470 (2008)

[4.48] P. T. A. Galek, *Knowledge-based Model of Hydrogen-bonding Propensity in Organic Crystals,* Acta Cryst., **B63**, 768-782 (2007)

[4.49] P. Galek, *CSD Solid Form Suite: Addressing Key Issues in Solid State Development,* White Paper www.ccdc.cam.ac.uk

[4.50] L. Fábián, Cryst. Growth Des., **9**, 1436-1443 (2009)

[4.51] M. A. Mohammad, *Hansen Solubility Parameter as a Tool to Predict Cocrystal Formation,* Int. J. Pharm, 407, 63-71 (2011)

[4.52] (a) C. V. S. Subrahmanyam, *Estimation of the Solubility Parameter of Trimethorprim by Current Methods,* Pharm. Acta Helv., **71**, 175-183 (1996); (b) B. C. Hancock, *The Use of Solubility Parameters in Pharmaceutical Dosage Form Design,* Int. J. Pharm., **148**, 1-21 (1997); (c) A. Güner, *The Algorithmic Calculations of Solubility Parameter for the Determination of Interactions in Dextran/Certain Polar Solvent Systems,* Eur. Polymer J., **40**, 1587-1594 (2004); (d) D. J. Greenhalgh, *Solubility Parameters as Predictors of Miscibility in Solid Dispersions,* J. Pharm. Sci., **88**(11), 1182-1190 (1999); (e) J. Breitkreutz, *Prediction of Intestinal Drug Absorption Properties by Three-Dimensional Solubility Parameters,* Pharm. Res., **15**(9), 1370-1374 (1998)

[4.53] C. B. Aakeröy, *Exploring the Hydrogen-Bond Preference of N-H Moieties in Co-crystals Assembled via O-H(acid)...N(py) Intermolecular Interactions,* Cryst. Eng. Comm., **9**, 46-54 (2007)

[4.54] J. H. Loehlin, *Supramolecular Alcohol-Amine Crystals and Their Hydrogen-Bond Patterns,* Acta Cryst., **B54**, 695-704 (1998)

[4.55] J. Wang, *Three Concomitant Polymorphs of 1:1 4,4'dihydroxybenzophenone/1,2-bis(4-pyridyl)-ethylene: Applications of Hydrothermal Method in Searching Polymorphs,* Cryst. Eng. Comm, **9**, 591-594 (2007)

[4.56] M. D. Eddleston, *Screening for Polymorphs of Cocrystals: A Case Study,* Cryst. Eng. Comm., **15**, 175-181 (2013)

[4.57] R. S. Dhumal, *Particle Engineering Using Sonocrystallization: Salbutamol Sulphate for Pulmonary Delivery,* Int. J. Pharm, **368**, 129-137 (2009)

[4.58] L. Padrela, *Formation of Indomethacin-Saccharin Cocrystals Using Supercritical Fluid Technology,* Eur. J. Pharm. Sci., **38**, 9-17 (2009)

[4.59] S. Karki, *Control and Interconversion of Cocrystal Stoichiometry in Grinding: Stepwise Mechanism for the Formation of a Hydrogen-Bonded Cocrystal,* Cryst. Eng. Comm., **11**, 470-481 (2009)

[4.60] N. Issa, *Screening for Cocrystals of Succinic Acid and 4-Aminobenzoic Acid,* Cryst. Eng. Comm., **14**, 2454-2464 (2012)

[4.61] A. V. Trask, *Solvent-Drop Grinding: Green Polymorph Control of Cocrystallization,* Chem. Commun., 890-891 (2004)

[4.62] S. A. Ross, *Engineering and Manufacturing of Pharmaceutical Co-crystals: A Review of Solvent-Free Manufacturing Technologies,* Chem. Commun., **52**, 8772-8786 (2016)

[4.63] K. Chadwick, *How Does Grinding Produce Co-crystal? Insights from the Case of Benzophenone and Diphenylamine,* Cryst. Eng. Comm., **9**, 732-734 (2007)

[4.64] A. Y. Sheikh, *Scalable Solution Cocrystallization: Case of Carbamazepine-Nicotinamide I,* Cryst. Eng. Comm., **11**, 501-509 (2009)

[4.65] I. Wang, *Anti-Solvent Co-crystallization of Carbamazepine and Saccharin,* Int. J. Pharm, **450**, 311-322 (2013)

[4.66] N. Shan, *Mechanochemistry and Co-crystal Formation: Effect of Solvent on Reaction Kinetics,* Chem. Commun., 2372-2373 (2002)

[4.67] W. W. Porter III, *Polymorphism in Carbamazepine Cocrystals,* Cryst. Growth Des., **8**(1), 14-16 (2008)

[4.68] V. R. Pedireddi, *Creation of Crystalline Supramolecular Arrays: A Comparison of Co-crystal Formation from Solution and by Solid-State Grinding,* Chem. Commun., 987-988 (1996)

[4.69] D. J. Berry, *Applying Hot-Stage Microscopy to Co-crystal Screening: A Study of Nicotinamide with Seven Active Pharmaceutical Ingredients,* Cryst. Growth and Des., **8**(5), 1697-1712 (2008)

[4.70] A. Alhalaweh, *Formation of Cocrystals from Stoichiometric Solutions of Incongruently Saturating Systems from Spray Drying,* Cryst. Growth and Des., **10**(8), 3302-3305 (2010)

[4.71] T. Rager, *Cocrystal Formation from Solvent Mixtures,* Cryst. Growth and Des., **10**(7), 3237-3241 (2010)

[4.72] N. Takata, *Cocrystal Screening of Stanolone and Mestanolone,* Cryst. Growth and Des., **8**(8), 3032-3037 (2008)

[4.73] T. Friščić, *The Role of Solvent in Mechanochemical and Sonochemical Cocrystal Formation: A Solubility-Based Approach for Predicting Cocrystallization Outcome,* Cryst. Eng. Comm., **11**, 418-426 (2009)

[4.74] S. Skovsgaard, *Co-crystallization of Benzoic Acid Derivatives with N-containing Bases in Solution and by Mechanical Grinding: Stoichiometric Variants, Polymorphism and Twinning,* Cryst. Eng. Comm., **11**, 443-453 (2009)

[4.75] N. Blagden, *Crystal Chemistry and Solvent Effects in Polymorphic Systems Sulfathiazole,* J. Chem. Soc., Faraday Trans., **94**(8), 1035-1044 (1998)

[4.76] N. Blagden, *Pharmaceutical Co-crystals – Are We There Yet?,* Cryst. Eng. Comm., **16**, 5753 – 5761 (2014)

[4.77] (a) S. Aitipamula, *Polymorphs, Salts, and Cocrystals: What's in a Name?*, Cryst. Growth Des., **12**, 2147-2152 (2012)

[4.78] U.S. Department of Health and Human Services*, Regulatory Classification of Pharmaceutical Co-crystals, Guidance for Industry*, April 2013.

[4.79] (a) C. A. Challener, *Scientific Advances in Cocrystals are Offset by Regulatory Uncertainty,* Pharm. Tech., 42- 45 May 2014; (b) S. L Childs, *The Salt-Cocrystal Continuum: The Influences of Crystal Structure on Ionization State,* Mol. Pharm., **4**(3), 323-338 (2007)

[4.80] European Medicines Agency, *Reflection Paper on the use of Cocrystals and Other Solid State Forms of Active Substances in Medicinal Products, EMA/CHMP/CVMP/QWP/136250/2014*

[4.81] (a) P. Vishweshwar, *Pharmaceutical Co-crystals,* J. Pharm. Sci., **95**(3), 499-156 (2006); (b) G. P. Stahly, *A Survey of Cocrystals Reported Prior to 2000,* Cryst. Growth Des., **9**, 4212-4229 (2009); (c) S. L. Childs, *The Reemergence of Cocrystals: The Crystal Clear Writing is on the Wall Introduction to Virtual Special Issue on Pharmaceutical Cocrystals,* Cryst. Growth Des., **9**, 4208-4211 (2009); (d) H. G. Brittain, *Cocrystal Systems of Pharmaceutical Interest: 2010,* Cryst., Growth Des., **12**, 1046-1054 (2012); (e) N. Qiao, *Pharmaceutical Cocrystals: An Overview,* Int. J. Pharm., **419** 1-11 (2011)

[4.82] https://www.novartis.com/news/media-releases/novartis-new-heart-failure-medicine-lcz696-now-called-entrestotm-approved-fda

[4.83] C. Saal, *Selection of Solid-State Forms: Challenges, Opportunities, Lessons and Adventures from Recent Years,* J. Pharm. Pharma., **67**, 755-756 (2015)

[4.84] S. Aitipamula, *Novel Solid Forms of the Anti-Tuberculosis Drug, Isoniazid: Ternary and Polymorphic Crystals,* Cryst. Eng. Comm., **15**, 5877-5887 (2013)

[4.85] B. S. Sekhon, *Drug-Drug Co-crystals,* DARU J. Pharm. Sci., **20**:45 (2012)

[4.86] (a) M. Kaftory, *Role of Hydrogen Bonding in Determining the Crystal Structures of the Adducts between Acetone and Urea Derivatives,* Chem. Mater., 6, 1245-1249 (1994); (b) H.G. McAdie, *Thermal Decomposition of Molecular Complexes I. Urea-n-Paraffin Inclusion Compounds,* Canadian Journal of Chemistry, 40(12), 2195-2203 (1962); (c) O. Redlich, *Addition Compounds of Urea and Organic Substances,* J. Am. Chem. Soc., 72(9), 4153-4160 (1950); (d) L.P Varakina, *Interaction of Urea with Dihydric Phenols in the Solid Phase,* Zhurnal Obshchei Khimii, 43(10), 2105-2110 (1973); (e) M.E. Jamróz, *The Urea-Phenol(s) Systems,* Fluid Phase Equilibria, 152, 307-

326 (1998); (f) K. Roberts, *Theoretical Investigation of Certain Urea Adduct Structures and Their Hydrogen Bonding,* Dept. Pure and Applied Chemistry, University of Strathclyde, Glasgow

[4.87] H. Gillier-Pandraud, *No 350. – Structure cristalline du phenol*, Bulletin de la Société Chimique de France, 5ᵉ Series (6), 1988-1995 (1967)

[4.88] A.L. MacDonald, *Structure of the 2:1 Complex of Phenol and Urea,* Acta Cryst., C43, 676-678 (1987)

[4.89] J. M. Robertson, *The Structure of Resorcinol A Quantitative X-Ray Investigation,* Proceedings of the Royal Society of London, Series A, Matmatical, Physical and Engineering Sciences, 157, 79-99 (1936); (b) J. Monteath Robertson, *A New Form of Resorcinol I. Structure Determination by X-Rays,* Proceedings of the Royal Society of London, Series A, Matmatical, Physical and Engineering Sciences, 167, 122-135 (1938); (c) G.E. Bacon, *A Neutron Powder Diffraction Study of Deuterated α and β-Resorcinol,* Acta Cryst., B36, 1908-1916 (1980)

[4.90] M. Pickering, *Structure of the 1:1 Complex of Resorcinol and Urea,* Acta Cryst., B38, 3161-3163 (1982)

[4.91] W.A. Caspari, *CCCXCIII - The Crystal Structure of Quinol. Part I.,* J. Chem. Soc., 2945 – 2948 (1926); S.C. Wallwork, *The Crystal Structure of the α Form of Quinol,* J. Chem. Soc. Perkin Trans. 2., 637-640 (1980); S.V. Lindeman, *The β-Modification of Hydroquinone $C_6H_6O_2$,* Cryst. Struct. Comm., 10, 1173-1179 (1981); W.A. Caspari, *CXLVII - The Crystal Structure of Quinol. Part II.,* J. Chem. Soc., 1093 - 1095 (1927); K. Maartmann-Moe, *The Crystal Structure of γ-Hydroquinone,* Acta Cryst., 21, 979 – 982 (1966)

[4.92] M.M. Mahmoud, *The Crystal Structure of the 1:1 Complex Between Quinol and Urea,* Acta Cryst., B31, 338-342 (1975)

[4.93] L. T. Zheng, *Anti-Inflammatory Effects of Catechols in Lipopolysaccharide-Stimulated Microglia Cells: Inhibition of Microglial Neurotoxicity,* Eur. J. Pharm, **588**, 106-113 (2008)

[4.94] (a) C. J. Brown, *The Crystal Structure of Catechol,* Acta. Cryst., **21**, 170-174 (1966); (b) H. Wunderlich, *Die Kristallstruktur von Brenzcatechin: Eine Neubestimmung,* Acta Cryst., B27, 1684 – 1686 (1971); (c) J. Clastre, *Cristallochimie – Structure cristalline du pyrocatéchol,* C. R. Acad. Sc. Paris, 260(8), 2518 – 2520 (1965)

[4.95] R. B. Hammond, *Application of Systematic Search Methods to Studies of the Structure of Urea-Dihydroxy Benzene Crystals,* J. Phys. Chem. B, **107,** 11820-11826 (2003)

[4.96] S. L. Mayo, *DREIDING: A Generic Force Field For Molecular Simulations,* J. Phys. Chem., **94**(26), 8897-8909 (1990)

[4.97]   J. P. M. Lommerse, *A Test of Crystal Structure Prediction of Small Organic Molecules,* Acta Cryst., **B56**, 697-714 (2000)

[4.98]   (a) W. D. S. Motherwell, *Crystal Structure Prediction of Small Organic Molecules: A Second Blind Test,* Acta Cryst., **B58**, 647-661 (2002); (b) J. A. R. P. Sarma, *The Supramolecular Synthon Approach to Crystal Structure Prediction,* Cryst. Growth Des., **2**(2), 93-100 (2002)

[4.99]   (a) G. M. Day, *A Third Blind Test of Crystal Structure Prediction,* Acta Cryst., **B61**, 511-527 (2005); (b) G. M. Day, *An Assessment of Lattice Energy Minimization for the Prediction of Small Organic Crystal Structures,* Cryst. Growth and Des., **4**(6), 1327-1340 (2004)

[4.100] G. M. Day, *Significant Progress in Predicting the Crystal Structures of Small Organic Molecules – A Report on the Fourth Blind Test,* Acta Cryst., **B65**, 107-125 (2009)

[4.101] D. A. Bardwell, *Towards Crystal Structure Prediction of Complex Organic Compounds – A Report on the Fifth Blind Test,* Acta Cryst., **B67**, 535-551 (2011)

[4.102] A. M. Reilly, *Report on the Sixth Blind Test of Organic Crystal Structure Prediction Methods,* Acta Cryst., **B72**, 439-459 (2016)

[4.103] W. H. Press, *Numerical Recipes in Fortran 77, The Art of Scientific Computing, Second Edition, Volume 1 of Fortran Numerical Recipes*, Cambridge University Press (1992)

[4.104] G. M. Day, *Current Approaches to Predicting Molecular Organic Crystal Structures,* Cryst. Rev., **17**(1), 3-52 (2011)

[4.105] (a) M. A. Newmann, *A Major Advance in Crystal Structure Prediction,* Angew. Chem. Int. Ed., **47**, 2427-2430 (2008); (b) M. A. Neumann, *Energy Ranking of Molecular Crystals Using Density Functional Theory Calculations and an Empirical van der Waals Correction,* J. Phys. Chem., **B109**, 15531-15541 (2005); (c) M. A. Neumann, *Tailor-Made Force Fields for Crystal-Structure Prediction,* J. Phys. Chem. **B112**, 9810-9828 (2008)

CHAPTER 5 – CONCLUSIONS AND FUTURE WORK

5.1   Conclusions and Future Work

Since carrying out this study there have been many advances made in the area of modelling potentials and these have been discussed throughout the previous chapters. From reading about the blind tests set by the CSD it is apparent that such developments have increased the capabilities of solving crystal structures as well as expanding the types of molecules these methods may be applied to [5.1].

The study carried out here appears to show a very simple and focused approach when compared to the methodologies used today. With increased complexities of potentials and increased computational power available it is clear that vast arrays of parallel searches may be carried out for larger and more complex systems [5.2]. Searches have progressed from seeking the single most thermodynamically stable structure at one set temperature. Modern methods tend more towards searching ranges of temperatures for all possible stable forms. The assessment of the results through clustering and comparing against other physical properties has developed a method of ranking structures using the crystal energy landscape [5.3]. This process has highlighted how close in energy two different structures can be and how accurate and reliable the models need to be when calculating relative energies.

The validation chapter in this study could be developed by recalculating the energy values using the newer more accurate potential models. It would be interesting to see a comparison between the atomistic approach and quantum mechanical calculations. A lot of work has been carried out on carbamazepine already and some work was carried out in this study at growing the triclinic form of carbamazepine crystals. Unfortunately the structure of the triclinic form was solved using single crystal data [5.4] before completion of the powder investigation but it might be interesting to complete this work as the initial results, such as the unit cell parameters correlated well with the published single crystal data.

The cocrystal work could again benefit from reassessment using methodologies which proved successful in the CSD blind tests [5.5]. It would be good to try expanding the series of tests carried out on the grinding method of producing the cocrystals to begin with. Is an intermediate form produced and what is its structure? Is the transition process

improved using drop assisted grinding? Does the nature of the solvent used change anything about the structure of the cocrystal? The grinding method was carried out using a simple pestle and mortar, can experimental conditions be improved or standardized using mechanized methods? For crystallization of the cocrystal from solvation only water was investigated. What happens when other solvents are used and are solvates formed in preference at any point? The solution was warmed gently during this experiment and standardization of the experimental process should be attempted. Using different solvents at varying temperatures would be an interesting study into whether different ratios of the cocrystal are made. This is effectively following a polymorph screening approach which is a routine first step in industries where polymorphs can make a big difference to the product performance [5.6 - 5.9].

Work also began on the trihydroxybenzene series of compounds and whether they would form cocrystals with urea. Most work was carried out using pyrogallol but completion of this study using the methods outlines above for catechol urea would be of interest.

Overall, for such simple systems, the approach in this study worked well. The validation of the simulated annealing method was successful for two forms of both primidone and carbamazepine. The cocrystal of catechol urea has been published elsewhere using different methodology but appears to compare well with the results reported here [5.10].

5.2 References

[5.1]   G. M. Day, *Current Approaches to Predicting Molecular Organic Crystal Structures,* Cryst. Rev., **17**(1), 3-52 (2011)

[5.2]   A. M. Reilly, *Report on the Sixth Blind Test of Organic Crystal Structure Prediction Methods,* Acta Cryst., **B72**, 439-459 (2016)

[5.3]   (a) S. L. Price, *Why Don't we Find More Polymorphs?,* Acta Cryst., **B69,** 313-328 (2013); (b) S. Z. Ismail., *Evaluating a Crystal Energy Landscape in the Context of Industrial Polymorph Screening,* Cryst. Growth and Des., **13,** 2396-2406 (2013); (c) S. L. Price, *Predicting Structures of Organic Compounds,* Chem. Soc. Rev., **43,** 2098-2111 (2014); (d) S. L. Price, *Can Computed Crystal Energy Landscapes Help Understand Pharmaceutical Solids?,* Chem. Commun., **52,** 7065-7077 (2016); (e) H. P. G. Thompson, *Which Conformations make Stable Crystal Structures? Mapping Crystalline Molecular Geometries to the Conformational Energy Landscape,* Chem. Sci., **5,** 3173-3182 (2014)

[5.4]   M. Lang, *Form IV of Carbamazepine,* J. Pharm. Sci., **91**(4), 1186-1190 (2002)

[5.5] (a) J. P. M. Lommerse, *A Test of Crystal Structure Prediction of Small Organic Molecules,* Acta Cryst., **B56**, 697-714 (2000); (b) W. D. S. Motherwell, *Crystal Structure Prediction of Small Organic Molecules: A Second Blind Test,* Acta Cryst., **B58**, 647-661 (2002); (c) J. A. R. P. Sarma, *The Supramolecular Synthon Approach to Crystal Structure Prediction,* Cryst. Growth Des., **2**(2), 93-100 (2002); (c) G. M. Day, *A Third Blind Test of Crystal Structure Prediction,* Acta Cryst., **B61**, 511-527 (2005); (d) G. M. Day, *An Assessment of Lattice Energy Minimization for the Prediction of Small Organic Crystal Structures,* Cryst. Growth and Des., **4**(6), 1327-1340 (2004); (e) G. M. Day, *Significant Progress in Predicting the Crystal Structures of Small Organic Molecules – A Report on the Fourth Blind Test,* Acta Cryst., **B65**, 107-125 (2009); (f) D. A. Bardwell, *Towards Crystal Structure Prediction of Complex Organic Compounds – A Report on the Fifth Blind Test,* Acta Cryst., **B67**, 535-551 (2011); (g) A. M. Reilly, *Report on the Sixth Blind Test of Organic Crystal Structure Prediction Methods,* Acta Cryst., **B72**, 439-459 (2016)

[5.6] (a) G. R. Desiraju, *Crystal Engineering: Structure, Design and Function,* Curr. Opin. Solid State Mat. Sci., **13**, 35 (2009); (b) J. Rantanen, *The Future of Pharmaceutical Manufacturing Sciences,* J. Pharm. Sci., **104**, 3612-3638 (2015); (c) S. Datta, *Crystal Structures of Drugs: Advances in Determination, Prediction and Engineering,* Nat. Rev. Drug Discovery, **3**, 42-57 (2004); (d) R. D. Bailey Walsh, *Crystal Engineering of the Composition of Pharmaceutical Phases,* Chem. Commun., 186-187 (2003); (e) G. R. Desiraju, *Designer Crystals: Intermolecular Interactions, Network Structures and Supramolecular Synthons,* Chem. Commun, 1475-1481 (1997); (f) J. Elliott, *Pharmaceutical Materials Science: An Active New Frontier in Materials Research,* MRS Bull., **31**, 869-873 (2006); (g) C. B. Aakeröy, *Recent Advances in Crystal Engineering,* Cryst. Eng. Comm., **12**, 22-43 (2010); (h) N. Feeder, *The Integration of Solid-Form Informatics into Solid-Form Selection,* J. Pharm. Pharmacol., **67**, 857-868 (2015); (i) M. D. Ticehurst, *Integration of Active Pharmaceutical Ingredient Solid Form Selection and Particle Engineering into Drug Product Design,* J. Pharm. Pharmacol., **67**, 782-802 (2015); (j) C.B. Aakeröy, *Building Co-crystals with Molecular Sense and Supramolecular Sensibility,* Cryst. Eng. Comm., 7(72), 439-448 (2005); (k) J. D. Wuest, *Engineering Crystals by the Strategy of Molecular Tectonics,* Chem Commun., 5830-5837 (2005); (l) G. R Desiraju, *Crystal Engineering: A Holistic View,* Angew Chemie Int. Ed., **46**(44) 8342-8356 (2007)

[5.7] (a) M. M. Parmar, *Polymorph Selection with Morphology Control Using Solvents,* Cryst. Growth and Des., **7**(9), 1635-1642 (2007); (b) A. T. M. Serajuddin, *Salt Formation to Improve Drug Solubility,* Adv. Drug Del. Rev., **59**, 603-616 (2007)

[5.8] (a) M. Hartenfeller, *Enabling Future Drug Discovery by de novo Design,* WIREs Computational Molecular Science, **1**, 742-759 (2011); (b) G. Schneider, *De Novo Design: From Models to Molecules,* De Novo Molecular Design, Chapter 1, G. Schneider (Ed.), 1-55 (2014)

[5.9] (a) G. Schneider, *De Novo Design of Molecular Architectures by Evolutionary Assembly of Drug-Derived Building Blocks,* J. Comput. Aid. Mol. Des., **14**, 487-494 (2000); (b) K. Loving, *Computational Approaches for Fragment-Based and De Novo Design,* Curr. Top. Med. Chem., **10**, 14-32 (2010); (c) A. S. Reddy, *Structure-Based De Novo Drug Design,* De Novo Molecular Design, Chapter 4, G. Schneider (Ed.), 97-123 (2014); (d) V. J. Gillet, *De Novo Molecular Design,* Evolutionary Algorithms in Molecular Design, Chapter 4, D. E. Clark (Ed.), 49-69 (2000)

[5.10] R. B. Hammond, *Application of Systematic Search Methods to Studies of the Structure of Urea-Dihydroxy Benzene Crystals,* J. Phys. Chem. B, **107,** 11820-11826 (2003)