

# Computational prediction of organic crystal structures

Yiming Ding



November 2022

Submitted in partial fulfilment of the requirements for the degree of

**Master of Philosophy (MPhil)**

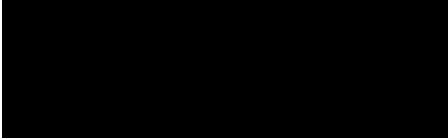
to

**University College London**

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

## **Declaration**

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



Yiming Ding

November 2022

## Abstract

Crystal Structure Prediction (CSP) is used by the pharmaceutical industry to assess whether other polymorphs of active pharmaceutical ingredients (API) might cause problems during manufacturing processes. In the 7th Blind Test of CSP, organized by the Cambridge Crystallographic Data Centre (CCDC), one of the targets (XXX) was to predict the possible stoichiometries of two co-crystals of cannabinol (CBN) and tetramethylpyrazine (TMP). This thesis describes the methodology used for the submission of predicted structures of these co-crystals, concluding that the likely stoichiometries were 1CBN:1TMP and 1CBN:2TMP, as these were more stable than the component structures and had plausible crystal packings.

Following submission, this thesis analysed the crystal structures of TMP and have proposed starting points for the crystal structure refinement of a structure (MPYRAZ03) on the Cambridge Structural Database (CSD) that has no atomic coordinates. The CBN search failed to find the  $Z'=2$  experimental crystal structure (CANNOL) that is on the CSD, which has a high energy molecular conformation. This failure was found to be due to the limits on the structure generation program (Sobol sequence and density setting) and was exacerbated by the point charge model failing to model the CANNOL hydrogen bonding adequately. Alternative strategies to find the experimental structure were proposed, but they were deemed too expensive to run a full search.

As this thesis was being completed, experimental co-crystal structures were provided by CCDC. After comparing with experimental structures, there was no experimental co-crystal structure in co-crystal CSP searches used in this thesis. This problem was caused by the folded pentane tail instead of the hydroxyl group.

## Impact Statement

The engineering of co-crystals can change many properties, such as stability in humid conditions, compared to the pure active pharmaceutical ingredient (API) without making and/or breaking covalent bonds to change the API molecule. Forming a co-crystal with a non-toxic conformer (another molecule) may be a route to finding a good solid form of the API to use in the pharmaceutical product. Many experimental co-crystal structures have been found in the laboratory and co-crystals are very common in industry, so co-crystal structure prediction is very important. Co-crystals have more components and more atoms than pure crystals, thus, it is more demanding to attempt to predict the existence and structure of co-crystals than the pure API crystal structure.

Since 1999 the Cambridge Crystallographic Data Centre has held Crystal Structure Prediction Blind Tests to bring together scientists and advance methods in this field. In these Blind Tests, the molecular diagrams are given to the groups involved in developing CSP methods, with the challenge of predicting the crystal structures of the molecule, by submitting their predictions by a strict deadline. The predictions are then compared with the experimental structures (which have been kept secret) so that everyone can see whether the CSP methods work. In industrial pharmaceutical development, time and cost are important. It is hoped that CSP can help industry reduce the risk in the drug development stage that the solid form experimental work has not identified all important polymorphs. There are multiple companies whose business is to provide CSP services to industry, who are participating in the current Blind Test, and so the CCDC tests have commercial as well as academic significance.

In the 7<sup>th</sup> Blind test, one of the challenges was to predict the stoichiometries and structures of two co-crystals of tetramethyl-pyrazine (TMP) and cannabinol (CBN). This thesis describes how I tackled this co-crystal challenge. It revealed that the methodology for CSP that was used in this thesis was not adequate for even predicting the crystal structure of CBN. The reason has been found for this and proposed a variation on the CSP approach.

The CCDC have already released the preliminary result that no group predicted one of the co-crystal structures successfully. There will be considerable

discussion of the results of the 7<sup>th</sup> Blind Test, which will eventually lead to a publication, but it is clear that there are still limitations on the ability of CSP to predict drug co-crystal structures. The importance of co-crystals as a way of delivering APIs to patients means that there will be further work on developing CSP methods.

## **Acknowledgements**

First, I thank my first supervisor Prof. Sarah (Sally) L. Price, many thanks for her support during this project. Many thanks for that she gave me an opportunity to join in her research group at UCL. I thank my second supervisor Martinus Antonius Zwijnenburg, many thanks for his support during this project.

I extend special thanks to Dr. Louise Price and Dr. Rui Guo for the training and help. In particular Dr. Louise Price carried out the CBN fragments analysis (Chapter 4 section 4.2.1 Figure 14). Dr Louise Price and Dr Rui Guo carried out the work on coronene, and discussions with them on this case is the basis for some of the work in Chapter 6.

Thank you very much to everyone in the Price group for their help. I can raise any questions about science or life during each group meeting and coffee break, and they will actively help me. I cannot complete this project without their help.

Finally, I want to thank my parents who supported the funding for this project.

## Table of Contents

1	Introduction.....	21
1.1	Co-crystals.....	24
1.1.1	The importance of co-crystals .....	24
1.1.2	Why is CSP for co-crystals more difficult than single components 24	
1.2	Blind Test.....	25
1.2.1	Previous Blind Tests.....	25
1.2.2	XXX target in 7 <sup>th</sup> Blind Test .....	26
1.3	Thesis structure .....	27
2	Theoretical background to Crystal Structure Prediction (CSP).....	29
2.1	Introduction.....	29
2.1.1	Definition of crystal internal structure by diffraction .....	30
2.1.2	The Cambridge Structural Database (CSD) .....	31
2.1.3	Mercury .....	32
2.1.4	PLATON.....	33
2.1.5	Packing Coefficients.....	33
2.1.6	Lattice energy and temperature effects .....	35
2.1.7	Co-crystal and co-crystallization energy .....	35
2.2	Quantum chemistry.....	36
2.2.1	Wavefunction based methods (HF, MP2).....	37
2.2.2	Density Functional Theory (DFT) .....	37
2.2.3	Semi-empirical periodic methods e.g., Density Functional Tight-Binding (DFTB).....	39
2.2.4	Basis sets for molecules and plane waves (with pseudopotential for periodic).....	41
2.2.5	Dispersion term .....	41
2.3	Electronic structure methods and codes used in this work .....	44

2.3.1	Periodic electronic structure ( $\Psi_{\text{cry}}$ ) methods .....	44
2.3.2	Molecular ( $\Psi_{\text{mol}}$ ) method of modelling organic crystal structures.	46
2.4	Crystal Structure Prediction Two steps method (General methodology)	48
2.4.1	Global search phase (Structure generation).....	49
2.4.2	Structure optimization stage (Energy evaluation and ranking) .....	51
2.5	Conclusion .....	52
3	CSP of tetramethyl pyrazine (TMP).....	54
3.1	Methodology .....	54
3.1.1	CSP methodology .....	54
3.1.2	Generation and optimization of crystal structures .....	54
3.2	Results.....	55
3.2.1	TMP conformation and its symmetry .....	55
3.2.2	TMP rigid search optimized by DMACRYS with rigid methyl groups	57
3.2.3	Use of CSP to suggest atomic coordinates for monoclinic polymorph	58
3.3	Conclusion .....	60
3.3.1	Lattice energies for comparisons for predicting co-crystal formation and stoichiometry .....	60
4	CSP of Cannabinol (CBN).....	62
4.1	Adding hydrogen atom positions to experimental structure .....	62
4.2	Methodology .....	63
4.2.1	Molecular flexibility .....	63
4.2.2	Generation and optimization of crystal structures .....	68
4.3	CBN CSP Results.....	70
4.4	Conclusion .....	73



4.4.1	Lattice energies for comparisons for predicting co-crystal formation and stoichiometry .....	74
5	Co-crystal CSP .....	76
5.1	Methodology for 7 <sup>th</sup> CCDC Blind Test .....	76
5.1.1	Generation of crystal structures .....	76
5.2	Results .....	78
5.2.1	1CBN:1TMP co-crystal .....	78
5.2.2	1CBN:2TMP co-crystal .....	83
5.2.3	2CBN:TMP co-crystal .....	86
5.3	Conclusion .....	91
5.3.1	Blind Test submission of structures in order of $\Delta E_{CC}$ . .....	91
5.3.2	Limitations of methods and confidence. ....	91
5.3.3	Completeness of search and energy rankings .....	92
5.3.4	What are submitted .....	92
5.3.5	Co-crystal confidence .....	93
6	Analysis of limitations of the CSP method used for CBN and TMP and co-crystals .....	97
6.1	Rigid methyl group assumption for TMP .....	97
6.1.1	A flexible CSP for TMP .....	97
6.1.2	Comparing TMP rigid search and flexible search generated in CrystalPredictor .....	98
6.1.3	Comparing TMP rigid search optimized by CrystalOptimizer for flexible methyl groups and DMACRYS with rigid methyl groups .....	99
6.2	CSP for TMP experimental structure without atomic coordinate in CSD 104	
6.2.1	Conclusion .....	106
6.3	Limitations of CrystalPredictor in CSP of CBN .....	108
6.3.1	Limitation of Sobol sequence (generation step) .....	108

6.3.2	Limitation of point charge + FIT model .....	109
6.3.3	Conclusion .....	112
6.4	Testing alternative methods.....	114
6.4.1	New search using Workflow 2 .....	114
6.4.2	New search using workflow 3.....	115
6.4.3	Conclusion .....	117
6.5	Sensitivity to potential model .....	117
6.5.1	Methodology.....	117
6.5.2	Results .....	119
6.6	Conclusion.....	121
6.7	Appendix: monoclinic TMP suggested .res file .....	123
7	Conclusions, before and after receiving experimental structures for co-crystals.....	124
7.1	Conclusions – towards a CSP method for CBN structures .....	125
7.2	The experimental information on XXX, the co-crystals of CBN and TMP. 126	
7.3	Outlook for CSP.....	129
8	References.....	130

## List of figures

Figure 1 Co-formers of XXX target (right: Tetramethylpyrazine (TMP) left: Cannabinol (CBN)) with the atomic numbering used in this work. In the co-crystal CSP work, the numbering starts with the CBN molecules and then TMP and follows the same sequence.

Figure 2 The grid used in packing coefficient calculation and voids space calculation. Black points are the crystal components molecules occupied space. Blue point is any point that is more than a certain distance (usually it is the water molecular radius) from the edges of the molecules, and green points are within this probe radius distance of such a point included too.

Figure 3 The thermodynamic cycle used to calculate co-crystal formation and co-crystallization energy, which is adapted from reference<sup>24</sup>.

Figure 4 Left figure is the molecular overlay of MPYRAZ02 (light green) and gas phase conformation (cyan) for TMP (RMSD<sub>1</sub>=0.0604 Å with hydrogen atoms and 0.0175 Å without hydrogen atoms) and the torsion angles shown in follow table.

Figure 5 Summary plot of the lattice energy landscape versus cell density for TMP from Z'=1 search after DMACRYS rigid refinement with PBE0/6-31G(d,p). The second lowest lattice energy structure orthorhombic polymorph is depicted by the red diamond, which is equal to lattice energy minimum starting from MPYRAZ02 with same optimization method (DMACRYS). The global minimum structure monoclinic depicted by a red circle is approximately a supercell of monoclinic MPYRAZ03.

Figure 6 Summary plot of the lattice energy landscape versus packing coefficient for TMP from Z'=1 search after DMACRYS rigid refinement with PBE0/6-31G(d,p). The second lowest lattice energy structure orthorhombic polymorph is depicted by the red diamond, which is equal to lattice energy minimum starting from MPYRAZ02 with same optimization method (DMACRYS). The global minimum structure monoclinic depicted by a red circle is approximately a supercell of monoclinic MPYRAZ03.

Figure 7 Left figure is the overlay (RMSD<sub>15</sub>=0.172 Å) of the experimental structure MPYRAZ02 and generated structure T131 optimized by DMACRYS using PBE0

6-31G(d,p), which is the second lowest lattice energy structure. Right figure is the overlay ( $\text{RMSD}_{15}=0.103 \text{ \AA}$ ) for MPYRAZ02 optimized by DMACRYS using PBE0/6-31G(d,p) and generated structure T131 (second lowest lattice energy structure) optimized by DMACRYS using PBE0/6-31G(d,p) shows that they are equivalent structures and the experimental orthorhombic polymorph has been found in this search.

Figure 8 Symmetry elements for generated T11 optimized by DMACRYS using PBE0 6-31G(d,p) (left) and symmetry elements for optimized T11 with symmetry improved by PLATON (middle) and symmetry elements for MPYRAZ03 (right) collected from CSD.

Figure 9 Diagram packing conformation of CBN experimental structure (CANNOL) showing the hydrogen atoms positions added between two oxygens (left) where added hydrogen atoms is coloured white, and oxygen atoms are coloured red. In experimental structure (CANNOL in CSD), there are two molecules in the asymmetric unit with different conformations (right).

Figure 10 Relaxed conformational energy scans of CBN for hydroxyl group (OH) torsion angle at various levels of theory. Basis set: 6-31G(d,p). The inset diagrams show the CBN conformations with torsion  $\text{H26\_O2\_C6\_C7} = 50^\circ$  (mol 1 in CANNOL),  $200^\circ$ ,  $360^\circ$  (mol 2 in CANNOL). The angles in the experimental structure of CANNOL are indicated by black diamonds.

Figure 11 Relaxed conformational energy scans of CBN for the pentane tail at different computational methods with 6-31G(d,p). The inset diagrams show the CBN conformations with torsion  $\text{C15\_C14\_C8\_C7} = 0, 90, 180, 270, 0/360$  ( $^\circ$ ), with two minimum points for  $90^\circ$  and  $270^\circ$ .

Figure 12 Relaxed conformational energy scans of CBN for the pentane tail at different DFT functional with 6-31G(d,p). The inset diagrams show the CBN conformations with torsion  $\text{C17\_C16\_C15\_C14} = 0, 60, 120, 180, 240, 300$  ( $^\circ$ ).

Figure 13 CSD Conformer Generator analysis for the pentane tail ( $\text{C15\_C14\_C8\_C7}$  and  $\text{C17\_C16\_C15\_C14}$ ) in CBN mol1, which matches the conformation energy scan results.  $\text{C17\_C16\_C15\_C14}$  is focused on  $180^\circ$  and  $60^\circ$  and  $300^\circ$ ;  $\text{C15\_C14\_C8\_C7}$  is focusing on  $90^\circ$  and  $270^\circ$  but caused by the

small energy barrier, there are large half peak width around those two minimum point. C15\_C14\_C8\_C7 is different from others as C7 and C8 are in the ring.

Figure 14 Fragments analyzed for CBN pentane tail shows that C17\_C16\_C15\_C14 is focused on  $90^\circ$  and  $270^\circ$  but caused by the small energy barrier, there are large half peak width around those two minimum point, which is similar to the Conformation Generator analysis and conformation energy scan analysis; C17\_C16\_C15\_C14, C18\_C17\_C16\_C15, and C16\_C15\_C14\_C8 have the similar distribution to each other, which is focused on the  $140^\circ$ – $220^\circ$ .

Figure 15 Left figure is the overlay of the experimental structure CANNOL (green) and experimental structure optimized using CrystalOptimizer with PBE0 6–31G(d,p) and FIT potential for hydroxyl group, pentane tail and ring discussed in the last part.  $\text{RMSD}_{15}$  is 0.17 Å. The middle and right figures are the overlay of CBN experimental conformations and optimized conformations using CrystalOptimizer with PBE0 6–31G(d,p) and FIT potential for hydroxyl group, pentane tail and ring discussed in the last part.  $\text{RMSD}_1$  for mol1 is 0.0922 Å and  $\text{RMSD}_1$  for mol2 is 0.116 Å without hydrogen atoms.

Figure 16 Summary plot of the lattice energy landscape versus cell density for CBN from  $Z'=2$  search after optimized using CrystalOptimizer with PBE0/6–31G(d,p). The lattice energy minimum starting from CANNOL with the same methodology is denoted as E1.

Figure 17 Summary plot of the lattice energy landscape versus packing coefficient for CBN from  $Z'=2$  search after optimized using CrystalOptimizer with PBE0/6–31G(d,p). The lattice energy minimum starting from CANNOL with the same methodology is denoted as E1

Figure 18 Generated CBN crystal structure (C5 (P21/c) shown in the appendix) formed by C1,1(6) a D1,1(2) b D3,3(9) >b>a<b hydrogen bond which is same to experimental structure, but their crystal structure are totally different and the hydrogen bond between mol1 and mol2 is too large (3.250 Å).

Figure 19 Comparing C5 to experimental structure. The left is the comparison of the chain; the right is the comparison of the asymmetry unit. CANNOL is coloured by elements, and C5 is coloured by blue and green with  $Z'=2$ .

Figure 20 Lattice energy landscape (density) for co-crystal 1CBN:1TMP search. Blue line:  $E(\text{CBN,exp,opt})+E(\text{TMP,matching exp,rig})=-214.499$  kJ/mol, the sum of experimental CBN and TMP lattice energy after CrystalOptimizer.

Figure 21 Lattice energy landscape (packing coefficient) for co-crystal 1CBN:1TMP search. Blue line:  $E(\text{CBN,exp,opt})+E(\text{TMP,matching exp,rig})=-214.499$  kJ/mol, the sum of experimental CBN and TMP lattice energy after CrystalOptimizer.

Figure 22 Lattice energy landscape (density) for co-crystal 1CBN:2TMP search. Blue line:  $2E(\text{TMP, matching exp,rig})+ E(\text{CBN,exp,opt})=-288.399$  kJ/mol, the sum of experimental CBN and 2 TMP lattice energy after CrystalOptimizer.

Figure 23 Lattice energy landscape (packing coefficient) for co-crystal 1CBN:2TMP search. Blue line:  $2E(\text{TMP, matching exp,rig})+ E(\text{CBN,exp,opt})=-288.399$  kJ/mol. Blue line is the sum of experimental CBN and 2 TMP lattice energy after CrystalOptimizer.

Figure 24 Lattice energy landscape (density) for co-crystal 2CBN:1TMP search. The line is not drawn like 1CBN:1TMP and 1CBN:2TMP co-crystal searches, because all of those structures are higher than this cutoff energy, which is  $\text{Cutoff} = 2E(\text{TMP,matching exp,rig})+ E(\text{CBN,exp,opt})=-354.948$  kJ/mol.

Figure 25 Lattice energy landscape (packing coefficient) for co-crystal 2CBN:1TMP search. The line is not drawn like 1CBN:1TMP and 1CBN:2TMP co-crystal searches, because all of those structures are higher than this cutoff energy, which is  $\text{Cutoff} = 2E(\text{TMP, matching exp,rig})+ E(\text{CBN,exp,opt})=-354.948$  kJ/mol.

Figure 26 Crystal cell and voids for the lowest energy 1CBN: 2TMP co-crystal structures. The lowest structure contains H-bond and  $\pi$ - $\pi$  stacking.

Figure 27 1CBN:2TMP GM structure with the hydrogen bond and  $\pi$ - $\pi$  stacking. The distance of hydrogen bond is 2.8 Å, hydroxyl group torsion angle is  $-6.99^\circ$ , which is near to 0 in the ring plane.

Figure 28 Relaxed conformational energy scan of one methyl group in TMP (H2\_C5\_C2\_C1) at PBE0 6-31G(d,p). The black line shows that there is very

little molecular energy difference (no more than 0.5 kJ/mol) between the conformations with the methyl dihedral between  $50^\circ \sim 70^\circ$ .

Figure 29 The torsion angles of the TMP CH<sub>3</sub> groups (H7\_C7\_C3\_C4, H10\_C8\_C4\_C3, H1\_C5\_C2\_C1, H4\_C6\_C1\_C2,) for generated structures after the CrystalPredictor flexible search and are optimized using point charge + FIT. The black arrows denote the conformation in the rigid search.

Figure 30 Workflow for comparing TMP rigid search (chapter 3) optimized by CrystalOptimizer for flexible methyl groups and DMACRYS with rigid methyl groups.

Figure 31 Lattice energy landscape summary plot for TMP from Z'=1 rigid molecule search after CrystalOptimizer (left) and after DMACRYS rigid refinement (right). The orthorhombic polymorph is depicted by the orange diamond, which is lattice energy minimum structure form MPYRAZ02, and the GM structure with monoclinic polymorph is depicted by orange round, which is approximately a supercell of MPYRAZ03.

Figure 32 RMSD<sub>15</sub> and RMSD<sub>1</sub> between the generated structures matching the experimental structures optimized by CrystalOptimizer with 4 methyl groups and DMACRYS.

Figure 33 The comparison among A362 and experimental structure optimized by CASTEP with PBE-TS and CrystalOptimizer with PBE0/6-31G(d,p) +FIT potential.

Figure 34 pXRD of MPYRAZ02 (which has atomic coordinates) and MPYRAZ02 optimized by different methods. The major peaks are labelled by the lattice planes.

Figure 35 pXRD of generated structure matching the monoclinic MPYRAZ03 optimized by different methods. The major peaks are labelled by the lattice planes.

Figure 36 Workflow for rigid CBN search with experimental conformations and experimental structure.

Figure 37 Figures from left to right are CANNOL, CANNOL optimized by FIT+point charge with non-polar hydrogen atomic potential parameters, CANNOL optimized by FIT + point charge with Beyer polar hydrogen atomic potential parameters respectively. The hydrogen bond contact distance is defined as smaller than the vdW radii (defaults in Mercury).

Figure 38 3 workflows for a CBN CSP search. Workflow 1 is the original search, and it is used in the co<sup>o</sup>-crystal search. Workflow 2 and Workflow 3 are investigated new CSP workflows using DFTB3-D3 based on the workflow1 and aiming to correct the problem caused by the point charge model.

Figure 39 Workflow of testing sensitivity to potential model in DMACRYS with different basis sets, functional, and potential parameters. The colours of the rectangles match the colours of the arrows in Figure 41.

Figure 40 Dispersion-repulsion potential for hydroxyl group (grey) along a non-polar hydrogen ( $A_{HH}=11971$  kJ/mol,  $B_{HH}=3.74$  Å<sup>-1</sup>,  $C_{HH}= 136.4$  kJmol<sup>-1</sup>Å), compared to that along Coombes polar H(orange) ( $A_{HH}=5029.68$  kJ/mol,  $B_{HH}=4.66$  Å<sup>-1</sup>,  $C_{HH}= 21.50$  kJmol<sup>-1</sup>Å) and Beyer polar hydrogen model (blue) ( $A_{HH}=2263.3$  kJ/mol,  $B_{HH}=4.66$  Å<sup>-1</sup>,  $C_{HH}= 21.50$  kJmol<sup>-1</sup>Å). Black arrows highlight the O...H bond distance in the hydrogen bonds in the experimental structure CANNOL. This shows the experimental hydrogen bond repulsion and dispersion energy is sensitive to the different hydrogen parameters.

Figure 41 Energy comparison among 5 structures using different methods. Green arrow meaning the single point calculation using Beyer potential shown in Figure 39. Red arrow means the single point calculation using 6-31G(d,p) shown in Figure 39. Blue arrow means the single point calculation using MP2 shown in Figure 39.

Figure 42 The CBN molecule with black curved arrows qualitatively indicates the range of the flexible group torsions used in the CSP searches (Table 3).



## List of Tables

Table 1 Cell parameters for T11 after DMACRYS refinement using PBE0 6-31G(d,p), and cell parameters for optimized T11 with symmetry improved by PLATON, and cell parameters of MPYRAZ03 collected from CSD whose atomic coordinates do not in the CSD.

Table 2 The lowest TMP crystal structures generated by rigid search and experimental structure (MPYRAZ02) optimized by DMACRYS with PBE0 6-31G(d,p). The generated structures are numbered by the rank after point charge + FIT optimized by CrystalPredictor. MPYRAZ02 opt is the lattice energy minimum using same potential model starting from the MPYRAZ02 structure using the experimental conformation held rigid.

Table 3 Torsion angle ranges and intervals applied in CrystalPredictor LAM generation

Table 4 The lowest CBN crystal structures generated by rigid search and experimental structure (MPYRAZ02) optimized by DMACRYS with PBE0 6-31G(d,p). The generated structures are numbered by the rank after point charge + FIT optimized by CrystalPredictor. CANNOL opt is the structure optimized by CrystalOptimizer starting from the CANNOL.

Table 5 The most stable CSP generated co-crystal structures of CBN:TMP are named as the rank after point charge + FIT optimized by CrystalPredictor, for example, CT1 is the global minimum structure after point charge + FIT optimization. Reference energy is:  $E(\text{CBN,exp,opt}) + E(\text{TMP,matching exp,rig}) = -214.499$  kJ/mol. Co-crystallization energy is calculated by  $\Delta E_{cc} = E_{latt}(\text{co-crystal}) - E_{latt}(\text{TMP}) - E_{latt}(\text{CBN})$ . Structures that co-crystallization energies are lower than 0 are coloured as green.

Table 6 The most stable CSP generated co-crystal structures of CBN:2TMP are named as the rank after point charge + FIT optimized by CrystalPredictor, for example, CTT1 is the global minimum structure after point charge + FIT optimization. Reference energy is:  $2E(\text{TMP, matching exp,rig}) + E(\text{CBN,exp,opt}) = -288.399$  kJ/mol. Co-crystallization energy is calculated by

$\Delta E_{cc} = E_{latt}(co-crystal) - 2E_{latt}(TMP) - E_{latt}(CBN)$ . Structures that co-crystallization energies are lower than 0 are coloured as green.

Table 7 The most stable CSP generated co-crystal structures of 2CBN:TMP are named as the rank after point charge + FIT optimized by CrystalPredictor, for example, CCT1 is the global minimum structure after point charge+FIT optimization. Reference energy is:  $2E(TMP, \text{matching exp,rig}) + E(CBN, \text{exp,opt}) = -354.948$  kJ/mol. Co-crystallization energy ( $\Delta E_{cc}$ ) is calculated by  $\frac{E_{latt}(co-crystal) - E_{latt}(TMP)}{2} - E_{latt}(CBN)$ . To compare stabilization of co-crystal with different stichometry, the co-crystallization energy is calculated with respect to 1 mol CBN.

Table 8 The submitted 100 co-crystal structures including the label, space group, density and co-crystallization energy.

Table 9 Torsion group ranges in CrystalPredictor LAM generation for TMP flexible search

Table 10 Cell parameters and details of A1193, A362 and MPYRAZ02 optimized by PBE-TS or single point optimization.

Table 11 TMP monoclinic polymorphs crystal parameters. MPYRAZ03 is the experimental crystal structure in CSD. A1193 is the structure generated by CrystalPredictor with rigid search and optimized by CrystalOptimzier with PBE0/6-31G(d,p). T11 is the structure generated by CrystalPredictor with rigid search and optimized by DMACRYS with PBE0/6-31G(d,p). A1193 opt by CASTEP is the structure optimized by CASTEP with PBE-TS starting from A1193.

## List of Symbols and Abbreviations

TMP	Tetramethylpyrazine
CBN	Cannabinol
CT	1CBN:1TMP co-crystal
CTT	1CBN:2TMP co-crystal
CCT	2CBN:1TMP co-crystal
CSP	Crystal Structure Prediction
CCDC	Cambridge Crystallographic Data Centre
CSD	Crystal Structure Database
APF	Atomic fill factor
API	Active pharmaceutical ingredient
DMA	Distributed multipole analysis
DFT	Density functional theory
XRD	X-ray diffraction
pXRD	Powder X-ray diffraction
LDA	Local density approximation
HF	Hartree-Fock
GGA	Generalized Gradient Approximation density functional
LSD	Local Spin Density
PBE	Perdew-Burke-Ernzerhof density functional
STO	Slater-Type Orbital
GTO	Gaussian-Type Orbital
LCAO	Linear-combination-of-atomic-orbitals
SCC	Self-consistent charge

DFTB	Density Functional Tight-Binding
D3(BJ) function	Third order dispersion model with Becke-Johnson-damping function
TS	Tkatchenko and Scheffler dispersion correction
$\Psi_{\text{cry}}$ crystals	Crystal modelling approach based on electronic structure of crystals
$\Psi_{\text{mol}}$ molecule	Crystal modeling approach based on electronic structure of molecule
LAM	Local approximate model
$Z'$	The number of molecules in the asymmetry unit
$E_{\text{latt}}$	Lattice energy
$\Delta E_{\text{intra}}$	Intramolecular energy relative to gas phase
$U_{\text{inter}}$	Intermolecular energy
$\Delta E_{\text{cc}}$	Co-crystallization relative to components
GM	Global minimum structure
$\text{RMSD}_n$	Root-mean-square deviation for $n$ molecules

## 1 Introduction

The original reason why it is necessary to predict crystal structures is to design new materials with desired physical properties<sup>1</sup>, e.g. porous materials<sup>2</sup>. Crystal Structure Prediction (CSP) is now used in industrial pharmaceutical development<sup>3</sup>, because of polymorphism. Polymorphs describe the existence of a molecule in more than one crystal possible structure. Polymorphs formed by the same molecules will have the same solution and melt structure<sup>4</sup>. Polymorphism is of practical relevance to pharmaceuticals, explosives, pigments, agrochemicals, dyestuffs, and foods<sup>5</sup>. A polymorphic transition is "A reversible transition of a solid crystalline phase at a certain temperature and pressure (the inversion point) to another phase of the same chemical composition with a different crystal structure."<sup>5</sup>. However, as many polymorphs do not undergo a solid-state transformation, some metastable crystals usually appear to be stable. In drug development, active pharmaceutical ingredients (API) solid form screening is critical. Drugs not only can form a variety of polymorphs but also can crystallize with variety of solvent molecules to form variety of solvates and crystallize with various of other molecules to form a variety of co-crystals. They can exhibit different physicochemical properties to the solid form selected for development. Therefore, in the pharmaceutical industry, the use of computational methods to predict the solvates<sup>6</sup>, co-crystals, and polymorphs will be valuable.

The pharmaceutical industry is very interested in drug crystal forms because they need a high degree of control over the physicochemical properties of their products (most in tablets) that affect the performance, quality, and safety of pharmaceutical products, as well as their intellectual property value. McCrone<sup>4</sup> has claimed that "Each compound has different polymorphic forms, and in general, the number of known forms is directly proportional to the time and money spent studying the compound". It is less clear to what extent molecules have been screened to find important crystal forms. Awareness of the polymorphism prevalence has increased for the industry, with regulators expecting solid form properties of drugs to be well understood and controlled, and a strong desire to minimise the risk of sudden delayed emergence of crystalline forms that could threaten market availability, e.g. ritonavir<sup>7</sup> and rotigotine<sup>8</sup>. CSP combining other computational modelling methods e.g. Mercury (section 2.1.3) can infer not only

polymorphic propensity, but also extrapolate the calculation of solid-state properties related to drug product performance.

The most likely crystal structure should be the global minimum point in the free energy surface defined by the geometry of the unit cell and the position, orientation, and conformation of the molecules in the unit cell. Ostwald's rule<sup>9,10</sup> of stage states that metastable crystal structures usually are kinetically advantaged, so they are easier to crystallize than thermal stable structures in solution. Metastable structures should transform into thermal stable structures, but the timescale sometime could be too large. Thus, it has been claimed<sup>11</sup> that predicting the metastable and thermal stable structure can help show whether there is a more stable form that has yet to be found. Late appearing forms and disappearing polymorphic events have been observed in the past<sup>12,13</sup>. However, the proportion of compounds estimated to be at risk of late appearing, more stable forms may be much larger than most crystallization scientists expected. It has been estimated<sup>11</sup> that as to all small-molecule drugs on the market currently, between 15 and 45% of them are not the thermodynamically most stable crystal structure. These results imply that the most stable crystal structures of these small-molecule drugs are kinetically hindered. Nucleation kinetics depend on the catalysts, templates, solvents, cooling rate, and so on. It is important to note that experimental observations will always greatly underestimate the proportion of missing polymorphs. Many drugs on the market may simply not have been used and crystallized long enough to allow a proper empirical understanding of the scope of the phenomenon.

However, due to the cost of calculation, most current CSP methods are used to calculate the lattice energy instead of free energy. An approach ( $\Psi_{\text{mol}}$ ) that combines anisotropic atom-atom intermolecular potential models with density functional theory intramolecular energies was used in a computational study<sup>14</sup> of 508 polymorphic organic molecules and their 1061 experimentally determined crystal structures. Using lattice dynamics calculations to estimate the contribution of vibrations to free energy showed the polymorphic energy differences were generally very small: more than half of the polymorphic pairs were separated by less than 2 kJ/mol, while in only 5% of the cases the energy differences exceeded 7.2 kJ/mol<sup>14</sup>. The relatively small vibrational free energy contributions are large

enough to cause polymorphic stability reordering for 9% of polymorphic pairs below or at room temperature<sup>14</sup>.

Currently, how to predict organic crystals accurately is still a challenge, especially for big organic molecules and multi-component co-crystal systems. The intermolecular interactions between organic molecules are very complex. Modelling the dispersion is very important for organics. The relationship between the dispersion energy and electron correlation means that highly accurate and expensive electronic structure theory methods will be required.

Current computational methods for co-formers selection can be divided as follows<sup>15</sup>: knowledge-based (structural informatics and thermodynamics), physics-based (surface site interaction points, Hansen solubility parameters, synthesis engineering, CSP and COSMO-RS), and ML (machine learning) methods. Furthermore, ML+COSMO-RS<sup>16</sup>, which is a combined physics and ML-based method, was recently proposed. ML+COSMO-RS was better than COSMO-RS and ML. Most virtual interpolymer screening methods lack the long-range order (crystallinity) contribution of co-crystals. This limitation was successfully solved<sup>15</sup> by virtual co-former screening calculations based on CSP to guide co-crystal design. Co-former screening based on CSP significantly improved the prediction of challenging indomethacin and paracetamol co-crystal cases<sup>15</sup>.

Sugden proposes a high-throughput and systematic computational approach<sup>17</sup> aimed to identify API/co-former pairs that are difficult to be observed in experiments. Building on the well-established CSP method, the proposed method<sup>17</sup> achieves its efficiency without any expensive quantum mechanical calculations. This computational program<sup>17</sup> is used to perform co-crystal screening of a given API against a panel of potential co-formers. The energy difference between the sum of the most stable pure constituents lattice energies and the most stable co-crystal lattice energy is used to assess the stabilization of the co-crystal formation. The most stable co-crystal forms were determined by Ab initio CSP methods.

CrystalPredictor<sup>18,19</sup> and CrystalOptimizer<sup>20</sup>, which are approaches used in this thesis, have advantages for co-crystals, as the same LAM databases can be

used for all the searches. This can be extended to investigate a variety of co-formers efficiently. There have been many discussions about CSP being suitable for co-crystals<sup>15</sup>, and the 7<sup>th</sup> Blind Test target XXX<sup>21</sup> tests the current reliability of the available methods.

## 1.1 Co-crystals

Co-crystals<sup>22, 23</sup> are single-phase crystals composed of different compounds in a stoichiometric ratio. Co-crystals are multicomponent solids that are neither solvates nor simple salts.

### 1.1.1 The importance of co-crystals

The co-crystal can change properties compared to the pure API<sup>24</sup>. As many drug candidates are eliminated due to their low solubility<sup>25, 26</sup>, the ability to form a co-crystal with a non-toxic conformer (other molecules) may be a route to finding a better solid form of the API to use in the pharmaceutical product<sup>22,27</sup>. Pharmaceutical co-crystals have different properties from pure APIs without breaking or making covalent bonds<sup>28</sup>.

The co-crystals engineering makes the most favourable conditions using advantages of the specific properties of each co-former<sup>29</sup>. The principal idea<sup>28</sup> is to develop superior physic-chemical properties of the API while keeping the medical properties of the drug molecule itself constant. The co-crystal former can also be used solely in the purification and isolation of the API<sup>30,31</sup>. Many co-crystals experimental structures have been found in the laboratory<sup>23</sup>. Co-crystals are very common in the industry, so co-crystal structure prediction is very important.

### 1.1.2 Why is CSP for co-crystals more difficult than single components

There are additional variables defining the relative position of the components in the asymmetric unit cell for co-crystal structures prediction, which is significantly different from the single component crystal structures prediction.  $Z'$  is the number of molecules in the asymmetry unit. Co-crystals, solvates, and  $Z' > 1$  polymorphs have additional variables of the relative position of the component molecules. Thus, the components will be put in the asymmetry units, and then generate the crystal cell through the symmetry.



The issue of co-crystal prediction is similar to that of solvate prediction. There are two important challenges<sup>6</sup> in solvate prediction. A survey<sup>6</sup> of approximately 12,000 drug-like molecular solvates, that are extracted from the Cambridge Structural Database (CSD, which will be introduced in 2.1.2), revealed<sup>6</sup> that most of them have stoichiometric and asymmetric unit numbers of components that are commonly available by CSP. Quantum mechanical calculations on a subset of these drug-like molecular solvates show that<sup>6</sup> thermodynamics generally drive their formation.

## 1.2 Blind Test

To bring together scientists in this field and advance methods, the Cambridge Crystallographic Data Centre (CCDC)<sup>32</sup> has held CSP Blind Tests since 1999.<sup>33</sup> In these Blind Tests<sup>33</sup>, the molecular diagrams are given to the groups involved in developing CSP methods, with the challenge of predicting the crystal structures of the molecule. All participating groups are required to submit their predictions by a strict deadline. The predictions are then compared with the experimental structures (which have been kept a secret) so that everyone can see whether the CSP methods work. CSP methods are developed by using known crystal structures.

Now that there are multiple companies whose business is to provide CSP services to industries. Some of them are participating in the current Blind Test, so the CCDC tests have commercial as well as academic interest. There are lots of commercial CSP software. For industry CSP, both time and cost are important. For example, Xtalpi<sup>34</sup>, GRACE<sup>35</sup>, which uses cloud technology to greatly improve computing speed and computing cost, and OpenEye Scientific Software<sup>36</sup> which performs computational crystal structure prediction on rigid molecules. It is very expensive to do a commercial CSP, but CSP can help industries reduce the risk in the drug development stage by finding other polymorphs.

### 1.2.1 Previous Blind Tests

In the 6<sup>th</sup> Blind Test<sup>37</sup>, there were 5 target systems: a chloride salt hydrate, a polymorphic former drug candidate, a small nearly rigid molecule, a co-crystal and a bulky flexible molecule.

It was clear that<sup>37</sup> sufficiently accurate calculation of lattice energies was so expensive that almost all methods had a hierarchical approach: requiring an easily evaluated potential energy model to generate the millions of possible crystal structures and then one of more methods of evaluating the lattice energy with increasing accuracy. The structure generation stage (section 2.4.1) often uses a force field which is an analytical expression for the energy in terms of the atomic positions. CSP started with rigid molecules when the lattice energy was assumed as only the intermolecular energy. Many force fields were developed by fitting to organic crystal structures of rigid molecules, such as the 2001 Williams potential<sup>38</sup> or the FIT potential<sup>85,86,87</sup>, which combines earlier potentials by Williams and parameters for polar hydrogens.

The 4<sup>th</sup> Blind Test<sup>40</sup> was the first in which periodic density function calculations were used, and this method predicted all four targets within the three predictions allowed. This was the first time that any method has been successful for all targets. The earlier Blind Tests had a low success rate, despite being on small molecules.

Notably, there was a great increase in industrial interest in CSP after the 5<sup>th</sup> Blind Test<sup>41</sup>, which was the first successful prediction with a target XX (benzyl-(4-(4-methyl-5-(p-tolylsulfonyl)-1,3-thiazol-2-yl) phenyl) that was more typical of small molecules in pharmaceutical development. Two groups successfully predicted this structure as the globe minimum structure<sup>42</sup>. Both groups used the  $\Psi_{\text{mol}}$  method (section 2.3.2) and reduced the search space to a manageable level. The use of  $\Psi_{\text{mol}}$  method with multipoles and empirical potentials performed is better than  $\Psi_{\text{cry}}$  method (section 2.3.1) with DFT-D in this case<sup>42</sup>.

### 1.2.2 XXX target in 7<sup>th</sup> Blind Test

In the 7<sup>th</sup> Blind Test<sup>21</sup>, the organic molecules are larger than previous Blind Tests and include: an optoelectronic molecule containing Si and I, a metal organic containing Cu, small organic, a co-crystal, and three pharmaceutical or agrochemical molecules, including ones containing S and F, and an organic salt.

In this project, our research group participated the 7th Blind Test held by Cambridge Crystallographic Data Centre<sup>32</sup>. This thesis mainly focuses on target XXX which is a co-crystal challenge.

The 7<sup>th</sup> Blind Test target XXX is aiming to identify the stoichiometries of the two experimentally known co-crystal forms, with different stoichiometries. By 29/06/2021, participants should submit a list containing landscapes of up to 1500 and a list of one hundred structures in total ranked in order of likelihood of observation and a statement outlining the most likely observed stoichiometries of the system (justifying if this differs to that of the first two stoichiometries within the ranked list).

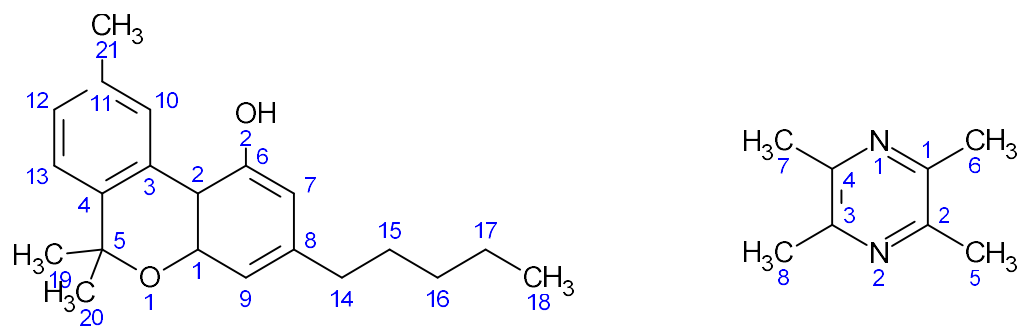


Figure 1 Co-formers of XXX target (right: Tetramethylpyrazine (TMP) left: Cannabinol (CBN)) with the atomic numbering used in this work. In the co-crystal CSP work, the numbering starts with the CBN molecules and then TMP and follows the same sequence.

Figure 1 shows the molecular structural formulas of two co-formers given by 7<sup>th</sup> Blind Test target XXX<sup>21</sup>. First, the spatial structure of each molecule needs to be considered, including the flexible groups. Those molecular structural formulas can be used to search in the CSD<sup>32</sup> (section 2.1.2 The Cambridge Structural Database (CSD)), which is helpful to understand CBN and TMP experimental structures. CSP search results on TMP and CBN are in chapter 3 and chapter 4.

### 1.3 Thesis structure

In thesis, Chapter 2 introduces the background theory used in this thesis. In this thesis, most of the search used the  $\Psi_{\text{mol}}$  method, but there are some  $\Psi_{\text{crys}}$  calculations to compare them and assess if the  $\Psi_{\text{mol}}$  method is good enough. Chapter 3 and Chapter 4 are the CSP searches for the conformers (TMP and CBN), which helped me decide how to predict the co-crystal. In Chapter 3, the TMP experimental structures are found, which improve the confidence in

co-crystal prediction and making sure that the PBE0/6-31G(d,p) in CrystalPredictor<sup>18,19</sup> and CrystalOptimizer<sup>20</sup> is suitable for TMP. Chapter 5 is the co-crystal structure prediction, which is the target XXX in 7<sup>th</sup> Blind Test. Chapter 6 is the analysis of the assumptions made in the Blind Test submission (Chapters 3-5) including the assumptions about the flexibility of the molecules, and whether the search method was appropriate for CBN as in Chapter 4, and why the CSP did not find the CBN experimental structure. The point-charge model used in the structure generation program is not adequate for modelling the conformational changes in CBN hydroxyl group when the hydrogen bonds are formed. A variant on the CrystalPredictor is proposed. It can model the CBN experimental structure, unlike the point charge model.

## 2 Theoretical background to Crystal Structure Prediction (CSP)

### 2.1 Introduction

The crystal is a solid whose constituents are arranged in a highly ordered microscopic structure and extended in all directions. The particles in the internal structure of the crystal are regularly and periodically repeated in the three-dimensional space, forming a lattice. Crystallization or solidification is the processes of crystal formation. Organic crystals are crystals formed by organic molecules, including organic ions.

The symmetry of the crystal is manifested in the shape of the crystal. This depends on its regular lattice structure. The symmetry of the crystal is not only manifested in the external shape, but also its internal structure is also symmetrical. It is the embodiment of its periodic repetition in three-dimensional space.

In the crystal structure<sup>43,44</sup>, the space lattice represents the symmetry related to translation. It can also contain symmetry related to reflection, rotation, and inversion. It can be reflected on the macroscopic level, which is called macroscopic symmetry. The crystal structure must coexist with the spatial lattice and restrain each other. So: (1) Only 1, 2, 3, 4 and 6 symmetry axes can exist in the crystal structure. (2) The spatial lattice can only be of 14 possible types of Bravais lattice. The basic rotation operation of the  $n$ -th symmetry axis is a rotation of  $360^\circ/n$ . (3) There are only 230 space groups for crystals. Space groups are generated from 14 Bravais lattices and 32 crystallographic point groups, each of which belongs to one of 7 lattice systems.

The properties of crystals include<sup>43,44</sup>: 1. Long-range order: the atoms regular arrangement within the crystal at least in the micrometre range. 2. Uniformity: the macroscopic properties of each part inside the crystal are the same. 3. Anisotropy: different directions in a crystal have different physical properties. 4. Symmetry: both the internal structure and the ideal external shape of the crystal have specific symmetry. 5. Self-limiting: crystals have the property of spontaneously forming closed geometric polyhedral. 6. Cleavage: the crystal has the property of splitting along certain oriented crystal planes. 7. Crystal plane

angle conservation: the angle between two corresponding crystal planes belonging to the same crystal is constant.

Crystal defects<sup>43,45</sup> are locations where the structural integrity of a crystal is damaged. There are point defects, line defects and surface defects. In the 0 K, the ideal crystal will not have any crystal defects, but with temperature increase, crystal defects can reduce the free energy through reducing the entropy. Lattice modelling in this thesis assumes a perfect infinite crystal.

#### 2.1.1.1 Definition of crystal internal structure by diffraction

Atoms in the crystal coherently scatter X-rays<sup>46,47</sup>. The total number of scatterings of all electrons in an atom can be considered as scattering from one scattering centre of the atom. These scatterings interfere with each other in space, resulting in a diffraction maximum in a certain direction. Each crystalline substance has its specific crystal structure and lattice parameters, so crystals of two different polymorphs will usually give different X-ray diffraction patterns, that is, give different diffraction beam directions and intensities. The diffraction pattern reflects the lattice structure characteristics of the atomic arrangement in the crystal (such as the coordination of atoms, lattice constants, structural symmetry, crystal internal defects, etc.). Therefore, on the one hand, the crystal can be used as a grating to observe the X-ray diffraction phenomenon; on the other hand, the spatial structure of the crystal can be deduced from the diffraction pattern. This is an important method in modern materials research.

Crystals are solids composed of atoms arranged periodically in three-dimensional space. The periodicity of the spatial distribution of crystalline substances can be represented by the law of spatial lattice node distribution. The nodes of a spatial lattice are abstractions of identical points in a crystal with identical surroundings and identical material content.

Analysis of the spatial lattice model shows that the same node appears repeatedly in three directions in space with periods  $a$ ,  $b$ , and  $c$ , and the entire spatial lattice can be established. The basis vectors of the lattice are the three repeating periods  $a$ ,  $b$ , and  $c$ . The unit cell of the lattice is the parallelepiped formed by the basis vectors. The spatial lattice can be formed by the translation of this parallelepiped in the three fundamental directions.

XRD refers to X-ray diffraction. Bragg's law: X-rays irradiate the particles in the crystal to cause diffraction, and the diffracted beams interfere with each other to produce interference fringes. The generation of interference fringes is related to the crystal structure and the X-ray incident angle. Bragg's law is expressed as:  $2d\sin\theta=n\lambda$  ( $\lambda$  is the wavelength of the incident wave,  $n$  is an integer multiple of the wavelength,  $d$  is the distance between parallel atomic planes, and  $\theta$  is the angle between the crystal plane and the incident light). Thus, it is the fingerprint of crystals. Single crystal diffraction can be used to solve a crystal structure. The hydrogen atomic positions are poorly determined by single crystal X-ray diffraction but can be located by neutron diffraction. The hydrogen atomic positions determined by X-rays can be standardized by elongating the bonds to hydrogen to the lengths given by the neutron method.

Single crystals are sometimes difficult to grow, so powder X-ray diffraction (Powder Diffraction Pattern Cumulative Intensity, pXRD) is used, but this averages over the orientation of the powder crystals. When X-rays are incident to a tiny crystal, the diffraction rays will be diffused and broadened. With the crystal grains smaller and smaller, the X-ray diffraction band is broadened. The relationship between the half-width of the diffraction peak and the grain size is described by Scherrer equation. The output from CSP can help solve a crystal structure measured by pXRD.

#### 2.1.2 The Cambridge Structural Database (CSD)

The Cambridge Structural Database (CSD)<sup>32</sup> was established in 1965. The CSD is the world's database of metal-organic and small-molecule organic crystal structures. CSD is a trusted resource containing over one million structures from neutron diffraction and X-ray analysis. Every entry added to the database is automatically checked and manually managed by CSD experts. Each entry contains a chemical representation, as well as physical and chemical property information, such as colour and morphology that is in the publication.

This rich data resource, along with the visualization, advanced search and analysis software provided by CCDC through CSD-Enterprise<sup>32</sup>, provides tools to leverage this data for further research. CCDC's free access structure service can be used to retrieve and view the individual structures in the CSD. 2D; 3D

datasets, and basic information about structures can be viewed and downloaded using Access Structures.

### 2.1.3 Mercury

Mercury<sup>48</sup> provides comprehensive tools for 3D structure visualization, statistical analysis, and crystal packing exploration of CSD search data. There is a wide range of options to aid in the study of crystal structures, which can be used to: 1. Define and visualize Miller planes, and slice through crystals in any orientation, generate packing diagrams. 2. Build and explore intermolecular contact networks. 3. Show space group symmetry elements. 4. Calculate and display voids (free spaces in the crystal structure) based on solvent accessible surfaces or contact surfaces. 5. Display, for example, the strongest interactions in the crystal structure, assuming an intermolecular potential model.

Sophisticated data analysis tools in Mercury provide statistics, analysis, and 3D structural visualization and charting and plotting options. This integration enables powerful visualization and analysis of ConQuest<sup>49</sup> (Crystal Structure directory tool for CSD) substructure searches, where geometric parameters have been defined in the query. As to statistical tools, there is a range of advanced features specific to structural data, including principal component analysis, cone angle correction in hydrogen bond analysis, and the ability to handle topological symmetries within molecular search fragments. Hence, it is possible to do the fragments analysed for CBN hydrocarbon pentane tail shown in section 4.2.1.

Root Mean Square Deviation (RMSD) is the quadratic mean of differences between the calculated value and real value. In Mercury,  $RMSD_n$  is used to calculate the distances (differences) between corresponding atoms to compare the similarity between different crystal structures and  $n$  is the number of the molecules used to compare.

Hydrogen bond graph set<sup>50</sup> method is based on the graph theory. This method is based on the topological nodes of molecules and lines of hydrogen bonds. It is used to categorize hydrogen bond motifs. This method can code and decode hydrogen bonds systematically and consistently. Using this method, a few parameters can be used to define the hydrogen bond. For example, a hydrogen bond graph set  $G_{a,b,(c)}$   $G$  includes:  $C$  which means chain;  $D$  which means



noncyclic dimers and other finite hydrogen bond sets; *R* which means ring. *a*, *b* are the number of acceptors and donors.

Mercury has several options for display settings. The crystal structures or 3D images can be saved and exported by Mercury in many common formats. Mercury has been used to produce most images in this thesis.

#### 2.1.4 PLATON

The PLATON program<sup>51</sup> is designed to be a multipurpose crystallography tool. It includes tools used as part of single crystal structure determination. The check CIF function in PLATON forms part of the small molecule crystal structure verification tool. PLATON can be used for the detection of possibly missed additional symmetry in each coordinate set. Hence, in this thesis, PLATON is used to find more symmetries in the crystal and check the true symmetry of CSP generated structures. Furthermore, PLATON can be used to calculate the voids and packing coefficients which is shown in section 2.1.5.

#### 2.1.5 Packing Coefficients

In crystallography, the atomic packing factor (APF)<sup>43,52</sup>, (or packing coefficient) is the volume fraction of a crystal structure occupied by constituent particles. The fill fraction is mathematically expressed as:

$$APF = \frac{V_{particle}}{V_{unit\ cell}}$$

where  $V_{particle}$  is the volume of particles in a unit cell, and  $V_{unit\ cell}$  is the volume occupied by a unit cell. It is a dimensionless quantity, and it is always less than one.

Mathematically it can be proved that the APF of the densest atomic arrangement is about 0.74 for the one-component hard-sphere structure. For multi-component structures such as organic crystals, the APF can exceed 0.74 but it is usually between 0.65 and 0.70.

Both CCDC<sup>32</sup> software and PLATON<sup>51</sup> use the same method to calculate the fill factor. The centre position of each atom, as well as the kind of atom, is known, and a model of the molecule can be generated.

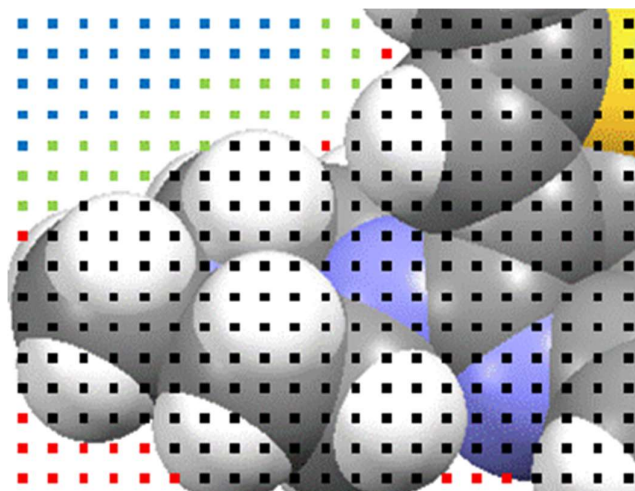


Figure 2 The grid used in packing coefficient calculation and voids space calculation. Black points are the crystal components molecules occupied space. Blue point is any point that is more than a certain distance (usually it is the water molecular radius) from the edges of the molecules, and green points are within this probe radius distance of such a point included too.<sup>52</sup>

Shown in Figure 2, to calculate the packing coefficient, firstly, a grid is generated with a spacing that is defined in the CCDC software or 0.2 Å in PLATON<sup>51</sup>. There is a distinction between points inside and outside the molecular sphere. The percentage of intramolecular dots (black points) is the packing coefficient.

PLATON includes additional steps where any points that are more than a certain distance (usually it is the water molecular radius) from the edge of the crystal component molecule (blue points), and those that are within the detection radius distance of that point are also included (green), shown in the Figure 2. The blue and green points will include the limits of any solvent molecule that might fit.<sup>53</sup> In other words black points include the crystal components (solute); the blue and green points will include the limits of any solvent molecule that might fit; the red points do not include any solute or solvent molecules. Counting the number of blue points can be used to calculate the voids, which means how much other molecules (usually it is water molecules) can be put in this crystal. Voids calculation is important for crystals. For example, there are sometimes some channels in the crystal. In this case, the crystal formed with the solvent in the channel will be more stable than the pure crystal.

The atomic radius is calculated by Gavezzotti<sup>54</sup> for the Mercury software (Mercury is introduced in 2.1.3), and the atomic radius is calculated by Bondi<sup>55</sup> for PLATON (section2.1.4).

### 2.1.6 Lattice energy and temperature effects

The lattice energy is defined as the energy change when an infinite static perfect lattice of the molecules is separated into infinitely separated molecules in the lowest energy conformation.  $E_{latt}=U_{inter}+\Delta E_{intra}$  where the  $E_{latt}$  is the lattice energy,  $U_{inter}$  is the intermolecular interaction energy and the  $\Delta E_{intra}$  is the intramolecular energy relative to the gas phase. If the molecular conformation in the crystal phase is assumed to be the same as the gas phase of the isolated molecule ( $\Delta E_{intra}=0$ ), then the lattice energy can be the sum of all the intermolecular interactions in the lattice.

Free energy can be converted into external work during a certain thermodynamic process. Free energy is the reduced system–internal energy. It measures the “useful energy” that the system can output to the outside during a specific thermodynamic process. The Helmholtz free energy is defined for constant volume and the Gibbs free energy for constant pressure.

Helmholtz free energy:  $F=U-TS$  and Gibbs free energy:  $G =U-TS + pV =H-TS$

$S$  is the entropy;  $T$  is the temperature and  $U$ , in the case of crystals is the lattice energy.  $H$  (enthalpy) is equal to  $U+pV$ .

### 2.1.7 Co-crystal and co-crystallization energy

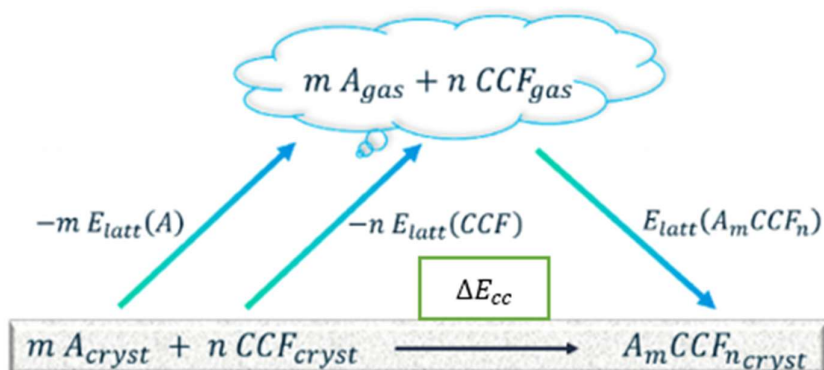


Figure 3 The thermodynamic cycle used to calculate co-crystal formation and co-crystallization energy, which is adapted from reference<sup>24</sup>.  $E_{latt}$  means the lattice energy.  $\Delta E_{cc}$  means the co-crystallization energy.

In crystal structure prediction, the lattice energy of the co-crystal should be lower than the pure component crystal structure for the co-crystal to exist. In other

words, the co-crystal should be more stable than the pure component's crystal structure. The co-crystallization growth process can be considered equivalent to the sublimation thermodynamic cycle<sup>56</sup> shown in Figure 3, for a co-crystal  $mA:nCCF$ , where  $A$  is the drug;  $CCF$  is the co-former;  $m:n$  is the stoichiometry.

As shown in Figure 3, the co-crystal forming process and co-crystallization energy calculation can be divided into two processes: 1. Co-formers sublime to the gas phase. 2. Gas phase co-formers form the co-crystal together. The thermodynamic method being used to estimate the thermodynamic driving force for co-crystallization is illustrated in Figure 3, which by approximating the free energies by lattice energies, gives that, the co-crystallization lattice energy can be calculated by:

$$\text{Equation 1 } \Delta E_{cc} = \frac{E_{latt}(A_m CCF_n) - nE_{latt}(CCF)}{m} - E_{latt}(A)$$

Where the  $A$  is the API and  $CCF$  means the co-former of the co-crystal.

Whether the co-crystal is formed can be justified by whether the co-crystallization lattice energy,  $\Delta E_{cc}$ , is lower than zero. The lower the co-crystallization energy, the easier the co-crystal can be formed. In a study<sup>24</sup>, although co-crystallization energies of some co-crystals are 8 kJ/mol when they use CSP results for co-crystal and drugs, co-crystals are still formed. As to the co-formers, there is not too much difference between the CSP co-former results and experimental results. But most co-crystallization energies are below 0.

## 2.2 Quantum chemistry

Quantum chemistry<sup>57,58</sup> is a basic science. It applies the basic principles and methods of quantum mechanics to solve chemical problems. The main methods of quantum chemistry used in this project are molecular orbital (MO) based methods and Density Functional Theory (DFT).

Ab initio is a first-principles calculation in a narrow sense. It refers to quantum calculations that do not use any empirical parameters and only use fundamental constants such as Planck constant, elementary charge and so on. But this type of calculation is very slow, so there are some semi-empirical methods where some empirical parameters are used to greatly speed up the calculation, but of course, it will inevitably sacrifice the accuracy of the calculation results.

All approaches use Born–Oppenheimer approximation (BO approximation)<sup>58</sup> i.e. assume that, as to the molecule, the energy and wavefunction of nuclei and electrons can be calculated separately, because the nuclei are heavier than the electron too much.

#### 2.2.1 Wavefunction based methods (HF, MP2)

The molecular orbital method is the generalization of atomic orbitals to molecules. In other words, it is assumed that in the molecule, each electron moves in the average potential field. A single electron function (the electron's coordinate function) can be used to represent each electron motion state, which is the single–electron function molecular orbital. The wavefunction of the entire molecule is composed of molecular orbitals of all electrons in the molecule.

The Hartree–Fock equation (HF equation)<sup>58</sup>, is an equation that calculates the wave function of the multi–electron system. Hartree–Fock approximation<sup>58</sup>, also known as a molecular orbital approximation or single determinant approximation, considers that the single Slater determinant, which is constructed by the molecular orbital wavefunction, can be used to represent the wave function of a multi–electron system. The HF equation is a one–electron eigen equation in form. The obtained eigenstate is a one–electron wave function, that is, a molecular orbital.

Møller–Plesset perturbation theory<sup>59,60</sup> is based on molecular orbital theory and it is an advanced quantum chemical calculation method. Møller–Plesset perturbation theory is based on the self–consistent field solution of the Hartree–Fock equation, and the approximate solution of the multi–electron system considering the correlation energy is obtained by applying the perturbation theory. MP2 (Møller–Plesset perturbation theory with 2<sup>nd</sup> correction) is the many–body perturbation theory with secondary correction. MP2 corrects HF for electron correlation. The electron correlation effects include the intermolecular dispersion term.

#### 2.2.2 Density Functional Theory (DFT)

Density functional theory<sup>61,62</sup> is used to study the electronic structure in multielectron systems. There are some classical methods of electronic structure theory including Hartree–Fock method and MP2 (post–Hartree–Fock methods)

and so on. Complex multi-electron wavefunctions are the basement for these classical methods of electronic structure theory. DFT theory uses electron density as the fundamental quantity of research to replace the wavefunction. The electron density is a function only relevant to three variables but there are  $3N$  variables ( $N$  is the number of electrons) for the multi-electron wavefunction. Thus, it is conceptually and practically more convenient to deal with multi-electron systems.

Local density approximation (LDA)<sup>61,62</sup> is an approximation that is used to calculate the contribution of the exchange-correlation (XC) energy functional in DFT. It starts from the non-interacting uniform electron gas (or non-interacting homogeneous electron gas), which can be used to calculate the exchange energy of the system, and the relevant energy contribution is processed by fitting the free electron gas. This method is not suitable for molecules and organic crystals as the electron density vary considerably. More suitable methods of modelling the exchange energy are described in 2.2.2.1.

DFT concept originated from the Thomas-Fermi model, but until the Hohenberg-Kohn theorem<sup>63,64</sup> using the inhomogeneous electron gas model, there was no solid theoretical basis. In the first Hohenberg-Kohn theorem, the ground state energy is only an electron density functional. As to the second Hohenberg-Kohn theorem, minimizing the system energy with the ground state density as the variable is used to calculate the ground state energy.

The original theory only applies to ground states where no magnetic field exists, although it has now been generalized<sup>63,64</sup>. It is in these exact correspondences that there are approximations (this theory can be generalized to the time-dependent domain, and thus used to calculate the properties of excited states).

#### *2.2.2.1 Functionals (PBE and PBE0) commonly used in organic solid*

There is an extension for LDA treating different electrons based on their spin projections, which is named local spin density approximation (LSDA). LDA is based on the non-interacting homogeneous electron gas. If the inhomogeneity of the charge distribution is considered, especially in some systems with localized electrons, the gradient of the charge density should be calculated, which is

Generalized Gradient Approximation (GGA). To the exchange–correlation energy, GGA is used to improve the LSDA.

Perdew, Burke and Ernzerhof proposed a simple GGA where all parameters (except those in LSD) are fundamental constants<sup>65</sup>. There is a linear response of the uniform electron gas. It includes the correct behaviour at uniform scaling and smoother potentials. This functional is named PBE (Perdew–Burke–Ernzerhof)<sup>65</sup>.

PBE0<sup>66</sup> is the hybrid functional that mixes the Hartree–Fock exchange energy and Perdew–Burke–Ernzerhof (PBE) exchange energy in a set 1:3 ratio, along with the full PBE correlation energy, which is expressed as:

$$E_{XC}^{PBE0} = \frac{1}{4}E_X^{HF} + \frac{3}{4}E_X^{PBE} + E_C^{PBE}$$

### 2.2.3 Semi-empirical periodic methods e.g., Density Functional Tight-Binding (DFTB)

DFTB<sup>67</sup> can be extended to Density Functional Tight-Binding which is the tight-binding model based on density functional. It is an approximate model for periodic solid systems. This model uses "repulsive potential + orbital energy" to represent total energy. It obtains empirical parameters by fitting advanced calculation results. This model is very timesaving for molecular system calculations, but there are too large errors for the heat of formation, bond length, and other properties. Further increasing the parametric potential function can be used to improve it. SCC-DFTB is a self-consistent charge based on DFTB. SCC-DFTB adds a polynomial to the energy expression of the original DFTB. In this model, Taylor's expansion of the atomic charge is used to correct the error on the correlation effects and exchange terms.

LCAO (Linear-combination-of-atomic-orbitals) is the basement of the DFTB. It is that every one-electron wavefunction  $\phi_i$  can be expressed as a linear combination of atomic orbitals  $\eta_\mu$  with linear parameters  $c_\mu^i$ ;  $\phi_i = \sum_\mu c_\mu^i \eta_\mu$ . It is also the basement of DFTB with the formalism uses the local-density approximation. The model charge density and potential<sup>68</sup> can be used to calculate Hamiltonian and overlap matrix elements in DFTB method. Initial charge density is constructed from the library of atomic DFT densities. Then it is used in DFTB method which is the zero-order calculation of spheres. Parametrization includes

the repulsive parameters (fitted to atomization energy) and electronic parameters (angular momentum–spin–polarization)<sup>69</sup>.

The calculation of the Hamilton in DFT will require a lot of computational resources. But using LCAO approximation, let the eigenfunctions of Hamilton assumed as the LCAO basis set, the DFTB Hamilton matrix elements can be calculated, which is the non–SCC (self–consistent charge) DFTB Hamiltonian. Therefore, the main computational cost of DFTB is the solution of the diagonalization of the Hamilton matrix. Furthermore, saving the atoms DFT results as the parameters can save more time.

There is no spin contribution and SCC contribution in the non–SCC<sup>70</sup> (self–consistent charge) DFTB Hamiltonian (zeroth–order). When the ground–state charge density is close to the reference charge density  $\rho_0$ , because of the LCAO approximation, the non–SCC DFTB will not lead to much error. There is no charge flow among atoms in non–SCC DFTB method, which will lead to the solution being different from the atomic–like densities assumed. The molecular charge density will be also changed. These errors need to be corrected.

SCC<sup>70</sup> uses the spherical neutral atoms as the start charge density. Then the monopole approximation and spherical neutral atoms will be used to calculate the first order and second–order terms. The high multipole terms will be the first term or ignored. SCC–DFTB correction extends the energy depending on the charge density change. Then the non–SCC result is corrected by second order (DFTB2), third order (DFTB3) correction. 2<sup>nd</sup> correction is the chemical hardness, Mulliken charges, and charge fluctuation. 3<sup>rd</sup> correction is the change in chemical hardness.

SCC–DFTB treats the input charge density fluctuations using the perturbation method. it is important when there are so many charges for atoms. The second critical approximation includes the second interaction among the electrons. However, it is not suitable for all situations. Then there is no extra change requirement for the third critical approximation term and higher terms, which are not necessary for many systems. Many organic and biological systems require the second approximation perturbation theory. Because of the self–consistent



solution of the eigenvalue, SCC method requires more time than non-SCC method. Furthermore, parametrizations are required in SCC-DFTB<sup>71</sup>.

There are some semi-empirical methods to save the calculation cost. 1) SCC calculation requires the integral among atomic wavefunctions in atom pairs, which can be saved as the data source. 2) The valence electrons are calculated by the smallest basis set and then corrected.

Generally, linear-combination-of-atomic-orbitals (LCAO) assumption can let the DFT Hamilton solution chapter but not accurately, furthermore, saving the atoms DFT results as the parameters can save more time. But obviously, it is not self-consistent, adding the SCC correction can improve the accuracy of DFTB results. DFTB3 is the SCC-DFTB method with the third order correction, which improves proton affinity, hydrogen transfer barrier, and hydrogen binding energy<sup>69</sup>.

#### 2.2.4 Basis sets for molecules and plane waves (with pseudopotential for periodic)

In quantum chemistry, a basis set is a set of mathematical functions used to describe the wave function of a system<sup>57</sup>. Gaussian-Type Orbital (GTO) refers to using a Gaussian-type function to simulate atomic orbitals. For example, as to 6-31G, a linear combination of 6 Gaussian functions forms each inner electron orbital, and each valence electron orbital is split into two basis functions, each consisting of 3 and 1 linear combination of a Gaussian function. The 6-31G(d) basis set adds a *d* polarization function for heavy atoms (i.e. not hydrogen or helium) on the basis of split-valence basis set 6-31G, (6-31G(d) is equivalent to 6-31G\*). The 6-31G(d,p) basis set adds a *d* polarization function for heavy atoms on the basis of split-valence basis set 6-31G, and adds a *p* polarization function for light atoms (hydrogen and helium atoms) (6-31G(d,p) is equivalent to 6-31G\*\*). 6-31+G(d,p) is equivalent to adding a diffuse function for heavy atoms to the basis set 6-31G(d,p). 6-31++G(d,p) is equivalent to adding a diffuse function for heavy atoms and light atoms (helium and hydrogen atoms) to the basis set 6-31G(d,p).

#### 2.2.5 Dispersion term

Due to the constant movement of nuclei and electrons in the molecule, the electron distribution of the molecule will always be changing. There is an

instantaneous relative displacement between the nucleus and electrons. A multipole generated through the displacement has interactions with other instantaneous multipoles<sup>72</sup>. This is the origin of the dispersion energy, a non-classical stabilising interaction. The DFT and DFTB<sup>67</sup> functionals used in periodic crystal calculations do not include the dispersion effects which operate at long range because of the approximate treatment of electron correlation. DFT methods could only be used for organic crystals once dispersion corrections were introduced by Neumann and Perrin<sup>73</sup>, as part of their development of CSP methods. In this study<sup>73</sup>, a parameterized hybrid approach provides high precision for optimizing structure and ordering energy of molecular crystals by combining empirical van der Waals corrections and high-level DFT. The difference between calculated and experimental unit cells is about 1% on average.

Tkatchenko and Scheffler (TS)<sup>74</sup> is a method used to describe an accurate long-range van der Waals interactions determination from the mean-field electronic structure.

$$E_{vdW} = -\frac{1}{2} \sum_{A,B} f_{damp}(R_{AB}, R_A^0, R_B^0) C_{6AB} R_{AB}^{-6}$$

Where  $R_A^0$  and  $R_B^0$  are the vdW radii;  $R_{AB}$  is the distance between atoms  $A$  and  $B$ ; and  $C_{6A}$  is the corresponding  $C_6$  coefficient. The  $R_{AB}^{-6}$  singularity at small distances is eliminated by the short-ranged damping function  $f_{damp}(R_{AB}, R_A^0, R_B^0)$ .

$$C_6^{ab} = \frac{2C_{6AA}C_{6B}}{\left[\frac{\alpha_B^0}{\alpha_A^0}C_{6A} + \frac{\alpha_A^0}{\alpha_B^0}C_{6BB}\right]}$$

$$f_{damp}(R_{ab}, R_{AB}^0) = \frac{1}{1 + \exp\left[-d\left(\frac{R_{ab}}{s_R R_{AB}^0} - 1\right)\right]}$$

Where  $d$  and  $s_R$  are free parameters and  $R_{AB}^0 = R_A^0 + R_B^0$ . The  $d$  parameter adjusts the damping function steepness.

All dispersion terms in D2 and TS are displayed as  $\frac{C_6^{ab}}{r_{ab}^6} f_{d,n}(r_{ab})$  term with the different  $f_{d,n}(r_{ab})$  for TS and D2,

$$f_{d,n}(r_{ab}) = \frac{1}{1 + e^{-d(\frac{r_{ab}}{r} - 1)}}$$

Where  $r$  is the sum of atomic vdW radii, and  $d$  is the experimental parameter.

D3 dispersion correction sums two- ( $E^{(2)}$ ) and three-body ( $E^{(3)}$ ) dispersion energies.

$$E^{disp} = E^{(2)} + E^{(3)}$$

There is a damping function with the damping cut-off<sup>72</sup> to avoid double-counting effects of correlation at the intermediate distance and near singularities for small  $r_{ab}$ .

In D3 the two-body term is shown as<sup>75</sup>:

$$E^{(2)} = \sum_{ab} \sum_{n=6,8,10,\dots} s_n \frac{C_n^{ab}}{r_{ab}^n} f_{d,n}(r_{ab})$$

Where

$$C_6^{ab} = \frac{3}{\pi} \int_0^\infty a^a(i\omega) a^b(i\omega) d\omega$$

$$C_8^{ab} = 3C_6^{ab} \sqrt{Q^a Q^b} \quad C_{n+4} = C_{n+2} \left(\frac{C_{n+2}}{C_n}\right)^3$$

The  $C_n^{ab}$  denotes the averaged (isotropic) 6th,8th,10th, .... dispersion coefficient for atoms  $A$  and  $B$ ;  $\sum_{ab}$  means over all atom pairs in the system, and  $r_{ab}$  is their internuclear distance. Global scaling factors  $s_n$  are free fit parameters and adjusted only for  $n > 6$  to ensure asymptotic exactness<sup>72</sup>.  $Q^a = s_{42} \sqrt{Z^a} \frac{\langle r^4 \rangle^a}{\langle r^2 \rangle^a}$  where heavier elements  $\sqrt{Z^a}$  is necessary to get consistent interaction energies and  $\langle r^4 \rangle$  and  $\langle r^2 \rangle$  are simple multipole-type expectation values which are from atomic densities;  $s_6 = 1$  and  $n = 6,8$ . The Becke–Johnson–damping function used in this report is:

$$f_{d,n}(r_{ab}) = \frac{r_{ab}^n}{r_{ab}^n + (\alpha_1 r_{0,ab} + \alpha_2)^n}$$

$\alpha_1, \alpha_2$  are the free fit parameters.

D2 dispersion does not have the three-body term dispersion correction. Three-body term dispersion correction of D3(BJ) for the atoms  $a$   $b$   $c$  can be expressed as<sup>72</sup>:

$$E^{(3)} = \frac{C_9^{abc}(3\cos\theta_a\cos\theta_b\cos\theta_c+1)}{(r_{ab}r_{bc}r_{ca})^3}$$

$$C_9^{abc} = \frac{3}{\pi} \int_0^\infty a^a(i\omega)a^b(i\omega)a^c(i\omega)d\omega$$

Because of the small three-body term dispersion correction,  $C_9^{abc}$  can be approximated to:  $C_9^{abc} = -\sqrt{C_6^{ab}C_6^{ac}C_6^{bc}}$

In general, two-body term in D3 dispersion is different from D2 and TS dispersion. Furthermore, in D3 there is the three-body term correction that is usually used in the big system.

## 2.3 Electronic structure methods and codes used in this work

### 2.3.1 Periodic electronic structure ( $\Psi_{\text{cry}}$ ) methods

Periodic electronic structure methods were introduced in the 4<sup>th</sup> Blind Test, by Neumann and Perrin. As each crystal structure is optimized at the electronic level, this approach is termed  $\Psi_{\text{cry}}$ . Such calculations are very expensive with computer resources. They have the advantage that the optimization allows the lattice parameters and positions of all atoms in the asymmetric unit cell to vary to find the minimum in the lattice energy, this crystal structure relaxed all atoms.

#### 2.3.1.1 CASTEP Plane waves, pseudopotential

CASTEP<sup>76</sup> is based on first-principles quantum mechanical descriptions for nuclei and electrons. It is a fully functional materials modeling code. It uses a robust method of plane wave basis sets and pseudopotentials. The code is developed by the CASTEP Developers Group of all UK academics. In this thesis, CASTEP version 21.1.1 is used.

In the plane wave method, the energy eigenstates of a single electron are calculated from the Schrödinger equation in a periodic potential. Pseudopotentials are needed for plane waves. CASTEP uses pseudopotentials to represent interactions between core and valence electrons. CASTEP supports norm conservation and Ultrasoft pseudopotentials. Pseudopotentials can be read

from files in various formats. CASTEP also has its own built-in generator that calculates the potential "on the fly" during the calculation process. There is a built-in database of well-tested potential. In particular, it is possible to generate high-precision semi-core pseudopotentials that have been used in high-pressure studies.

#### *2.3.1.2 CRYSTAL14 and DFTB+*

The high cost of CASTEP ab initio periodic density functional methods means that it is very costly to apply such calculations on hundreds of crystal structures, as needed in the early stages of a CSP study. Semi-empirical periodic electronic structure codes are much cheaper and can be used to optimise large numbers of structures.

A new Fortran95 implementation of the DFTB<sup>67</sup> method has been developed, which exploits the sparsity of the DFTB system of equations<sup>77</sup>. Stored matrices are true for both periodic and aperiodic computations. Many operations during DFTB computation can be evaluated directly in this sparse format, resulting in significant speed improvements. Furthermore, there are many operations on aperiodic and periodic systems to be treated the same, so most routines do not need to consider the boundary conditions of the system.

DFTB+<sup>77</sup> is an open-source software package that provides a fast and efficient way to perform atomic quantum mechanical simulations. It is capable of simulating large systems and long-time scales reasonably by implementing various approaches to approximate DFT, such as extended tight-binding methods and DFTB. It is much faster in terms of the corresponding ab initio method. It also provides approximate versions of various DFT extensions based on the DFTB framework, including electron transport using non-equilibrium Green's functions, time-dependent forms for handling excited systems, hybrid functionals, and so on. DFTB+ can also be used as a stand-alone application and it can be embedded in other software packages as a library.

DFTB+ is an atomic quantum mechanics simulation package that allows long-term fast and efficient simulation of large systems. It can be used either as a library or as a standalone application and has been linked to several other emulation packages.

Crystal14<sup>78</sup> is an ab initio code using Gaussian basis sets: that allows the use of pseudopotentials and all-electron strategies. A variety of density functionals are available. A mixture of various properties can be used (global, range-separated, dual). In particular, very efficient implementations of hybrid DFT functionals allow performing such computations at relatively low computational costs. The program can handle zero-, one-, two-, and three-dimensional systems (molecules, polymers, plates, crystals) on the same basis. When plane waves are used as basis sets, spurious 3D periodicity is not required for low-dimensional systems. Crystal14 can be used to perform the geometry steps and call the DFTB+ code.

### 2.3.2 Molecular ( $\Psi_{\text{mol}}$ ) method of modelling organic crystal structures

The CSP modeling approach based on the molecular electronic structure is called a ( $\Psi_{\text{mol}}$ ) method. It is an approach that originated for CSP studies on rigid molecules, where the molecular conformation and charge in the crystal is that of the isolated molecule in the gas phase to a good approximation.  $\Psi_{\text{mol}}$  approach uses the molecular wave function calculation results as the input for atomistic crystal modeling codes such as DMACRYS<sup>39</sup>.

The thermodynamic stability structure is determined by the lattice energy calculation. The lattice energy can be calculated by the sum of the intermolecular energy  $U_{\text{inter}}$  and the energy of the molecular conformation relevant to the most stable conformation (gas phase conformation)  $\Delta E_{\text{intra}}$ . The charge density of single molecules is calculated by GAUSSIAN<sup>79</sup>. The charge density is subjected to a distributed multipole analysis (DMA) using GDMA<sup>80,81</sup> to obtain a representation in terms of atomic multipoles. Then, the intermolecular energy can be calculated by DMACRYS<sup>39</sup> using the multipoles and a repulsion–dispersion potential. As the atomic dipoles and quadrupoles are important for modelling the directionality of  $\pi\dots\pi$  interactions and hydrogen bonding, CSP studies using distributed multipoles are more successful than those using just atomic charges<sup>82</sup>.

#### 2.3.2.1 Codes used in modeling organic crystals by GAUSSIAN, GDMA & point charge models

GAUSSIAN<sup>79</sup> is a powerful comprehensive package for molecular quantum chemistry. The executable program of GAUSSIAN can run on different models of personal computers and supercomputers and has different versions accordingly.

GAUSSIAN can be used to calculate the molecular charge density using DFT and wavefunction methods.

For  $\Psi_{\text{mol}}$  atomistic modelling of organic crystal, DMACRYS<sup>39</sup> is started from the code NEIGHCRYST<sup>88</sup> to form the input file. NEIGHCRYST extracts the molecule conformation from crystal defined by cell parameters, symmetry operations, and fractional atomic coordinates. In this thesis, DMA<sup>80,81</sup> (distributed multipole analysis) is calculated by GDMA program<sup>80,81</sup>. GDMA calculate DMA starting from the charge density calculation, and then the DMA will be calculated following the coordinate definition (it has been proved that different direction of the DMA will not lead to much error in lattice energy calculation). Originally, the distributed multipoles were calculated using the Gaussian that corresponds to every pair of Gaussian atomic orbitals in the basis set. However, for basis sets that include diffuse functions, a method that partitions the charge density into atomic densities and then integrates converges better.<sup>80,81</sup>

The best point atomic charge models<sup>18,19</sup> of a molecular charge are those calculated by fitting the atomic charges to the electrostatic potential on a grid of points outside the molecule in the region where other molecules could be. Reference the CHELG method in Gaussian that is used to provide the atomic charges in CrystalPredictor<sup>18,19</sup>.

#### 2.3.2.1.1 Intermolecular lattice energy calculation in DMACRYS

The lattice energy is calculated by the sum of intermolecular energies for a rigid molecule and intramolecular energies relevant to the gas phase conformation  $\Delta E_{\text{intra}}$ , if the conformation in the gas phase is in crystal.

DMACRYS<sup>39</sup> performs the lattice sums over the electrostatic interaction energy<sup>83,84</sup> using

$$\text{Equation 2} \quad U_{\text{electrostatic}} = \frac{1}{2} \sum_m \sum_{n \neq m} Q_t^a T_{tu}^{ab} Q_u^b$$

$Q_t^a$  refers to the multipoles  $t$  on the atomic site  $a$  of molecule  $m$ ;  $Q_u^b$  refers to the multipoles  $u$  on the atomic site  $b$  of molecule  $n$ ;  $T_{tu}^{ab}$  is the interaction matrix, and the non-electrostatic interaction<sup>39</sup> is:

$$\text{Equation 3} \quad U_{\text{rep-disp}}^{mn} = \sum_{a \in m, b \in n} \sqrt{A_{\tau\tau} A_{\kappa\kappa}} \exp\left(-\left(\frac{B_{\tau\tau} + B_{\kappa\kappa}}{2}\right) R_{ab}\right) - \frac{\sqrt{C_{\tau\tau} C_{\kappa\kappa}}}{R_{ab}^6}$$

where  $R_{ab}$  refers to the distance between atom  $a$  and atom  $b$ ;  $A, B, C$  are parameters for atom  $a, b$  of type  $\tau, \kappa$  in molecule  $m, n$ .  $\sqrt{A_{ii}A_{kk}} \exp\left(-\left(\frac{B_{ii}+B_{kk}}{2}\right)R_{ab}\right)$  is the repulsive term and the  $\frac{\sqrt{C_{ii}C_{kk}}}{R_{ab}^6}$  is the dispersion term. The repulsion–dispersion parameters in Equation 3 are those of the FIT<sup>85,86,87</sup> or W99<sup>38</sup> potential.

FIT potential<sup>85,86,87</sup> has evolved through using Williams parameterisations<sup>38</sup>. It had each element in conjunction with carbon and hydrogen only. Furthermore, Coombes<sup>86</sup> and Beyer<sup>87</sup> develop the FIT potential to make it adequate for hydrogen bonds forming with different hydrogen (N–H<sup>86</sup> and O–H<sup>87</sup>). The hydrogen nuclei are at the interaction sites. Both empirical FIT potentials effectively model the total intermolecular potential by removing the electrostatic component and sampling it in the crystal structure used for verification and fitting. The results can be very poor for atypical short contacts and can be sensitive to the electrostatic model. Because they are based on the experiment, the choice of which potential and parameters to use should be made by empirical testing of the relevant crystal structures.

#### 2.4 Crystal Structure Prediction Two steps method (General methodology)

Since the lattice energy surface shows many local minima, CSP is often very sensitive to the lattice energy calculation accuracy. Therefore, the CSP method is divided into two steps: (1) Global search phase (Structure generation). This requires generating plausible crystal structures usually involving a relatively crude lattice energy optimisation. (2) The structure optimization stage (Energy evaluation and ranking), at this stage, the more computationally expensive but more accurate lattice energy models are used to correctly identify which of the generated structures are lowest in energy. This can be hierarchical, involving increasingly more accurate calculations on smaller numbers of crystal structures at each step.

The structure generation step is critical to ensure that there is none of the low energy and experimental structures are missed in the first step. It is also vital that none of the lattice energy minimisation models are so crude that the important



structures have a high energy above the cut-off used for going on to the next stage. The first stage usually needs to generate one hundred thousand to one million or more crystal structures. The more accurate the calculation, the more expensive computer time is. So, the balance between accuracy and cost should be considered. At the same time, although the first step will generate one hundred thousand or one million structures, many of these will be duplicates when relaxed with a crude energy model, so they can be clustered by comparing the lattice structures and removing duplicates.

#### 2.4.1 Global search phase (Structure generation)

In the 5<sup>th</sup> Blind Test<sup>41</sup>, the successful prediction team adopted these two steps. In the structure generation step, GRACE, UPACK, and CrystalPredictor<sup>18,19</sup> are used. In addition, MOLPAK, and methods based on genetic algorithms can also be used for structure generation. In the 6<sup>th</sup> Blind Test<sup>37</sup>, MC simulated annealing CSD analogues, quasi-random search. Grid search and random search, random generation, genetic algorithm, MC parallel tempering, and evolutionary algorithm, were all used to generate a wide range of crystal structures.

In this project, the CrystalPredictor program is used to generate structures, which will be introduced in the next section.

##### 2.4.1.1 CrystalPredictor

The CrystalPredictor program<sup>18,19</sup> is based on using crystallographic symmetry. It is only necessary to generate the asymmetry unit (which may contain more than one molecule) and use the space group symmetry to produce possible structures. Analysis of organic crystal structures in the CSD<sup>32</sup> shows that most organic structures are found in just 61 space groups. Thus, in this thesis crystal structures are only generated in these 61 space groups, weighted by the frequency of observation in the CSD (unless otherwise specified).

(Most probable 61 space groups for organic crystals used in this thesis CSP: *P1 P-1 P2<sub>1</sub> P2<sub>1</sub>/c P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> Pna2<sub>1</sub> Pbca C2/c Cc Pnma P2<sub>1</sub>2<sub>1</sub>2 Pca2<sub>1</sub> Pbcn C2 Pc Cm P2<sub>1</sub>/m C2/m P2/c C222<sub>1</sub> Pmn2<sub>1</sub> Cmc2<sub>1</sub> Aba2 Fdd2 Iba2 Pnna Pccn Pbcm Pnnm Pmmn Cmcm Cmca Fddd Ibam P4<sub>1</sub> P4<sub>3</sub> I-4 P4/n P4<sub>2</sub>/n I4/m I4<sub>1</sub>/a P4<sub>1</sub>2<sub>1</sub>2 P4<sub>3</sub>2<sub>1</sub>2 P-42<sub>1</sub>c I-42d P3<sub>1</sub> P3<sub>2</sub> R3 P-3 R-3 P3<sub>1</sub>21 P3<sub>2</sub>21 R3c R-3c P6<sub>1</sub> P6<sub>3</sub> P6<sub>3</sub>/m P2<sub>1</sub>3 Pa-3 P222<sub>1</sub> Pba2)*

For flexible molecules, the range of possible values of each of the  $N$  flexible torsion angles is chosen, and a  $N$  dimensional grid of molecular conformations is set up. At each point on the grid, an ab initio calculation is run to generate a LAM (local approximate model) point<sup>18</sup>. These LAMs make up the database that describes the intramolecular energy surface  $\Delta E_{intra}$  and forces. Additional intermediate points can be calculated (where the energy difference is above a certain threshold) and make the database non-uniform. At each LAM point, the atomic point charges for the conformation are also stored.

CrystalPredictor program uses random numbers generated by low-discrepancy Sobol sequences to generate crystal structures. The numbers generated are used to determine: the space group and the allowed lattice parameters, the values of the flexible torsion angles, position and orientation of each molecule. If the first search is incomplete and it is necessary to continue the search based on the first search, the Sobol sequence can be continued.

Density is used within Crystal Predictor to remove unfeasible structures, with two methods. In Chapter 3,4,5, the upper limit of density was set to 300 kg/m<sup>3</sup> and any generated structures that were denser than this were rejected, as the molecules may be overlapping. This approach had a problem when the specified ranges of lattice parameters are not suitable for some of the higher symmetry space groups, leading to too many structures being rejected. An improved method within CrystalPredictor, used in Chapter 6, uses a density range. If the generated crystal structures have a density below the lower bound, the lattice lengths are decreased until this limit is reached. Conversely, structures with a density above the upper bound are expanded.

In the CrystalPredictor program the next step is called MINIMISE, which crudely optimises the structure to a minimum in the total lattice energy. The FIT potential<sup>85,86,87</sup> is used for the repulsion dispersion, and the point charges from the weighted-average LAM points<sup>19</sup> in the database to model the electrostatic in the intermolecular lattice energy. The intramolecular energy contribution is calculated by weighted-average LAM points<sup>19</sup> in the database, this method is not accurate enough but can screen out the impossible structure with high energy to reduce the output crystal structure candidates effectively. It could be seen as the

first step in the hierarchy of optimization lattice energy optimisations. After this optimisation, Crystal Predictor has a step called ANALYSE which removes duplicate structures based on the interatomic distance matrix for each structure, thereby removing both duplicates and structures with alternative lattice settings. ANALYSE also can be used to check if the search is completed through how many times the global minimum structure has been found.

#### 2.4.2 Structure optimization stage (Energy evaluation and ranking)

DMACRYS<sup>39</sup> is accurate but it is used for rigid molecules shown in section 2.3.2.1.1, In order to model the molecular conformation in the crystal affected by the intermolecular forces, it needs to be combined with the intramolecular energy calculation model to calculate lattice energy and optimize the crystal structures. CrystalOptimizer<sup>20</sup> is used in this step in this thesis, which combines GAUSSIAN<sup>79</sup>, GDMA<sup>80,81</sup>, DFTB<sup>67</sup>, NEIGHCRY<sup>88</sup> and DMACRYS codes.

##### 2.4.2.1 CrystalOptimizer

For flexible molecules, the intramolecular energy  $\Delta E_{intra}$  and the atomic charge density will change when the molecular conformation changes. In order to reduce the load of the computer and save calculating resources, the choice is made as to which molecular bond angles and torsion angles are likely to be affected. The intermolecular lattice energy is particularly sensitive to the molecular torsion angles that define the position of the protons involved in hydrogen bonding such as NH<sub>2</sub> and OH<sup>89</sup>.

Thus, when the charge density is calculated the first time, the distributed multipoles of each atom will be calculated. Each atom is assigned a local axis system using two directly connected atoms once the distributed multipoles of each atom have been calculated for a reference molecular conformation, and then the multipoles will be put in this atomic axis system. When the conformational structure changes limited, the multipole for this conformational structure will be calculated through the tensor algebra, in other words, the multipoles of each atom will not change in the atomic axis system, but this axis system moves.<sup>20</sup>

As to NH<sub>2</sub> and OH that have electron lone pairs, the multiples of the atoms having the electron lone pairs will be directly influenced by the position of the electron

lone pairs which cannot be described accurately by the analytical rotation of the atomic multipole moments. Thus, as to the atoms containing the electron lone pairs, the linear correction using first-order Taylor expansion is applied.<sup>20</sup>

CrystalOptimizer<sup>20</sup> calculations become quite efficient by reusing the results of the QM  $\Psi_{\text{mol}}$  calculations when optimising different crystal structures. A LAM/QM Database is constructed and reused as the thousands of different crystal structures in a CSP study are optimised<sup>20</sup>. In this database, the rotation matrix and the atomic matrix for the special molecule will be saved, which can be used in the optimization of the same molecule with different crystal structures.

In this project, the crystal structure is optimized using CrystalOptimizer to refine the crystal structure candidates to the lowest energy. (CrystalOptimizer switches between GAUSSIAN<sup>79</sup> runs and DMACRYS<sup>39</sup> runs. GAUSSIAN calculates the intramolecular energy penalty and multipoles for the conformation. DMACRYS calculates the intermolecular lattice energy.) However, there will be many repeated structures, thus similar structures are clustered with similar energy to reduce the results. Finally, landscapes and possible structures will be listed.

## 2.5 Conclusion

In general,  $\Psi_{\text{cry}}$  method is accurate but expensive, so it is limited in PBE-D with the use of PBE0<sup>90</sup> being limited to a small number of structures, and is mainly used by commercial CSP companies with large clusters or using cloud computing. The  $\Psi_{\text{mol}}$  method is cheaper and can use better quality molecular charge densities than can be used in  $\Psi_{\text{crys}}$  calculations. Hence  $\Psi_{\text{mol}}$  is used in chapter 3, 4, 5 to predict the crystal structures.

However, CrystalOptimizer<sup>20</sup> requires the choice of which torsion angles, angles, and bond lengths should be changed and optimized. If some angles should be optimized but were not optimized, or could not for cost, such as CBN in chapter 4, then the molecules will not pack as densely, and the hydrogen bonds could be longer than in the experiment. The  $\Psi_{\text{cry}}$  optimization relaxes all atoms and so may be more accurate than  $\Psi_{\text{mol}}$  method and you do not need to think about the choice. Thus, the periodic electronic structure method can help analyze if there is something missed in the molecular conformation optimization in crystal.

The lattice energy modelling used on CSP structures to rank their thermodynamic stability is far less accurate than the state-of-the-art methods that can be applied to a few crystal structures. Currently the prediction of the phase diagram of methanol<sup>91</sup> using fragment-based hybrid many-body interaction model,<sup>92</sup> using the highest level of theory currently applied and very high quality electronic structure methods, was still not completely accurate. CSP needs methods that can be applied to hundreds of crystal structures, not just the known polymorphs. Indeed, CSP and polymorph phase diagrams depend on relative energies and so benefit from the cancellation of errors. Progress with absolute lattice energies, as will be needed for predicting solubility are less advanced. A fragment based method was the first to prove that they had calculated the lattice energy of benzene to within 1 kJ/mol accuracy.<sup>93</sup>

### 3 CSP of tetramethyl pyrazine (TMP)

#### 3.1 Methodology

A CSD search for tetramethyl pyrazine crystal structures showed the two experimental crystal structures: monoclinic polymorph (MPYRAZ03 with no atomic coordinates,  $P2_1/c$ ,  $Z=2$   $Z'=0.5$ ) and orthorhombic polymorph (MPYRAZ, MPYRAZ01, MPYRAZ02,  $Pbca$ ,  $Z=4$   $Z'=0.5$ ). The orthorhombic form determination MPYRAZ was measured in 1951<sup>94</sup> with a larger error than MPYRAZ02(100K) and MPYRAZ01(300K) measured in 1981<sup>95</sup>. The MPYRAZ02 experimental structure will be used as the orthorhombic form, because with temperature increasing, methyl groups libration will be increased, which will lead to thermal expansion. MPYRAZ03, a monoclinic polymorph, was found during hydrolysis at room temperature in an inert environment as an organic fragment when this group characterizes the product<sup>96</sup>. The information in MPYRAZ03 is limited to cell parameters.

In the CSD, there are some co-crystals of TMP, such as XUYHUW, whose hydrogen bond donor hydroxyl group is the same as the hydrogen bond donor in CBN. The nitrogen atoms are involved in hydrogen bonding to other co-former. Thus, in the co-crystal structure formed by CBN and TMP, there may be a hydrogen bond to the nitrogen atoms in TMP.

##### 3.1.1 CSP methodology

In this chapter, a rigid model is used to predict TMP crystal structures to correspond to the methodology in co-crystal prediction. When the number of flexible degrees of freedom increases, the cost will increase drastically. All searches in this thesis must use the same assumption and methodology for the co-crystallization energies to be comparable. A search considering the methyl groups as flexible can be found in Chapter 6.1, for assessing the rigid TMP approximation.

##### 3.1.2 Generation and optimization of crystal structures

The TMP molecular geometry was extracted from experimental structure MPYRAZ02, and optimized by GAUSSIAN09<sup>79</sup> using PBE0/6-31G(d,p) methodology. This optimized molecular geometry was used as the input file for CrystalPredictor<sup>18,19</sup>. The PBE0/6-31G(d,p) method was used to calculate the

charge density in GAUSSIAN09 for TMP which was used to obtain the CHELPG potential-derived point charges for CrystalPredictor. The exp-6 FIT repulsion-dispersion potential<sup>85,86,87</sup> was used to complete the intermolecular potential model in CrystalPredictor.

This study used CrystalPredictor2.4.3.2 to generate crystal structures for TMP for the 61 default space groups. The crystal parameter ranges are considered as: cell angles: 50–130°, cell lengths: 3–40 Å, minimum cell density: 300 kg/m<sup>3</sup>, maximum intermolecular and intramolecular energies: 20 kJ/mol, region of polymorphism: 15 kJ/mol, structures generated: 100,000. After ANALYSE removed duplicate structures, the GM structure has been found many times and the number of unique crystal structures is reduced to 502, so this search is completed.

Before the ranking of the generated structures using a more accurate but expensive method, there is a clustering step to avoid duplicate calculations. In this step, structures were considered as duplicates when: energy difference < 1.00 kJ/mol, density difference < 0.005 g/cm<sup>3</sup>; RMSD<sub>30</sub> < 0.1 Å. This reduced the number of structures from 502 to 438.

The intermolecular lattice energy model was improved to the accurate but expensive model. The distributed atomic multipoles up to hexadecapole calculated by GDMA2.2<sup>80,81</sup> with PBE0/6-31G(d,p) were used in DMACRYS<sup>39</sup> to model the electrostatic component of the intermolecular lattice energy. An atom-atom exp-6 repulsion-dispersion potential with the FIT parameters was used to calculate other terms in the intermolecular energies. In the last clustering step, structures were considered as duplicates when: energy difference < 0.50 kJ/mol, density difference < 0.01 g/cm<sup>3</sup>; pXRD similarity > 0.9; RMSD<sub>30</sub> < 0.25 Å. This further reduced the number of crystal structures to 377.

## 3.2 Results

### 3.2.1 TMP conformation and its symmetry

In Figure 4, experimental conformation is collected from CSD MPYRAZ02 crystal using Mercury. The collected structure is optimized using GAUSSIAN09<sup>79</sup> with PBE0/6-31G(d,p) shown in 3.1. Figure 4 shows that there is not too much difference between experimental conformation and gas phase (RMSD<sub>1</sub>=0.0604

Å with hydrogen atoms and 0.0175 Å without hydrogen atoms), which will not lead to much energy difference in the inter- and intra- molecular interaction. Thus, the rigid search is applied to TMP search to save time.

Rotation of CH3	Exp (light green)	Gas phase (Cyan)
Upper left	175.56°	-179.97°
Lower left	174.83°	179.95°
Lower right	175.56°	179.96°
Upper right	-174.83°	180°

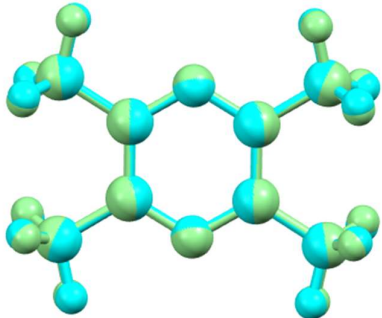


Figure 4 Left figure is the molecular overlay of MPYRAZ02 (light green) and gas phase conformation (cyan) for TMP (RMSD<sub>1</sub>=0.0604 Å with hydrogen atoms and 0.0175 Å without hydrogen atoms) and the torsion angles shown in follow the table.

There is a high symmetry of the TMP, which is shown in the CSD as the Z'=0.5 for the MPYRAZ02. Another half molecule must be added when TMP molecule is extracted. With the methyl conformations shown in Figure 4, there are 3 approximate symmetry planes (difference no more than 1°) and 1 inversion centre for the molecular structure.

After optimization by GAUSSIAN09, there is not too many changes for TMP apart from the hydrogen atoms, shown in Figure 4. As to the location of hydrogen atoms and methyl orientations, after optimization, rotation angles of methyl groups are changed a little, but the symmetry of methyl group is changed a little no more than 5°. MPYRAZ02 is measured at 100K by XRD, but the conformation is affected by the packing forces and may be affected by the libration or the measurement errors. Thus, using optimised TMP molecular structure as rigid in CSP avoids any bias to a particular packing.



### 3.2.2 TMP rigid search optimized by DMACRYS with rigid methyl groups

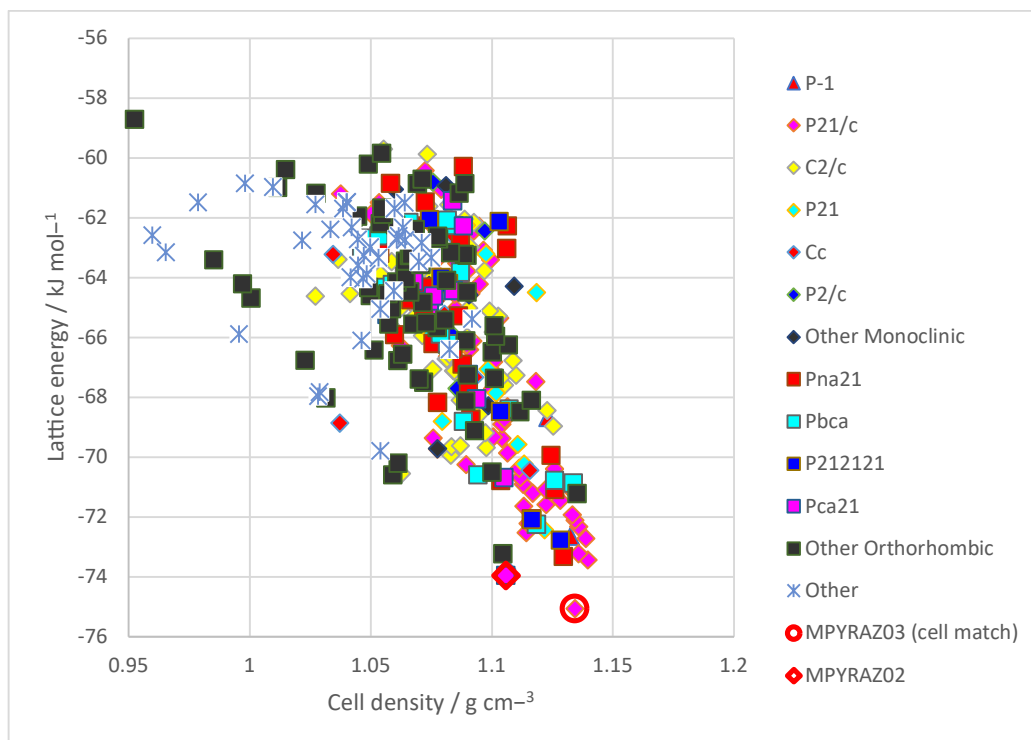


Figure 5 Summary plot of the lattice energy landscape versus cell density for TMP from  $Z'=1$  search after DMACRYS rigid refinement with PBE0/6-31G(d,p). The second lowest lattice energy structure orthorhombic polymorph is depicted by the red diamond, which is equal to lattice energy minimum starting from MPYRAZ02 with same optimization method (DMACRYS). The global minimum structure monoclinic depicted by a red circle is approximately a supercell of monoclinic MPYRAZ03.

The results of rigid molecules CSP for TMP are summarized in Figure 5 and Figure 6. In this search, the experimental structure, the orthorhombic MPYRAZ02, has been found as the second lowest energy structure labeled as T131 (The generated structures are numbered by the rank after point charge+FIT optimized by CrystalPredictor) with  $E(\text{TMP, matching exp, rig}) = -73.95$  kJ/mol. As can be seen in Figure 7, the  $\text{RMSD}_{15}$  between MPYRAZ02 and T131 optimized by DMACRYS<sup>39</sup> using PBE0/6-31G(d,p) is 0.172 Å, which means they are equivalent structures. The full list of structures is in Appendix to Chapter 3, also has the MPYRAZ02 after optimisation with the experimental confirmation held rigid which has a sufficient energy difference from the T131 that the effect of methyl group flexibility is considered in Chapter 6.

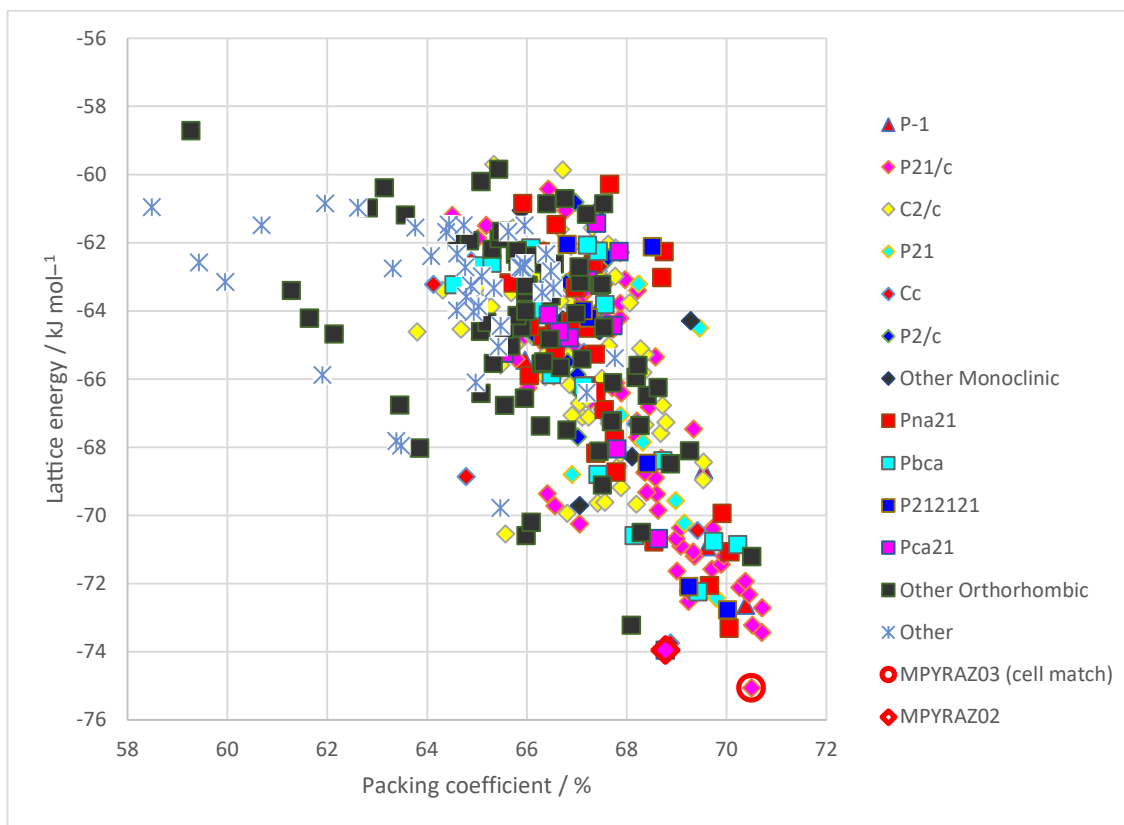


Figure 6 Summary plot of the lattice energy landscape versus packing coefficient for TMP from  $Z'=1$  search after DMACRYS rigid refinement with PBE0/6-31G(d,p). The second lowest lattice energy structure orthorhombic polymorph is depicted by the red diamond, which is equal to lattice energy minimum starting from MPYRAZ02 with same optimization method (DMACRYS). The global minimum structure monoclinic depicted by a red circle is approximately a supercell of monoclinic MPYRAZ03.

### 3.2.3 Use of CSP to suggest atomic coordinates for monoclinic polymorph

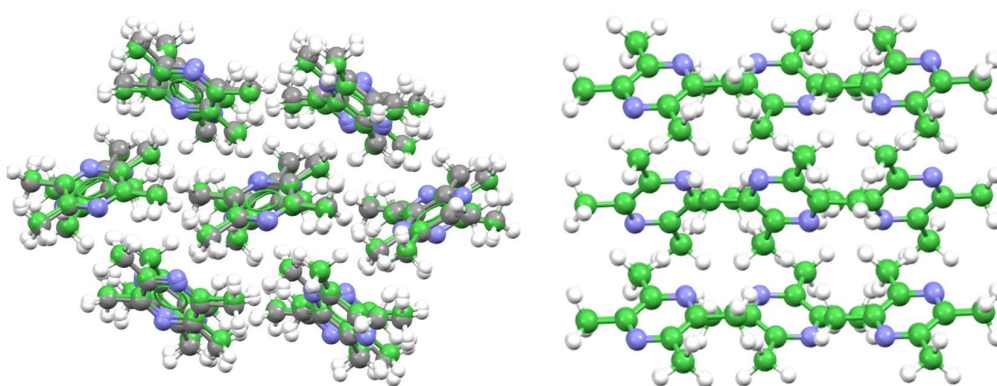


Figure 7 Left figure is the overlay ( $RMSD_{15}=0.172 \text{ \AA}$ ) of the experimental structure MPYRAZ02 and generated structure T131 optimized by DMACRYS using PBE0 6-31G(d,p), which is the second lowest lattice energy structure. Right figure is the overlay ( $RMSD_{15}=0.103 \text{ \AA}$ ) for MPYRAZ02 optimized by DMACRYS using PBE0/6-31G(d,p) and generated structure T131 (second lowest lattice energy structure) optimized by DMACRYS using PBE0/6-31G(d,p) shows that they are equivalent structures and the experimental orthorhombic polymorph has been found in this search.

As shown in Figure 5 and Figure 6, GM structure is labeled as T11 with  $E(\text{TMP,GM,rig}) = -75.06 \text{ kJ/mol}$ . As can be seen from Table 1, there are similar  $b$ ,  $c$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ , and the same space group between T11 and MPYRAZ03. Furthermore, the  $a$  parameter of T11 is twice the length of MPYRAZ03. There are no MPYRAZ03 coordinates in CSD with  $Z'=0.5$ , and TMP is a very symmetric molecule. T11 with  $Z'=1$  possible will match MPYRAZ03 with higher symmetric structure ( $Z'=0.5$ ).

Table 1 Cell parameters for T11 after DMACRYS refinement using PBE0 6-31G(d,p), and cell parameters for optimized T11 with symmetry improved by PLATON, and cell parameters of MPYRAZ03 collected from CSD whose atomic coordinates do not in the CSD.

	Space group	Z'	a/Å	b/Å	c/Å	$\alpha/^\circ$	$\beta/^\circ$	$\gamma/^\circ$
MPYRAZ03	$P2_1/c$	0.5	5.481(2)	6.936(3)	10.302(3)	90	99.90(3)	90
T11 after opt	$P2_1/c$	1	11.2077	6.8098	10.4740	90	98.0346	90
T11 after opt and PLATON	$P2_1/c$	0.5	5.6537	6.8098	10.4740	90	98.0346	90

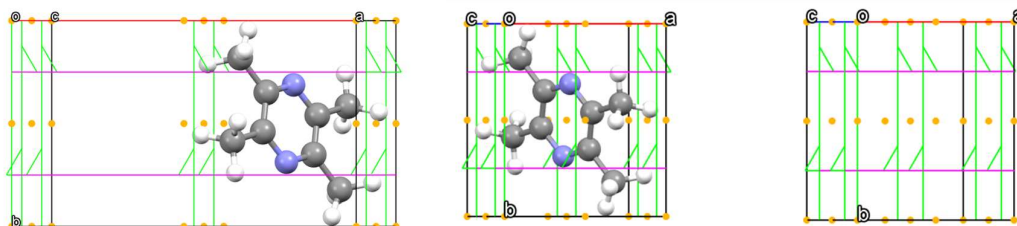


Figure 8 Symmetry elements for generated T11 optimized by DMACRYS using PBE0 6-31G(d,p) (left) and symmetry elements for optimized T11 with symmetry improved by PLATON (middle) and symmetry elements for MPYRAZ03 (right) collected from CSD.

After PLATON<sup>51</sup> calculation, a new symmetry has been found in T11, which causes the new crystal parameters shown in Table 1. Comparing those two symmetry elements and cell parameters in Table 1 and Figure 8 shows that they are equivalent crystal structures, and the monoclinic polymorph has been found in this search. The centre of TMP is in the centre of the cell and an inversion symmetry point, which can be used to explain a parameter in T11 is twice MPYRAZ03.

Whether an even better match to the cell parameters of MPYRAZ03 can be obtained by allowing methyl group flexibility will be investigated further with the flexible TMP model in Chapter 6.

### 3.3 Conclusion

The rigid body CSP for TMP is successful in that the orthorhombic form was found as the 2<sup>nd</sup> lowest energy structure. The lowest energy structure looks like a possible starting point for structure solution for the monoclinic polymorph if a powder X ray diffraction pattern could be obtained for the monoclinic form, as further discussed in section 6.1. However, using a rigid TMP molecule in a search with PBE0/6-31G(d,p) DMA<sup>80,81</sup> and FIT exp-6 potential<sup>85,86,87</sup> is adequate for the co-crystal CSP.

This is close to the ideal CSP outcome, as the global minimum structure is probably an observed polymorph, and the second lowest energy structure is the orthorhombic structure.

#### 3.3.1 Lattice energies for comparisons for predicting co-crystal formation and stoichiometry

As this rigid TMP model and potential is used in the co-crystal search (chap 5), the lattice energy that is used in evaluating the co-crystallization energy is the orthorhombic polymorph form (2<sup>nd</sup> lowest lattice energy structure matching experimental structure) of  $E(\text{TMP, matching exp, rig}) = -73.95 \text{ kJ/mol}$ , because it matches the experimental structure with the atomic coordinate. Although T11 is the GM structure and probably matches monoclinic experimental structure, it still cannot make sure about it.

## Appendix of Chapter 3

Table 2 The lowest TMP crystal structures generated by rigid search and experimental structure (MPYRAZ02) optimized by DMACRYS with PBE0 6–31G(d,p). The generated structures are numbered by the rank after point charge + FIT optimized by CrystalPredictor. MPYRAZ02 opt is the lattice energy minimum using same potential model starting from the MPYRAZ02 structure using the experimental conformation held rigid.

Label	Space Group	a /Å	b /Å	c /Å	$\alpha$ /°	$\beta$ /°	$\gamma$ /°	Density /gcm <sup>-3</sup>	Lattice Energy /kJmol <sup>-1</sup>
MPYRAZ03	P21/c	5.481(2)	6.936(3)	10.302(3)	90	99.90(3)	90		
T11	P21/c	11.3075	6.7404	10.5452	90	82.881	90	1.1343	-75.06
MPYRAZ02	P 21 21 21	8.331	9.225	10.148	90	90	90		
MPYRAZ02 opt	P 21 21 21	8.5816	9.1758	10.2507	90	90	90	1.1207	-77.2096 *
T131	Pca21	10.107	9.2944	8.7085	90	90	90	1.1058	-73.95
T372	Cc	11.8368	13.0753	7.0259	90	48.767	90	1.1062	-73.74
T19	P21/c	11.6811	10.9055	6.2681	90	83.736	90	1.1397	-73.43
T20	Pna21	20.9653	5.7528	6.6404	90	90	90	1.1295	-73.31
T1	Pbcn	6.9611	18.5039	12.7166	90	90	90	1.1045	-73.22
T24	P21/c	5.6925	6.7804	20.7957	90	82.823	90	1.1359	-73.22
T374	P212121	5.6836	6.7937	20.7661	90	90	90	1.1282	-72.77
T13	P21/c	8.9871	10.8072	8.2144	90	84.677	90	1.1388	-72.71
T208	P-1	5.5734	16.3541	8.5216	63.183	53.773	39.602	1.1325	-72.64
T64	P21/c	8.9921	7.5376	12.4525	90	74.144	90	1.1142	-72.52
T31	P21	6.9397	5.651	10.2821	90	89.689	90	1.1218	-72.43
T14	P21/n	6.3324	22.1745	5.7174	90	82.755	90	1.1359	-72.32
T32	Pbca	18.0894	10.6484	8.3976	90	90	90	1.1185	-72.24

\*Minimum with the experimental conformation does not include the conformational energy penalty.

## 4 CSP of Cannabinol (CBN)

There is only one CBN experimental crystal structure in CSD, which is CANNOL ( $Z=8$   $Z'=2$ ,  $P2_1/c$ ) without any hydrogen atoms located.

### 4.1 Adding hydrogen atom positions to experimental structure

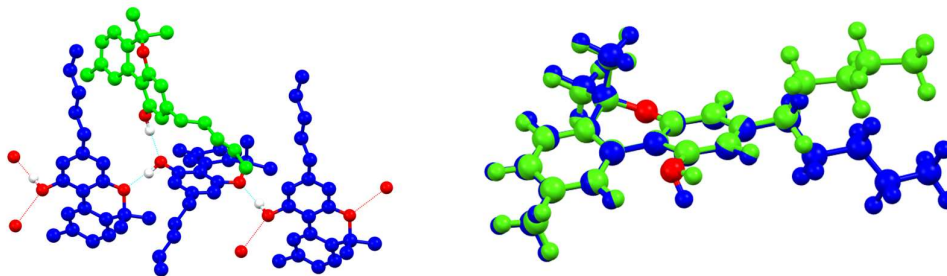


Figure 9 Diagram packing conformation of CBN experimental structure (CANNOL) showing the hydrogen atoms positions added between two oxygens (left) where added hydrogen atoms is coloured white, and oxygen atoms are coloured red. In experimental structure (CANNOL in CSD), there are two molecules in the asymmetric unit with different conformations (right).

As shown in Figure 9 left, molecule 1 (mol1) forms the chain (blue conformation); molecule 2 (mol2) forms the branch in the chain (green conformation). However, there are O...O (shown in Figure 9 as the red atoms) distances that correspond to hydrogen bonds, which implies that it is possible to add the polar (H-O) hydrogens as well as the non-polar (H-C) with confidence using Mercury, which is shown in Figure 9 left. There are 2 conformations in the asymmetric unit with different hydroxyl group and hydrocarbon pentane tail positions shown in Figure 9 right. In this thesis, mol1 (C1-C22) is the blue conformation and another green one is mol2 (C23-C44).

As can be seen in Figure 1, in CBN molecule, there are some flexible torsion angles: H26\_O2\_C6\_C7(OH), C18\_C17\_C16\_C15(dih20 or tail4), C17\_C16\_C15\_C14(dih19 or tail3), C16\_C15\_C14\_C8(dih18 or tail2), C15\_C14\_C8\_C7(dih17 or tail1) and ring C1\_C2\_C3\_C4\_C5\_O1.

## 4.2 Methodology

### 4.2.1 Molecular flexibility

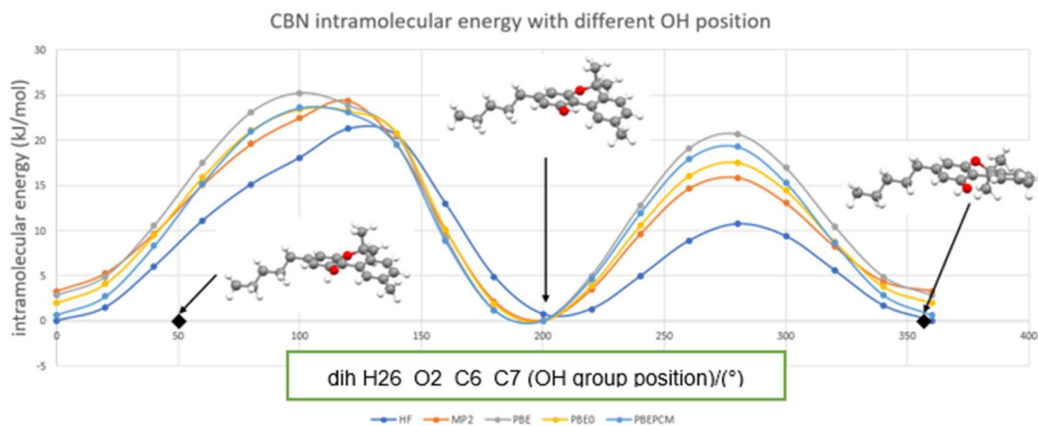


Figure 10 Relaxed conformational energy scans of CBN for hydroxyl group (OH) torsion angle at various levels of theory. Basis set: 6-31G(d,p). The inset diagrams show the CBN conformations with torsion  $H26\_O2\_C6\_C7 = 50^\circ$  (mol 1 in CANNOL),  $200^\circ$ ,  $360^\circ$  (mol 2 in CANNOL). The angles in the experimental structure of CANNOL are indicated by black diamonds.

Firstly, the conformation mol1 is extracted from the CBN experimental structure CANNOL with added hydrogen atoms by Mercury. Then, it is optimized by GAUSSIAN09<sup>79</sup> using PBE0/6-31G(d,p) to the gas phase conformation.

Relaxed conformational energy scans of mol1 in CBN for hydroxyl group torsion angle at various levels of theory are shown in Figure 10. There are not too many differences among the different levels of the theory shown in Figure 10.

Shown in Figure 9 right, the hydroxyl group of mol2 in the asymmetry unit is out of the plane of the rings and the hydroxyl group in mol1 is in the plane, which has been marked in Figure 10. Thus, shown in Figure 10, mol2 in CANNOL is not the lowest energy gas phase conformation, and there is an energy penalty between 8 and 14 kJ/mol with different computational models.

Furthermore, with the hydroxyl group scan the C1\_C2\_C3\_C4\_C5\_O1 ring flips to the alternative conformation when the hydroxyl group is close to the C10\_H5. After the flipping, there is a mirror symmetry between those two enantiomers. As to some space groups with mirror symmetry, both two enantiomers will be used to form crystals together. As to some space groups without mirror symmetry, although the two crystal structures formed by those two enantiomers are different,

they will have the same lattice energy. Thus, one enantiomer only needs to be considered in the search.

Another difference between mol1 and mol2 is the pentane tail (Figure 9), which is also very flexible. The conformation mol1 is optimized by GAUSSIAN09 using PBE0/6-31G(d,p) to the gas phase conformation. Relaxed conformational energy scans for CBN with different pentane tail positions and shapes were calculated by GAUSSIAN09 with HF/6-31G(d,p) and PBE0/6-31G(d,p).

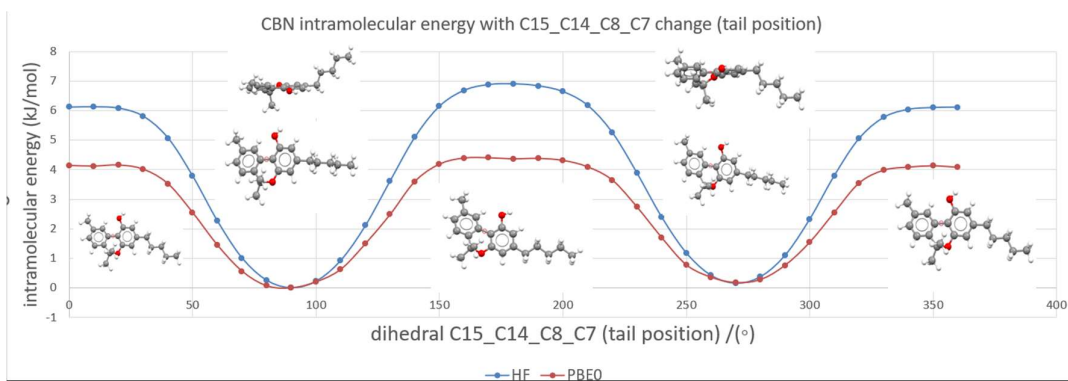


Figure 11 Relaxed conformational energy scans of CBN for the pentane tail at different computational methods with 6-31G(d,p). The inset diagrams show the CBN conformations with torsion C15\_C14\_C8\_C7 = 0, 90, 180, 270, 0/360 (°), with two minimum points for 90° and 270°.

As shown in Figure 11, as to the position of the pentane tail (dihedral C15\_C14\_C8\_C7), there are two minimum points (90° and 270°). The difference between the maximum and minimum points is no more than 4 kJ/mol using PBE0 or 7 kJ/mol using HF. Thus, considering the intermolecular interaction, all angles between 0°-360° are possible. There are two different methods have been used to scan, which are HF and PBE0. There is not too much difference between these two scan results.



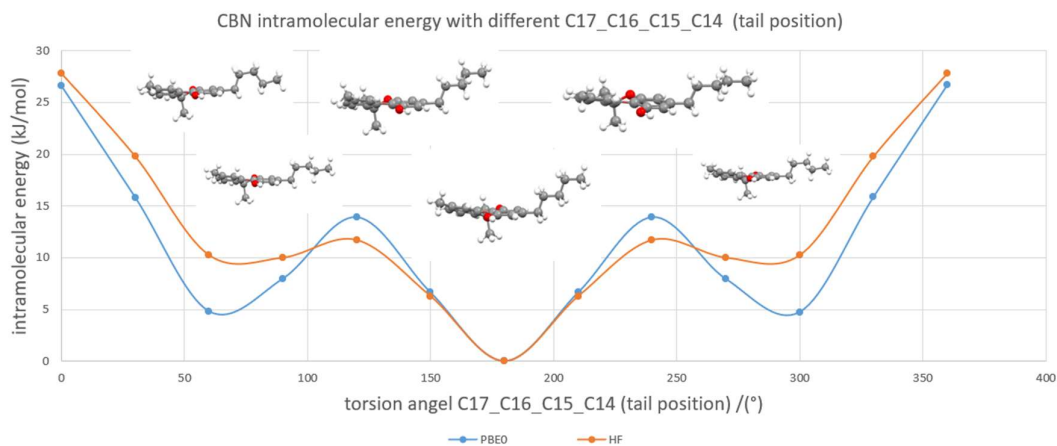


Figure 12 Relaxed conformational energy scans of CBN for the pentane tail at different DFT functional with 6-31G(d,p). The insert diagrams show the CBN conformations with torsion C17\_C16\_C15\_C14 = 0, 60, 120, 180, 240, 300 (°).

Within the hydrogen added pentane tail, there are three similar torsion angles (C17\_C16\_C15\_C14, C18\_C17\_C16\_C15, and C16\_C15\_C14\_C8) in the pentane tail. The scan of C17\_C16\_C15\_C14 has been shown because it is representative. As shown in Figure 12, with the pentane tail shape changed, the intramolecular energy changed much (about 27 kJ/mol from 0° to 180°). There are 3 local minimum points (60°, 300° and 180° (global minimum)). Between minimum points, there are about 14 kJ/mol energy barriers using HF or 12 kJ/mol using PBE0 at 120° to 240°. In the scan using PBE0/6-31G(d,p), the barrier is smaller than the HF, which means that the intramolecular energy of the pentane tail is sensitive to the computational method. As PBE0 is a more accurate method than HF, it is used for the LAM generation when the LAMs are generated for CSP of CBN. Because C17\_C16\_C15\_C14, C18\_C17\_C16\_C15, and C16\_C15\_C14\_C8 are far away from the ring, thus they have similar conformation energy curves. Based on the conformation energy scan, it should be considered that the flexible pentane tail around the 3 local minimum points of those 3 rotation angles, which will require a high computational load.

CSD Conformer Generator<sup>97</sup> analysis gives a similar distribution of angles in experimental crystal structures as will be expected from the scans. As to C17\_C16\_C15\_C14, C18\_C17\_C16\_C15, and C16\_C15\_C14\_C8, most conformations are focused around the 3 local minimum points. For

C15\_C14\_C8\_C7, where the pentane tail attaches the ring, in searched conformations, they are distributed from 0° to 360° which is shown in Figure 13.

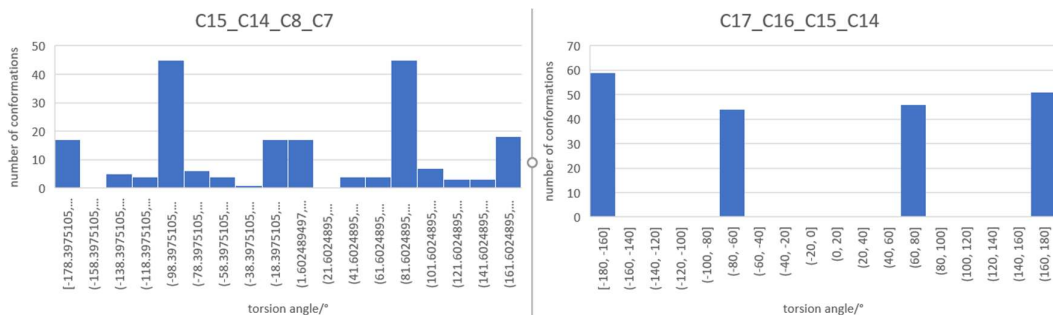


Figure 13 CSD Conformer Generator analysis for the pentane tail (C15\_C14\_C8\_C7 and C17\_C16\_C15\_C14) in CBN mol1, which matches the conformation energy scan results. C17\_C16\_C15\_C14 is focused on 180° and 60° and 300°; C17\_C16\_C15\_C14 is focusing on 90° and 270° but caused by the small energy barrier, there are large half peak width around those two minimum point. C15\_C14\_C8\_C7 is different from others as C7 and C8 are in the ring.

However, torsion angle scans of this pentane tail show that there is a wide range of angles with a small energy penalty. This will lead to a massive database of LAM points, with a lot of calculation time, and a lot of time needed by CrystalPredictor<sup>18,19</sup> to find the correct model for each step of the minimizations. Hence, it is necessary to look at the CSD to see what this sort of pentane tail does in the experimental crystal structures and see if searches can be limited somewhat. Thus, Dr. Louise Price did a fragments analysis.

The fragment in Figure 14 (a) was searched for in the CSD. There were only 60 hits, but because a number were higher Z' structures and a number had the fragment of interest multiple times, there were 113 fragments analyzed. A script was written to use the CCDC Python API, to extract the moiety shown in Figure 14 (a) and measure the angles of interest, as described in Figure 14.

As shown in Figure 14, C17\_C16\_C15\_C14 distribution is similar to the Conformation Generator analysis and conformation energy scan analysis, but as to other 3 pentane tail torsion angles are predominantly 140°–220°. This reflects the tendency of hydrocarbon chains to be extended in crystal structures (180°), which is missed by the conformer generator (Figure 13). Thus, based on the fragments analysis, the ranges of C17\_C16\_C15\_C14, C18\_C17\_C16\_C15, and C16\_C15\_C14\_C8 in CrystalPredictor are limited to 140°–220°. This means the CSP cannot generate tails conformations corresponding to the local minima in Figure 12, which are rarely observed.

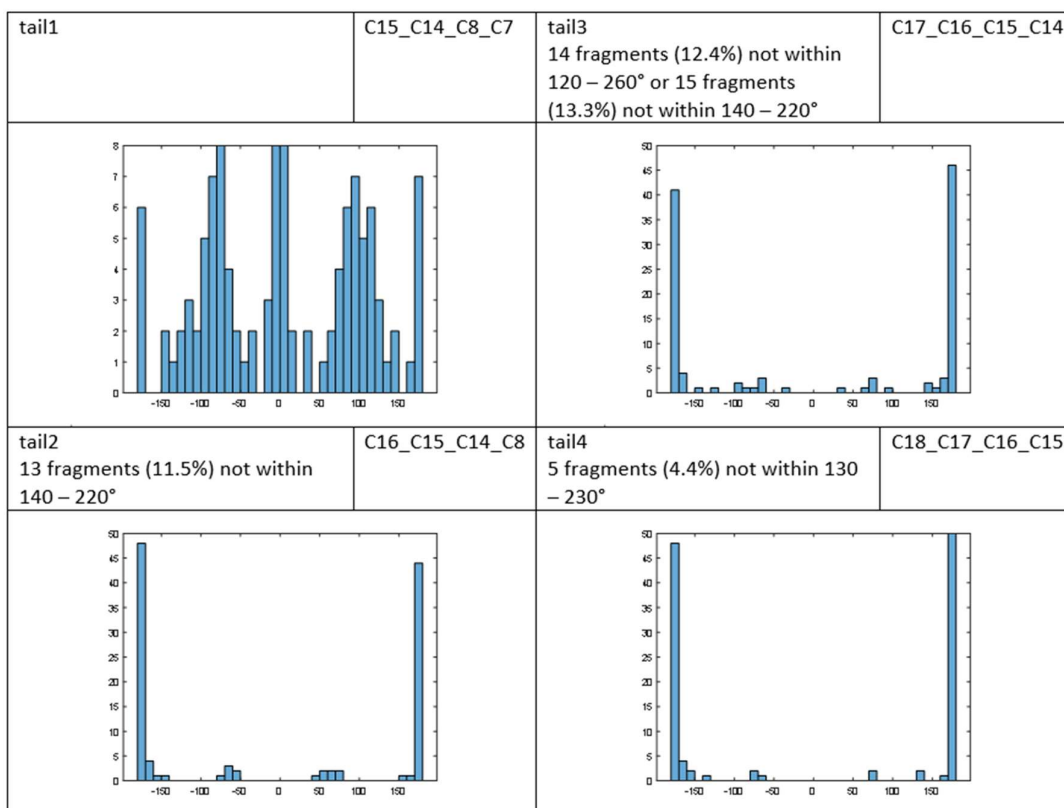
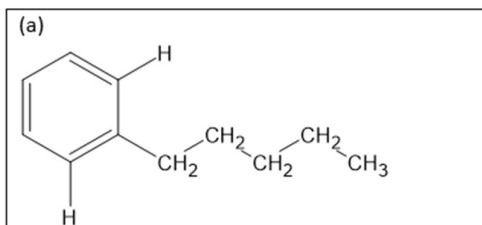


Figure 14 Fragments analyzed for CBN pentane tail shows that C17\_C16\_C15\_C14 is focused on 90° and 270° but caused by the small energy barrier, there are large half peak width around those two minimum point, which is similar to the Conformation Generator analysis and conformation energy scan analysis; C17\_C16\_C15\_C14, C18\_C17\_C16\_C15, and C16\_C15\_C14\_C8 have the similar distribution to each other, which is focused on the 140°–220°.

#### 4.2.1.1 LAM database calculation for CSP of CBN

CBN is highly flexible, however, all methyl groups were kept fixed to keep the calculations feasible. (This approximation is also applied to TMP and discussed in chapter 6.) Because the hydroxyl is far away from the pentane tail, these two groups can be described by separate LAM databases, which saves the calculation resource and time. The LAMS was calculated at PBE0/6–31G(d,p).

Table 3 Torsion angle ranges and intervals applied in CrystalPredictor LAM generation

Torsion angle	Start point(°)	interval(°)	Finish point(°)	range
Torsion group 1				
C18_C17_C16_C15	140.0°	40.0°	220.0°	120°–240°
C17_C16_C15_C14	140.0°	40.0°	220.0°	120°–240°
C16_C15_C14_C8	140.0°	40.0°	220.0°	120°–240°
C15_C14_C8_C7	200.0°	40.0°	520.0°	180°–540°
Torsion group 2				
H26_O2_C6_C7	160.0°	30.0°	430.0°	145°–445°

#### 4.2.2 Generation and optimization of crystal structures

The conformation of mol1 of the CBN experimental structure CANNOL was optimised using GAUSSIAN09<sup>79</sup> using PBE0/6–31G(d,p) to the gas phase conformation. It is used to calculate the CHELPG potential–derived point charges which were kept constant in CrystalPredictor<sup>18,19</sup>. The FIT exp–6 repulsion–dispersion potential<sup>85,86,87</sup> was used to complete the intermolecular potential model in CrystalPredictor.

This study used CrystalPredictor2.4.3.2 to generate Z'=2 crystal structures for CBN with 61 default space groups. The cell parameters are as follows: cell angles: 50–130°, cell length: 3–40 Å, minimum cell density: 300 kg/m<sup>3</sup>, maximum intermolecular/intramolecular energies: 20 kJ/mol, region of polymorphism: 15 kJ/mol, structures generated: 2,500. This was far fewer structures than will normally be generated, but the LAM database was so large that these calculations proved very expensive. The Blind Test deadline meant that using these LAMS in the co–crystal CSP (Chapter 6) was the priority. After ANALYSE removed duplicate structures, the GM structure has been found 1 time, so this search is not completed.

CrystalOptimizer<sup>20</sup> was used to refine structures after generation and optimization using point charge model and FIT potential in CrystalPredictor, allowing the

torsion angles (hydroxyl group and pentane tail shown in Table 3, and ring C5–O1–C1–C2–C3–C4) change. The distributed atomic multipoles up to hexadecapole are calculated by GDMA2.2<sup>80,81</sup> with PBE0/6–31G(d,p), which were used in CrystalOptimizer to calculate the electrostatic component of the intermolecular lattice energy and the PBE0/6–31G(d,p) ab initio method was also used for the intramolecular energy penalty  $\Delta E_{\text{intra}}$  in GAUSSIAN. Atom–atom exp–6 repulsion–dispersion potential with the FIT parameters was used to calculate all other terms in the intermolecular energies. Because of the time limitation, only 100 structures have been optimized by CrystalOptimizer. The initial parameters for CrystalOptimizer are as follows: Intermolecular local approximation settings: PBE0/6–31G(d,p), torsion cut off: 5.0°; bond angle cut off 5.0°; bond length cut off 0.2 Å; intramolecular local approximation settings: PBE0/6–31G(d,p), torsion cut off: 5.0°; bond angle cut off 5.0°; bond length cut off 0.2 Å; multipole moment derivatives: torsion: 0.2; bond angle 0.2; bond length 0.01; perturbations for numerical gradients of intermolecular energy with respect to flexible degrees of freedom; torsions: 0.5, bond angles: 0.5 and bond lengths: 0.2.

In the last clustering step, structures were considered as duplicates when: energy difference < 1.00 kJ/mol, density difference < 0.1 g/cm<sup>3</sup>; pXRD similarity > 0.9; RMSD<sub>30</sub> < 0.25 Å.

#### 4.2.2.1 Testing the CrystalOptimizer methodology for CBN on CANNOL

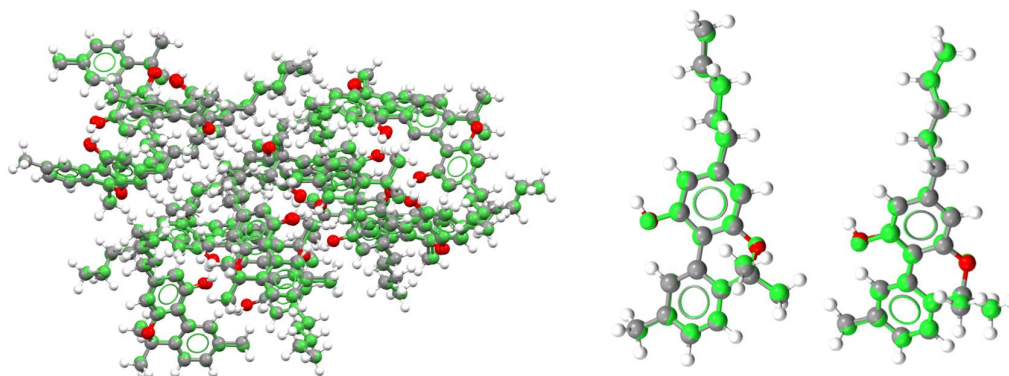


Figure 15 Left figure is the overlay of the experimental structure CANNOL (green) and experimental structure optimized using CrystalOptimizer with PBE0 6–31G(d,p) and FIT potential for hydroxyl group, pentane tail and ring discussed in the last part. RMSD<sub>15</sub> is 0.17 Å. The middle and right figures are the overlay of CBN experimental conformations and optimized conformations using CrystalOptimizer with PBE0 6–31G(d,p) and FIT potential for hydroxyl group, pentane tail and ring discussed in the last part. RMSD<sub>1</sub> for mol1 is 0.0922 Å and RMSD<sub>1</sub> for mol2 is 0.116 Å without hydrogen atoms.

To assess whether the choice of degrees of freedom to be refined by CrystalOptimizer<sup>20</sup> is suitable, it was tested on the CBN experimental structure CANNOL. CANNOL is optimized using CrystalOptimizer with the PBE0/6-31G(d,p) value of  $\Delta E_{\text{intra}}$  and distributed multipoles with the flexible hydrogen carbon pentane tail and hydroxyl group and ring. The resulting lattice energy minima had a  $\text{RMSD}_{15}=0.17 \text{ \AA}$  overlay with CANNOL, excluding the hydrogen atoms as calculated in Mercury. Furthermore, the resulting CBN conformations minima had  $\text{RMSD}_1=0.0922 \text{ \AA}$  and  $0.116 \text{ \AA}$  separately overlay with CANNOL, excluding the hydrogen atoms as calculated in Mercury. Especially for the pentane tail and hydroxyl group in each conformation, after the optimization they approximately keep their position. This shows that the CrystalOptimizer methodology used for CBN in all the CSP studies is suitable for CBN experimental structure CANNOL.

#### 4.3 CBN CSP Results

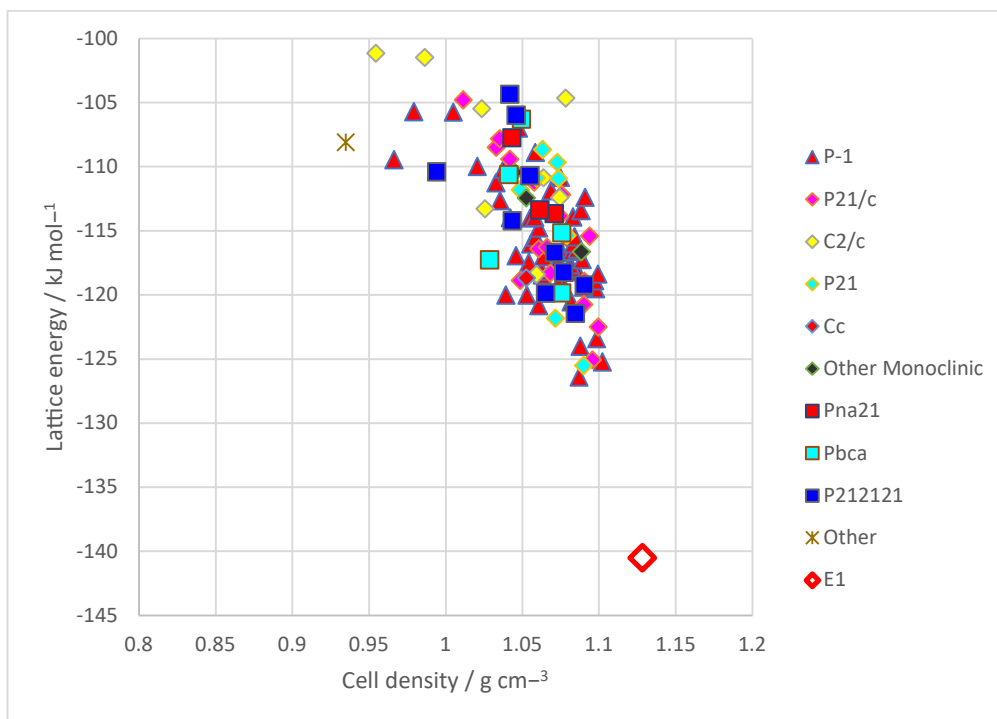


Figure 16 Summary plot of the lattice energy landscape versus cell density for CBN from  $Z'=2$  search after optimized using CrystalOptimizer with PBE0/6-31G(d,p). The lattice energy minimum starting from CANNOL with the same methodology is denoted as E1.

CANNOL was not found in this limited search. (The reasons for this are analysed in chapter 6.)

Generated crystal structures are numbered by the rank after point charge + FIT potential<sup>85,86,87</sup> optimized by CrystalPredictor<sup>18,19</sup>, for example, C1 is the global minimum structure after point charge + FIT optimization. All those total lattice energies in this search are higher than the experimental structure lattice energy optimized by CrystalOptimizer with the same flexible group to this search ( $E(\text{CBN,exp,opt}) = -140.499 \text{ kJ/mol}$ ). The lattice energy of GM structure (C9 shown in the appendix of Chapter 4) in this search is  $-126.381 \text{ kJ/mol}$ , shown in Figure 16.

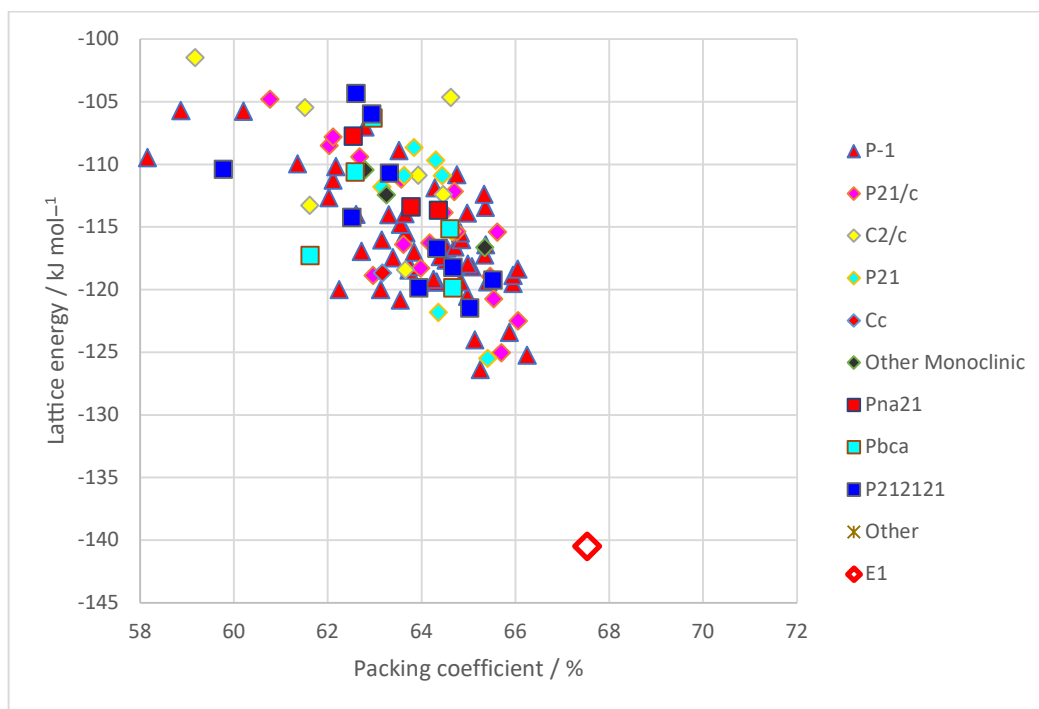


Figure 17 Summary plot of the lattice energy landscape versus packing coefficient for CBN from  $Z'=2$  search after optimized using CrystalOptimizer with PBE0/6-31G(d,p). The lattice energy minimum starting from CANNOL with the same methodology is denoted as E1.

As shown in Figure 17 and Figure 16, the experimental structure has the maximum density and packing coefficient and there is a gap between the experimental structure and other generated structures, shown in Table 4 in the Appendix of chapter 4. Meanwhile, optimized CANNOL has significantly more stable energy. The stability of the experimental structure is caused by the more stabilising intermolecular energy. The intramolecular energy for mol1 is destabilising, comes pending to the hydroxyl group being out of the plane. The

lowest energy structure generated in the search is C9, which has no hydrogen bond (defined as Mercury default parameter), but C5 does have a similar hydrogen bond system to CANNOL (mol1 form the chain and mol2 form the branch), but the relevant position between leads to the long and weak hydrogen bond and the low packing coefficient.

Generated GM structure is formed by the gas phase conformation which is totally different from the experimental conformation. Most generated structures are formed by the low energy conformation with the hydroxyl group in the plane of the ring. Some structures formed a hydrogen bond, but they are different from the experimental hydrogen bond system because of the direction of the hydroxyl group. Because of the time limitation, the lattice energy calculated with the same method as the CSP search start from the experimental structure is used to calculate the co-crystallization energy. The issue details of why the experimental structure was not found are investigated in chapter 6.

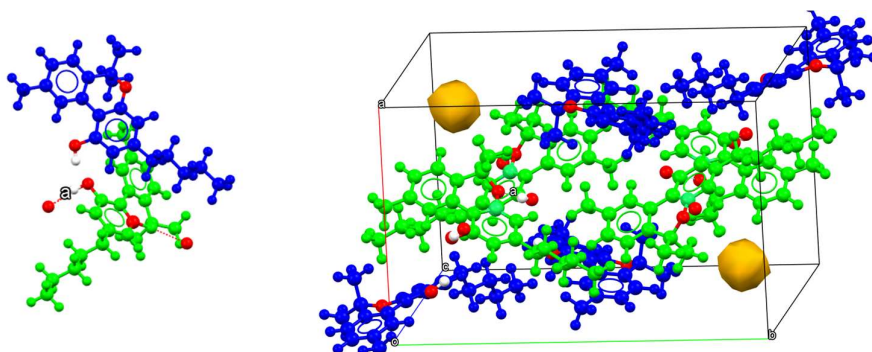


Figure 18 Generated CBN crystal structure (C5 (P2<sub>1</sub>/c) shown in the appendix) formed by C1,1(6) a D1,1(2) b D3,3(9) >b>a<b hydrogen bond which is same to experimental structure, but their crystal structure are totally different and the hydrogen bond between mol1 and mol2 is too large (3.250 Å).

C5 has the most similar hydrogen bond structure to the experimental structure (CANNOL), which has been highlighted in Table 4. The main difference between C5 generated structure and the experimental structure is the relevant position among mol1 forming chain, which causes the mol 1 and mol2 far away from each other (distance between O<sub>(mol1)</sub>...O<sub>(mol2)</sub> is 3.25 Å larger than the experimental structure 2.921 Å) shown in Figure 19. More details about hydrogen bond energy will be discussed in chapter 6. A reasonable hydrogen bond system can reduce the intermolecular energy, but bad packing with high void (Figure 18) and low



density (packing coefficient), will reduce the stabilising dispersion. More detail will be discussed in limitation chapter 6.

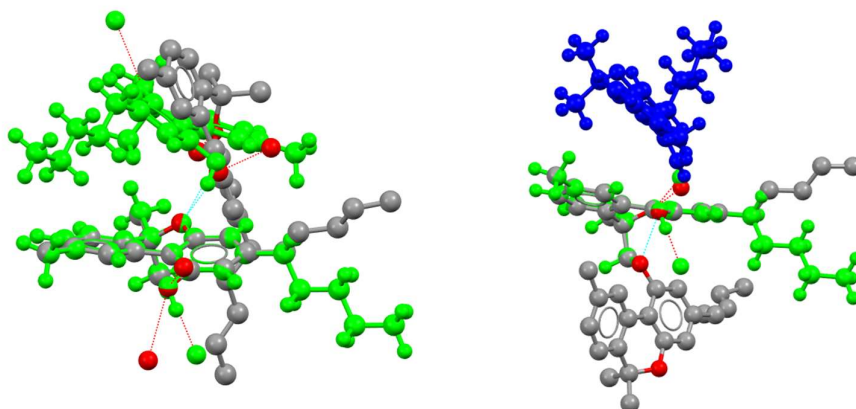


Figure 19 Comparing C5 to experimental structure. The left is the comparison of the chain; the right is the comparison of the asymmetry unit. CANNOL is coloured by elements, and C5 is coloured by blue and green with  $Z'=2$ .

In general, C5 has a similar hydrogen bond to the experimental structure, but 1) bad relative position between mol1 and mol2 (bad packing) leads to mol1 being far away from mol2, this hydrogen bond will lead to a higher energy than the experimental structure. 2) packing coefficient and density of C5 are lower than CANNOL, which shows that CANNOL has better dispersion stabilization than C5.

#### 4.4 Conclusion

The  $Z'=2$  experimental crystal structure of CBN ( $-140.499$  kJ/mol) was significantly lower in energy than the lowest energy structure ( $-126.381$  kJ/mol) generated in a small CrystalPredictor<sup>18,19</sup> search. A full  $Z'=2$  search will be very resource-intensive and was not a good use of resources given the timetable of the Blind Test submission for XXX. The reason why the CBN experimental structure was not found in this search is investigated in chapter 6.

Furthermore, the model used in CrystalOptimizer<sup>20</sup> is adequate for the balance between intramolecular energy and intermolecular interactions and it is adequate for reproducing the experimental structure CANNOL with the approximations of its flexibility. However, the variation in the density and hydrogen bonding and conformations between structures, for example the relative lattice energies of structures C5 and CANNOL and GM, are likely to be sensitive to the lattice energy model, as investigated in Chapter 6.

In this small flexible CBN search after the point charge model optimization, CANNOL is ranked as the 6th lowest structure, but the CrystalOptimizer models calculate the CANNOL lattice energy to be much lower (~ 15 kJ/mol) than GM in the CBN search. The difference in ranking between CrystalOptimizer and CrystalPredictor shows the limitation of CrystalPredictor, which will be discussed in chapter 6.

#### 4.4.1 Lattice energies for comparisons for predicting co-crystal formation and stoichiometry

The experimental structure was not found in the limited search and the lattice energy of GM in this search is far higher than the experimental structure, the lattice energy of the experimental structure (optimizing the pentane tail, hydroxyl group and ring by CrystalOptimizer<sup>20</sup>)  $E(\text{CBN,exp,opt}) = -140.499 \text{ kJ/mol}$ , will be used to evaluate the co-crystallization energy to analyse the co-crystal searches.

## Appendix of Chapter 4–The crystal structure of CBN after CrystalOptimizer refinement.

Table 4 The lowest CBN crystal structures generated by flexible search and experimental structure (CANNOL) optimized by DMACRYS with PBE0 6–31G(d,p). The generated structures are numbered by the rank after point charge + FIT optimized by CrystalPredictor. CANNOL opt is the experimental structure CANNOL optimized by CrystalOptimizer. C5 has the similar hydrogen bond graph sets to experimental structure, which has been highlight by green and shown in Figure 18 and Figure 19.

\* Hydrogen bond graph sets: Ga,b,(c) G: C means chain, D means noncyclic dimers and other finite hydrogen bonds sets R means ring. a, b are number of acceptors and donors.

Label	hydrogen bond graph sets*	Space Group	a /Å	b /Å	c /Å	$\alpha$ /°	$\beta$ /°	$\gamma$ /°	Density /gcm <sup>-3</sup>	intra energy mol1/ kJmol <sup>-1</sup>	Intra energy mol2/ kJmol <sup>-1</sup>	Intermolecular Lattice Energy /kJmol <sup>-1</sup>	Repulsion Dispersion Energy /kJmol <sup>-1</sup>	Lattice Energy /kJmol <sup>-1</sup>
CANNOL	C1,1(6) a D1,1(2) b D3,3(9) >b>a<b	P21/c	20.658	18.166	13.517	90	18.166	90						
CANNOL opt	C1,1(6) a D1,1(2) b D3,3(9) >b>a<b	P21/c	20.4386	18.1765	13.6051	90	133.677	90	1.1281	13.98761	3.49	-151.34	-243.03	-140.499
C9	D1,1(2) a	P-1	10.1774	10.7714	17.7754	83.151	83.305	80.245	1.0867	5.47028	4.006	-131.12	-242.4	-126.381
C39	C1,1(6) a	P21	12.4663	27.8993	5.4794	90	96.872	90	1.0898	4.42206	6.18015	-130.79	-242.37	-125.488
C56	D1,1(2) a	P-1	10.6008	8.7326	20.6606	92.434	78.371	92.075	1.1021	9.86595	7.52588	-133.92	-249.72	-125.223
C5	C1,1(6) a D1,1(2) b D3,3(9) >b>a<b	P21/c	13.4917	21.8264	12.9664	90	80.322	90	1.0956	16.38315	6.77632	-136.62	-243.86	-125.04
C12	no	P-1	26.3125	6.316	11.7811	76.779	84.628	91.006	1.0875	2.51226	7.0598	-128.79	-246.45	-124.003
C52	D1,1(2) a	P-1	6.0712	28.7433	11.0623	97.647	79.498	95.045	1.0984	9.0191	10.00684	-132.92	-249.42	-123.406
C48	no	P21/c	23.4517	6.2219	26.6178	90	74.979	90	1.0993	4.62593	8.24376	-128.92	-250.3	-122.484
C14	no	P21	7.6431	20.658	12.3079	90	82.017	90	1.0714	4.29189	4.25037	-126.09	-240.46	-121.818
C4	D1,1(2) a	P212121	14.3139	17.36	15.3032	90	90	90	1.0845	5.33913	10.34877	-129.29	-251.78	-121.445
C75	D1,1(2) a	P-1	26.9887	8.7565	8.6729	75.573	99.578	86.132	1.0604	4.01057	8.06425	-126.87	-239.68	-120.832
C7	no	P21/n	25.8395	5.3378	27.4367	90	91.083	90	1.0899	9.15042	7.80498	-129.21	-247.29	-120.732
C31	no	P-1	8.7234	15.2646	14.7036	80.146	82.043	84.88	1.0817	4.17153	2.99241	-124.1	-239.3	-120.517
C38	no	P-1	15.6739	17.5525	7.6542	98.259	93.597	72.242	1.0391	7.687	7.21375	-127.45	-236.76	-119.999

## 5 Co-crystal CSP

In section 1.2.2, the challenge of the Blind Test target XXX has been introduced, that is the stoichiometry challenge for the co-crystal structure prediction formed with flexible co-former, cannabinalol (CBN) and tetramethyl pyrazine (TMP).

### 5.1 Methodology for 7<sup>th</sup> CCDC Blind Test

Based on the last two chapters, it is decided to keep TMP rigid and make CBN flexible in co-crystal search to save cost. In chapter 1, the method of assessing the co-crystal stability has been introduced, which is the co-crystallization energy<sup>24</sup> (Equation 1). To compare the co-crystal stability among different stoichiometries of CBN<sub>m</sub>:TMP<sub>n</sub> the co-crystallization energy is

$$\text{Equation 4 } \Delta E_{cc} = \frac{E_{latt}(CBN_mTMP_n) - nE_{latt}(TMP)}{m} - E_{latt}(CBN)$$

so that all co-crystallization energies are relative to CBN.

#### 5.1.1 Generation of crystal structures

CrystalPredictor2.4.3.2<sup>18,19</sup> is used to generate structures for TMP:CBN= 1:1, 1:2, 2:1 with 1 formula unit in the asymmetric unit cell requesting 1,000,000 structures. All co-crystal searches cover the 61 most possible space groups (shown in 2.4.1.1) for organic molecules in the CSD, adapting the method used for TMP in chapter 3 and CBN in chapter 4. Generated structures are optimized using a point charge model in CrystalPredictor first and ranked. CrystalOptimizer2.4.7.1<sup>20</sup> is used to refine the selected torsion angles (same as in chapter 4) alongside the cell parameters for the ranked low energy structures in the last step. The conformational databases for the molecular flexibility of CBN used in Chapter 4 can be reused for the co-crystals, as discussed in sections 2.4.1.1 and 2.4.2.1. When the number of flexible degrees of freedom increases, the cost will increase drastically. However, all searches in this thesis must use the same methodology for the energies to be comparable. The intermolecular interactions between TMP and CBN were given by the combining rules for the exp-6 repulsion -dispersion and the electrostatic term is additive. The PBE0/6-31G(d,p) ab initio method was used to provide the CHELPG potential-derived point charges which were used in CrystalPredictor.

In these searches, CrystalOptimizer is used to rank the generated structures as the accurate model. The distributed atomic multipoles up to hexadecapole are calculated by GDMA2.2<sup>80,81</sup> with PBE0/6-31G(d,p), which were used in CrystalOptimizer to calculate the electrostatic component of the intermolecular lattice energy and the PBE0/6-31G(d,p) ab initio method was also used for the intramolecular energy penalty  $\Delta E_{\text{intra}}$  in GAUSSIAN<sup>79</sup>. Atom-atom exp-6 repulsion-dispersion potential with the FIT parameters<sup>85,86,87</sup> were used to calculate all other terms in the intermolecular energies. The initial parameters for CrystalOptimizer are as follows: Intermolecular local approximation settings: PBE0/6-31G(d,p), torsion cut off: 5.0°; bond angle cut off 5.0°; bond length cut off 0.2 Å; intramolecular local approximation settings: PBE0/6-31G(d,p), torsion cut off: 5.0°; bond angle cut off 5.0°; bond length cut off 0.2 Å; multipole moment derivatives: torsion: 0.2; bond angle 0.2; bond length 0.01; perturbations for numerical gradients of intermolecular energy with respect to flexible degrees of freedom: torsions: 0.5, bond angles: 0.5 and bond lengths: 0.2.

In the clustering step, structures were considered as duplicates when: energy difference < 1.00 kJ/mol, density difference < 0.1 g/cm<sup>3</sup>; pXRD similarity > 0.9; RMSD<sub>30</sub> < 0.25 Å.

#### *5.1.1.1 Some changes in 2CBN:TMP and 2TMP:CBN searches*

A problem was observed for 2CBN:TMP and 2TMP:CBN searches that is caused by the density limitation in CrystalPredictor shown in section 2.4.1.1. So the cell length parameters in those two searches had been increased and each search has been divided into two batches with different space groups.

The search parameters for 1TMP:1CBN are as follows: cell angles: 50–130°; cell length: 3–40 Å; maximum intermolecular/intramolecular energies: 20 kJ/mol; minimum cell density: 300 kg/m<sup>3</sup>, region of polymorphism: 15 kJ/mol. Also, compared to the CBN search in Chapter 4 and TMP search in Chapter 3 and 1CBN:1TMP co-crystal search, in the 1CBN:2TMP and 1TMP:2CBN co-crystal searches, the cell length limitation is increased from 40 Å to 50 Å in CrystalPredictor. As can be seen from the molecular structure shown in Figure 1, there are 49 atoms and a large pentane tail in CBN conformation. There are 3 molecules including 2 CBN and 1 TMP for 2CBN:1TMP co-crystal. In the

CrystalPredictor step, based on the log file, many candidates of some space groups (e.g R-3c) are eliminated. The reason why CrystalPredictor became stuck is that the limitation of the cell length makes many candidates eliminated within 40 Å cell length limitation for some space groups (e.g R-3c). Longer cell length was applied to 1CBN:2TMP and 2CBN:1TMP only, not to CBN and 1CBN:1TMP.

Furthermore, it is decided to run CrystalPredictor in two batches. The first batch calculated the ten most probable space groups, and other 51 space groups were calculated in another batch, which improved the CrystalPredictor efficiency. After both batches are finished, the stable structures in those two batches will be added together and run ANALYSE together. Two batches were for 1CBN:2TMP and 2CBN:1TMP only. For 1CBN:2TMP and 1TMP:2CBN searches, the searches were done with 10 most common space groups for 100,000 structures and other 51 groups for 10,000 structures separately and then put them together. Furthermore, the cell length range is changed to 3–50 Å. As to other parameters, they are the same as the parameters in chapters 3 and 4.

## 5.2 Results

In these co-crystal searches, the unit kJ/mol for lattice energy means kJ per mol asymmetry unit cells instead of kJ per mol molecules. In other words, as to 1CBN:1TMP, kJ/mol means kJ per 1 mol CBN molecules and 1 mol TMP molecules. As to 1CBN:2TMP, kJ/mol means kJ per 1 mol CBN molecules and 2 mol TMP molecules. As to 2CBN:1TMP, kJ/mol means kJ per 2 mol CBN molecules and 1 mol TMP molecules.

### 5.2.1 1CBN:1TMP co-crystal

Based on Equation 4, the structures under the blue line (Figure 20) are more stable than the separate co-formers, because their co-crystallization energy is lower than 0. The lattice energies of only 11 structures are lower than the sum of experimental CBN and TMP lattice energy. However, in the study<sup>24</sup>, some 1CBN:1TMP co-crystal structures whose crystallization energy is about 8 kJ/mol have been found. If the result in this study are applied, the co-crystallization energy of stable co-crystal can be extended to 8 kJ/mol, indicates more than 50 potential co-crystal structures. The most stable structures are shown in Table 5.

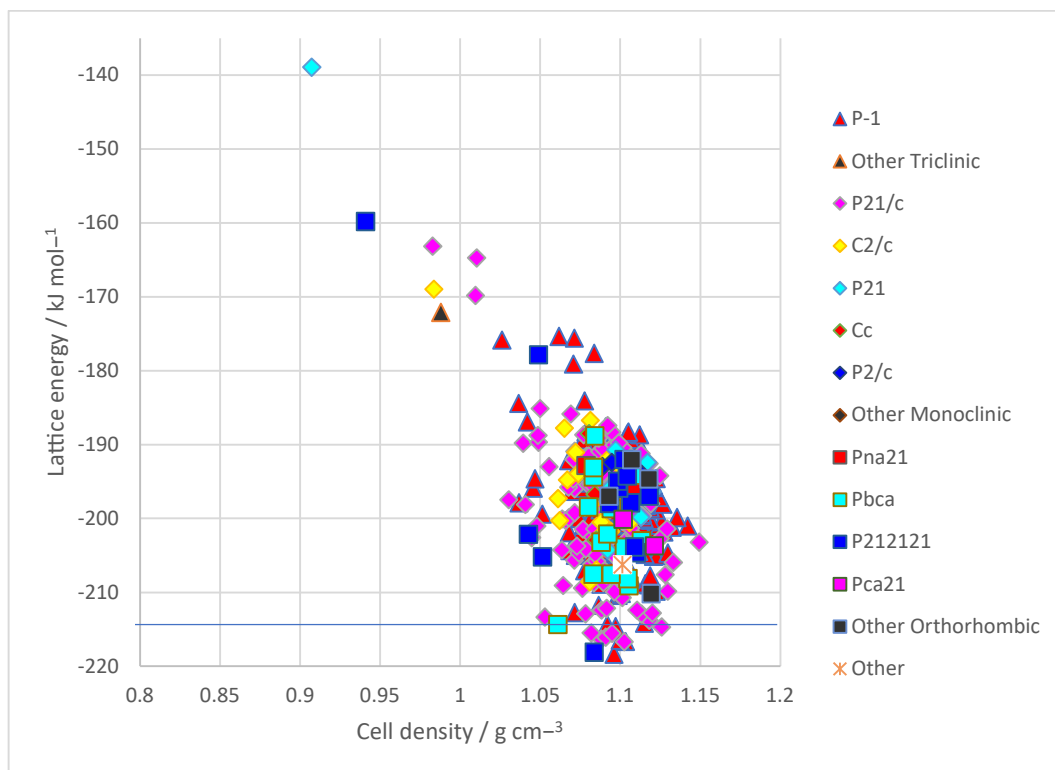


Figure 20 Lattice energy landscape (density) for co-crystal 1CBN:1TMP search. Blue line:  $E(\text{CBN,exp,opt}) + E(\text{TMP,matching exp,rig}) = -214.499 \text{ kJ/mol}$ , the sum of experimental CBN and TMP lattice energy after CrystalOptimizer.

There is a balance between packing (dispersion) and hydrogen bonding. Figure 21 is the lattice energy landscape against packing coefficient in %. In Figure 21, some of the lowest energy structures have relatively low packing coefficients, and inspection (Table 5) shows that these structures contain hydrogen bonds. Generated 1CBN:1TMP co-crystals with low lattice energy have a lower packing coefficient than experimental CBN (67.5%) and TMP (68.7% and 70.5%). The 5 highest packing coefficient structures (in the right of Figure 21 landscape) do not have any hydrogen bonds. Forming CBN(OH)...TMP(N) hydrogen bond will lead to the voids and low packing coefficient, because of the big difference in size of those two molecules. But forming hydrogen bonds will stabilize the intermolecular energy. All the most stable structures have hydrogen bonds as shown in Table 5. So, considering the balance between hydrogen bond and density (packing coefficient), the landscape shape is reasonable, although it is different from the other CSP result shape caused by dispersion, such as the TMP landscape in Chapter 3.

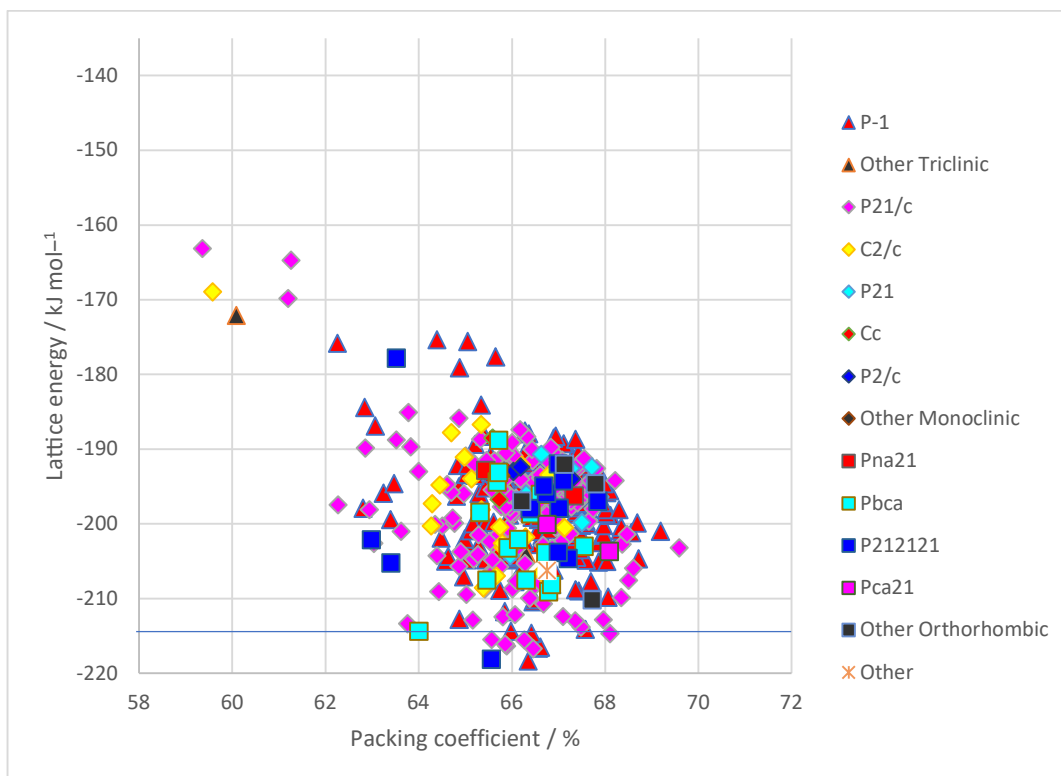


Figure 21 Lattice energy landscape (packing coefficient) for co-crystal 1CBN:1TMP search. Blue line:  $E(\text{CBN}, \text{exp}, \text{opt}) + E(\text{TMP}, \text{matching exp}, \text{rig}) = -214.499 \text{ kJ/mol}$ , the sum of experimental CBN and TMP lattice energy after CrystalOptimizer.

There is also a balance between the intermolecular energy and CBN intramolecular energy caused by the hydrogen carbon pentane tail. The intramolecular energies of CBN for most low crystal energy structures are between 4 kJ/mol and 10 kJ/mol shown in Table 5. CT279 is the global minimum structure in this 1CBN:1TMP search with  $U_{\text{inter}} = -225.25 \text{ kJ/mol}$  and  $\Delta E_{\text{intra}}(\text{CBN}) = 6.883 \text{ kJ/mol}$ . CT5 has the lowest intermolecular energy in Table 5. It is the third lowest energy structure in this search with  $U_{\text{inter}} = -226.21 \text{ kJ/mol}$  and  $\Delta E_{\text{intra}}(\text{CBN}) = 9.521 \text{ kJ/mol}$ . CT119 has the most stable CBN conformation in Table 5. It is the 9<sup>th</sup> lowest energy structure in this search with the  $U_{\text{inter}} = -220.03 \text{ kJ/mol}$  and  $\Delta E_{\text{intra}}(\text{CBN}) = 4.538 \text{ kJ/mol}$ . All structures in Table 5 have hydrogen bonds between CBN and TMP.

The CBN conformation will affect the lattice packing, shown in Table 5, all structures' CBN intramolecular energy are higher than 4 kJ/mol, which is caused by the hydrogen carbon pentane tail conformation. To reduce the intermolecular energy, the hydrogen carbon pentane tail cannot be in the lowest conformation.



The limited flexibility of the hydrogen carbon pentane tail may have led to a packing problem in the CBN:TMP co-crystal search, since the packing coefficients are all low.

Table 5 The most stable CSP generated co-crystal structures of CBN:TMP are named as the rank after point charge + FIT optimized by CrystalPredictor, for example, CT1 is the global minimum structure after point charge + FIT optimization. Reference energy is:  $E(\text{CBN,exp,opt})+E(\text{TMP, matching exp,rig}) = -214.499 \text{ kJ/mol}$ . Co-crystallization energy is calculated by  $\Delta E_{\text{cc}} = E_{\text{latt}}(\text{co-crystal}) - E_{\text{latt}}(\text{TMP}) - E_{\text{latt}}(\text{CBN})$ . Structures that co-crystallization energies are lower than 0 are coloured as green.

\* Hydrogen bond graph sets: Ga,b,(c) G: C means chain, D means noncyclic dimers and other finite hydrogen bonds sets R means ring. a, b are number of acceptors and donors.

Label	Hydrogen bond graph sets*	Space Group	a /Å	b /Å	c /Å	$\alpha /^\circ$	$\beta /^\circ$	$\gamma /^\circ$	Density /gcm <sup>-3</sup>	$\Delta E_{\text{intra}} / \text{kJmol}^{-1}$	Intermolecular Lattice Energy /kJmol <sup>-1</sup>	Repulsion Dispersion Energy /kJmol <sup>-1</sup>	Lattice Energy /kJmol <sup>-1</sup>	co-crystallization energy kJmol <sup>-1</sup>
CT279	D1,1(2) a	P-1	13.0864	9.2408	11.8221	83.101	73.823	81.494	1.0961	6.883	-225.25	-170.95	-218.367	-3.918
CT93	D1,1(2) a	P212121	25.5543	8.8093	12.1603	90	90	90	1.0837	7.195	-225.31	-168.04	-218.115	-3.666
CT5	D1,1(2) a	P21/n	19.0071	8.4776	16.7069	90	88.138	90	1.1025	9.521	-226.21	-170.53	-216.689	-2.24
CT59	D1,1(2) a	P-1	10.0731	8.5162	18.3574	85.122	107.472	66.909	1.1033	9.131	-225.75	-174.46	-216.619	-2.17
CT60	D1,1(2) a	P-1	9.4488	14.8139	10.6521	72.696	72.814	91.704	1.0999	7.998	-224.49	-170.12	-216.492	-2.043
CT814	D1,1(2) a	P21/c	12.492	29.7938	7.3436	90	86.015	90	1.088	6.017	-222.35	-169.42	-216.333	-1.884
CT459	D1,1(2) a	P21/c	8.8795	25.2144	12.1848	90	85.473	90	1.0908	7.239	-223.32	-165.6	-216.081	-1.632
CT229	D1,1(2) a	P21/n	13.8442	11.7398	16.6711	90	90.444	90	1.0949	9.262	-224.81	-168	-215.548	-1.099
CT119	D1,1(2) a	P21/c	12.3564	30.6036	7.2637	90	86.709	90	1.0818	4.538	-220.03	-166.19	-215.492	-1.043
CT67	D1,1(2) a	P21/c	13.8928	12.6925	15.6965	90	72.174	90	1.1258	5.4	-220.09	-184.04	-214.69	-0.241
CT674	D1,1(2) a	P-1	8.4703	18.3144	12.4792	54.404	92.462	116.105	1.0971	9.377	-224.02	-171.18	-214.643	-0.194

### 5.2.2 1CBN:2TMP co-crystal

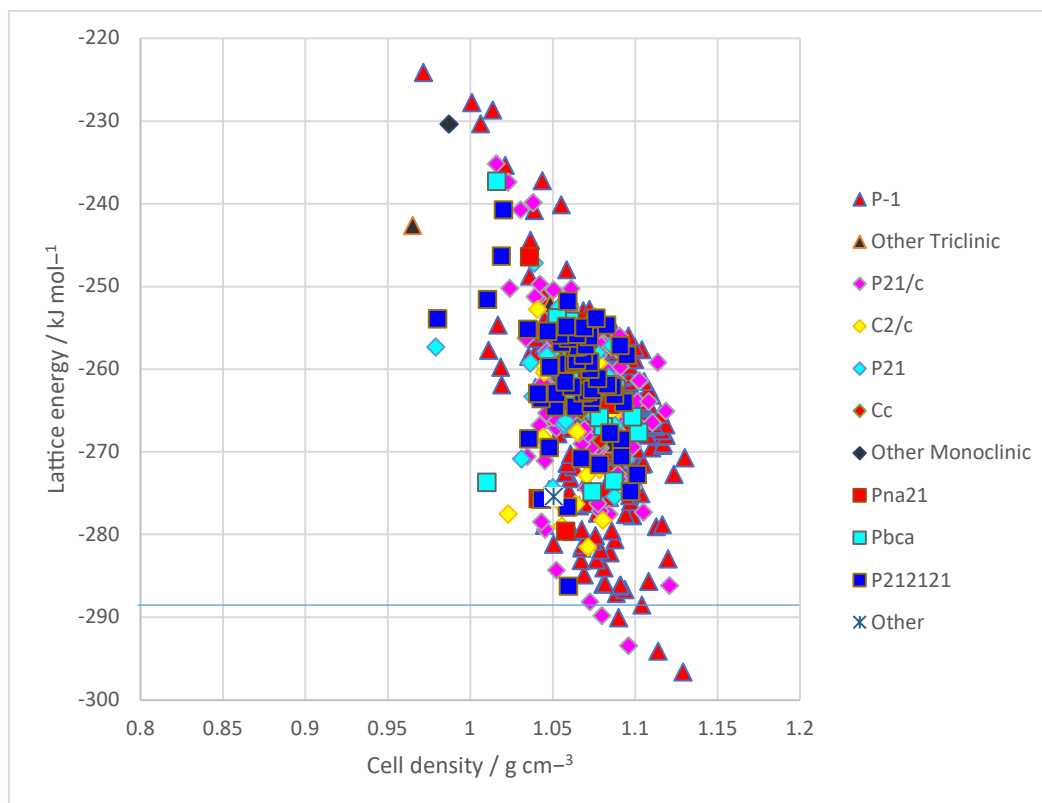


Figure 22 Lattice energy landscape (density) for co-crystal 1CBN:2TMP search. Blue line:  $2E(\text{TMP, matching exp,rig}) + E(\text{CBN,exp,opt}) = -288.399 \text{ kJ/mol}$ , the sum of experimental CBN and 2 TMP lattice energy after CrystalOptimizer.

The lattice energies of 6 CBN:2TMP structures are lower than the sum of experimental CBN and 2 TMP lattice energies (Equation 4). More than 20 structures' co-crystallization energies are lower than 8 kJ/mol.

Shown in Figure 23 and compared to Figure 6 TMP landscape against packing coefficient and Figure 17 CBN landscape against packing coefficient, the two lowest energy CBN:2TMP structures' packing coefficients are higher than CBN experimental structure and near to TMP orthorhombic experimental structure and are denser than the stable CBN:TMP structures.

The most stable CBN:2TMP co-crystal structures (Table 6) have a hydrogen bond between CBN and TMP. There is no packing problem in CBN:2TMP search, because one more small TMP improves the packing coefficient. Furthermore, some of the structures have some  $\pi$ - $\pi$  stacking between TMP and CBN.

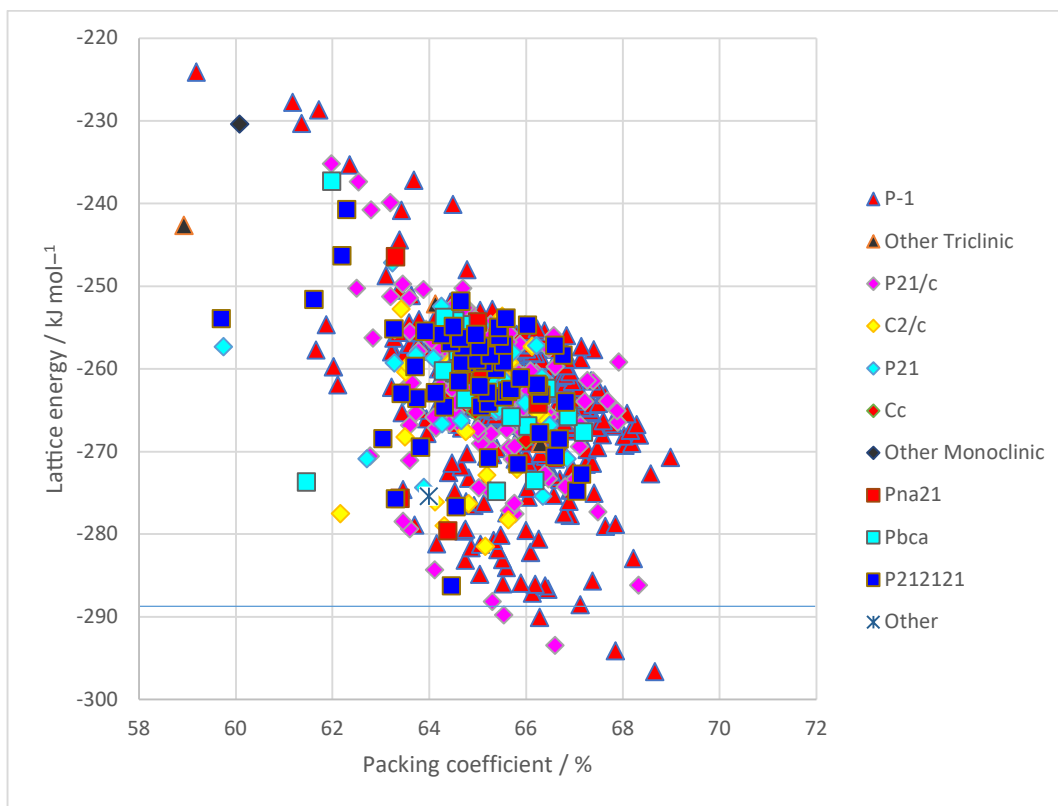


Figure 23 Lattice energy landscape (packing coefficient) for co-crystal 1CBN:2TMP search. Blue line:  $2E(\text{TMP, matching exp,rig}) + E(\text{CBN,exp,opt}) = -288.399 \text{ kJ/mol}$ . Blue line is the sum of experimental CBN and 2 TMP lattice energy after CrystalOptimizer.

There is also the balance between the intramolecular energy and intermolecular energy and the packing problem caused by the CBN hydrogen carbon conformation. Shown in Table 6, CTT32 is the GM structure in this search with  $U_{\text{inter}} = -303.94 \text{ kJ/mol}$  and  $\Delta E_{\text{intra}}(\text{CBN}) = 7.337 \text{ kJ/mol}$ . CTT347 has the lowest CBN intramolecular energy in Table 6 and it is the 5<sup>th</sup> most stable structure in this search with  $U_{\text{inter}} = -293.82 \text{ kJ/mol}$  and  $\Delta E_{\text{intra}}(\text{CBN}) = 4.032 \text{ kJ/mol}$ . As shown in Table 6, the CBN intramolecular energies of most CBN:2TMP co-crystal structures are near 10 kJ/mol, which is caused by the hydrogen carbon pentane tail conformation. Similar to CBN:TMP search, low energy structures do not have the CBN gas phase conformation, which means that there is a balance between the inter and intra molecular energy and CBN hydrogen carbon pentane tail conformation will lead to the packing problem in the limited search.

Table 6 The most stable CSP generated co-crystal structures of CBN:2TMP are named as the rank after point charge + FIT optimized by CrystalPredictor, for example, CTT1 is the global minimum structure after point charge + FIT optimization. Reference energy is:  $2E(\text{TMP, matching exp,rig}) + E(\text{CBN,exp,opt}) = -288.399 \text{ kJ/mol}$ . Co-crystallization energy is calculated by  $\Delta E_{\text{cc}} = E_{\text{latt}}(\text{co-crystal}) - 2E_{\text{latt}}(\text{TMP}) - E_{\text{latt}}(\text{CBN})$ . Structures that co-crystallization energies are lower than 0 are coloured as green.

\*Hydrogen bond graph sets: Ga,b,(c) G: C means chain, D means noncyclic dimers and other finite hydrogen bonds sets R means ring. a, b are number of acceptors and donors.

Label	Hydrogen bond graph sets*	Space Group	a /Å	b /Å	c /Å	$\alpha /^\circ$	$\beta /^\circ$	$\gamma /^\circ$	Density /gcm <sup>-3</sup>	$\Delta E_{\text{intra}}$ /kJmol <sup>-1</sup>	Intermolecular Lattice Energy /kJmol <sup>-1</sup>	Repulsion Dispersion Energy /kJmol <sup>-1</sup>	Lattice Energy /kJmol <sup>-1</sup>	co-crystallization energy/ kJmol <sup>-1</sup>
CTT32	D1,1(2) a	P-1	15.687	9.3902	11.8734	97.221	81.715	94.102	1.129	7.337	-303.94	-246.58	-296.60	-8.204
CTT332	D1,1(2) a	P-1	9.2575	11.9235	16.3007	76.902	92.856	97.357	1.1139	7.29	-301.39	-237.6	-294.1	-5.701
CTT514	D1,1(2) a	P21/n	9.552	26.4774	14.1671	90	80.294	90	1.0961	8.567	-302.02	-233.03	-293.45	-5.054
CTT130	D1,1(2) a	P-1	9.0984	9.2343	22.8779	86.258	91.91	68.06	1.0901	9.569	-299.63	-232.38	-290.06	-1.662
CTT374	D1,1(2) a	P21/n	16.5741	8.9165	25.4451	90	72.439	90	1.0798	4.032	-293.82	-220.87	-289.79	-1.389
CTT572	D1,1(2) a	P-1	8.1128	15.3993	23.1806	59.978	65.723	107.497	1.104	8.419	-296.94	-236.39	-288.52	-0.122
CTT370	D1,1(2) a	P21/c	13.6229	8.8472	30.8536	90	76.052	90	1.0727	7.345	-295.47	-229.69	-288.12	0.274
CTT752	D1,1(2) a	P-1	19.2787	8.7166	10.6885	96.766	93.016	86.325	1.0885	7.419	-294.54	-238.24	-287.12	1.278
CTT157	D1,1(2) a	P-1	16.6872	15.6076	7.9865	72.846	86.569	63.328	1.0938	8.168	-294.82	-233.7	-286.65	1.747
CTT349	D1,1(2) a	P212121	8.9341	30.3757	13.4632	90	90	90	1.0595	7.112	-293.34	-229.42	-286.23	2.171

### 5.2.3 2CBN:TMP co-crystal

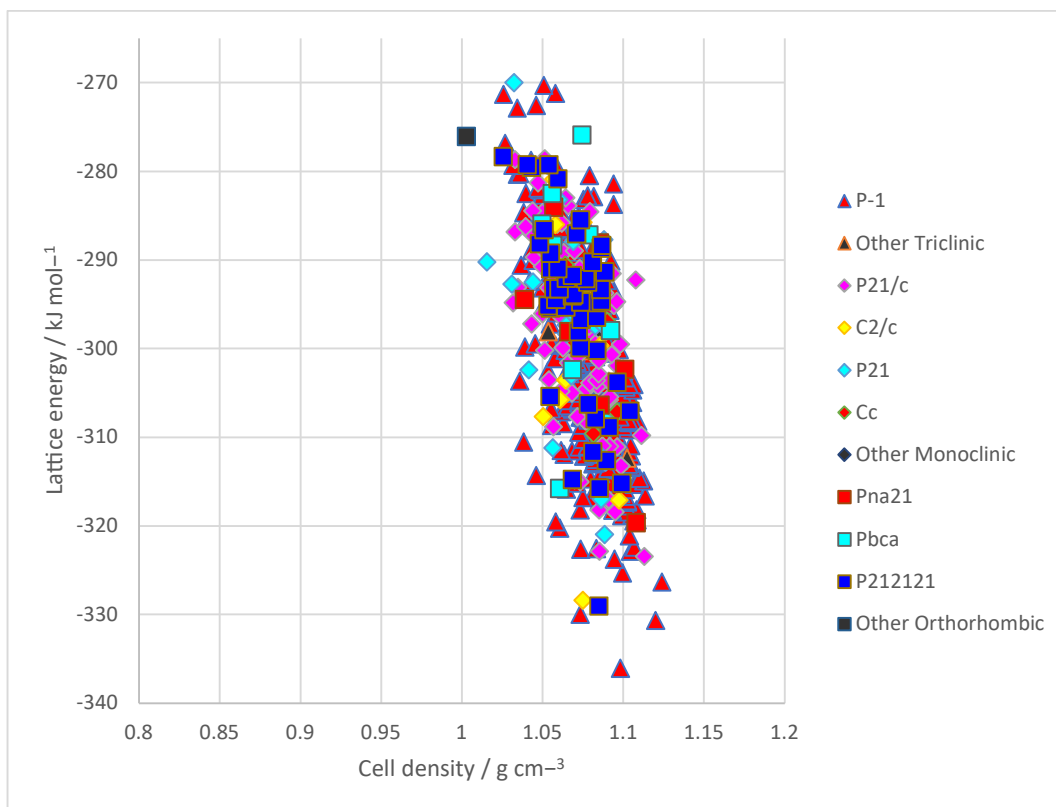


Figure 24 Lattice energy landscape (density) for co-crystal 2CBN:1TMP search. The line is not drawn like 1CBN:1TMP and 1CBN:2TMP co-crystal searches because all of those structures are higher than this cut-off energy which is  $Cutoff = 2E(TMP, matching\ exp, rig) + E(CBN, exp, opt) = -354.948\ kJ/mol$ .

Because of the limitation of the calculation resource, only 572 2CBN:TMP co-crystal structures have been optimized by CrystalOptimizier<sup>20</sup>.

In Figure 24 and Figure 25, there are no 2CBN:TMP stable co-crystal structures. The lowest lattice energy is  $-336.046\ kJ/mol$  in 572 structures optimized, whose co-crystallization energy is  $9.451\ kJ/mol$ . The co-crystallization analysis shows that no stable 2CBN:1TMP co-crystal structures have been found in this search.

Comparing Figure 21, Figure 22, Figure 23 and Figure 25, the packing coefficients of 2CBN:TMP are lower than CBN:TMP and CBN:2TMP, which shows the bad packing in the limited 2CBN:TMP search. The 2CBN:TMP structures have one or no hydrogen bonds (Table 7) and such hydrogen bonds are between the hydroxyl groups in the CBN, which is similar to CBN search. No structures with two CBN molecules forming OH...N hydrogen bonds to the same TMP molecule were generated.

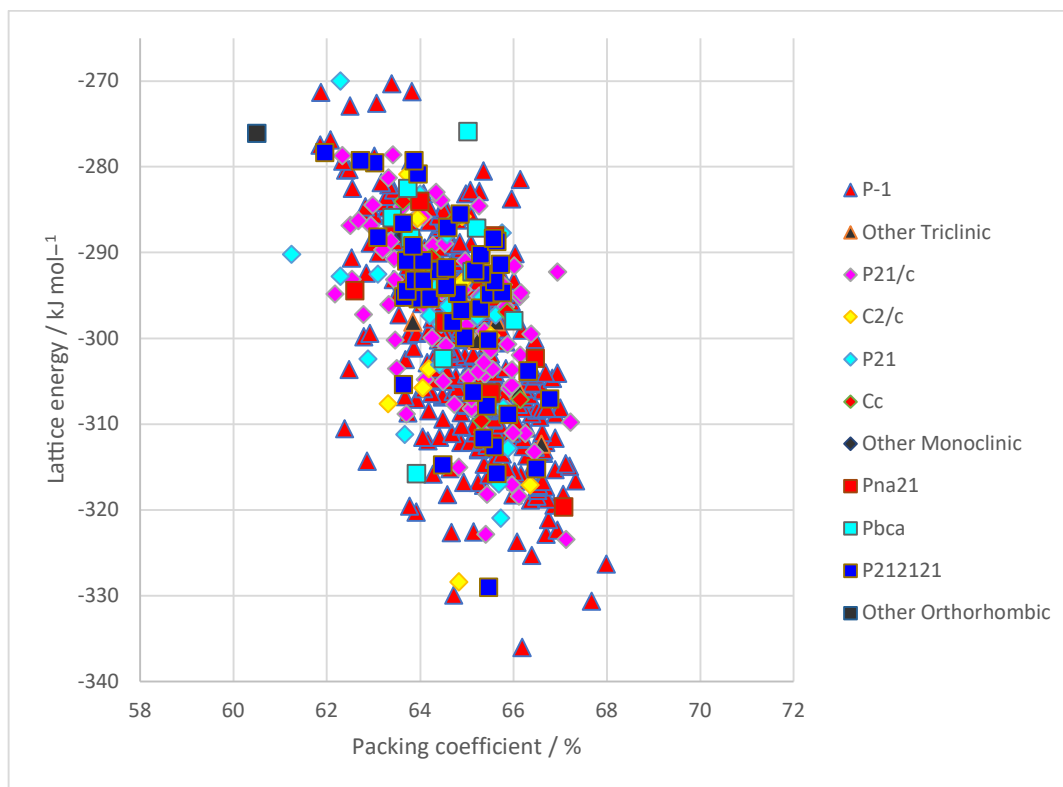


Figure 25 Lattice energy landscape (packing coefficient) for co-crystal 2CBN:1TMP search. The line is not drawn like 1CBN:1TMP and 1CBN:2TMP co-crystal searches, because all of those structures are higher than this cut-off energy, which is  $Cutoff = 2E(TMP, matching\ exp, rig) + E(CBN, exp, opt) = -354.948\ kJ/mol$ .

As shown in Figure 26 and Table 7, the GM structure CCT166 has a hydrogen bond between hydroxyl groups in CBN, which is similar to the low energy structures in the CBN search results. But this hydrogen bond structure is poorer than CANNOL, so there is more intermolecular interaction for CCT166 than CANNOL. Furthermore, this structure has a significant lower density than other 1CBN:1TMP and 1CBN:2TMP and CANNOL, which is also shown in the voids analysis. In general, hydrogen bonds can stabilize the intermolecular interaction, but it is not enough for the limited 2CBN:TMP co-crystal search, meanwhile, the low density and packing coefficient means it is not stabilised by dispersion energy. Especially compared with CANNOL, CANNOL will have a more stable structure than GM structure in the search. In GM structure, the CBN molecules are not in the gas phase conformation but as shown in Figure 26, one molecule has the high intramolecular energy caused by the hydroxyl group out of the plane and another one is caused by the hydrogen carbon pentane tail.

As to the second lowest structure, this structure has a higher density than GM, but it is still smaller than CANNOL. Furthermore, there is no hydrogen bond, which causes the intermolecular higher than others that have hydrogen bond.

As to the 3<sup>rd</sup> lowest structure shown in Figure 26, there is a hydrogen bond between TMP and CBN, which is similar to CBN:TMP and CBN:2TMP search, but its intermolecular energy is still higher than GM structure too much. Furthermore, there are the lowest CBN intramolecular energy (3.41 kJ/mol and 2.76 kJ/mol), which are near to gas phase conformation.

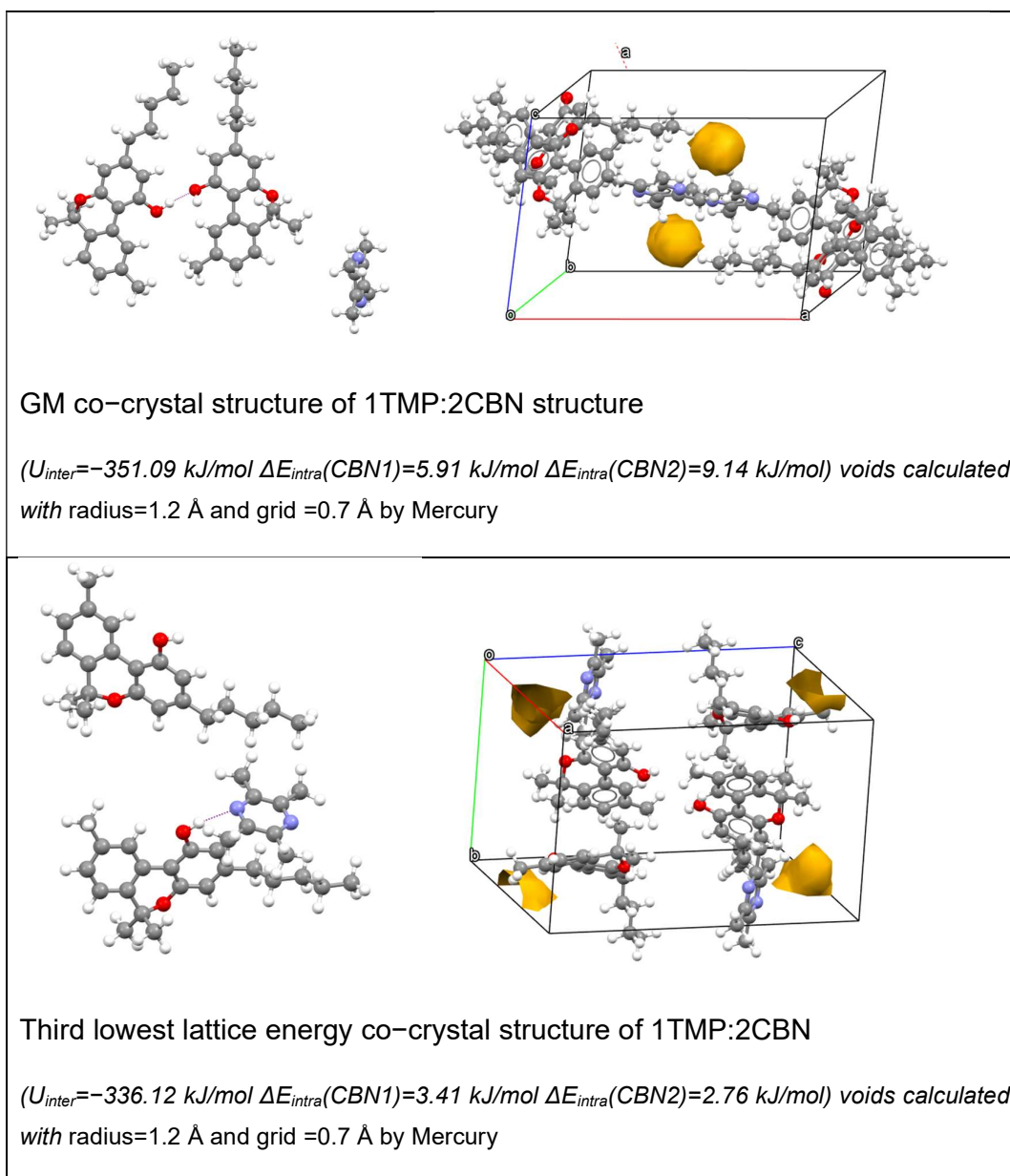


Figure 26 Crystal cell and voids for the lowest energy 1CBN: 2TMP co-crystal structures. The lowest structure contains H-bond and  $\pi$ - $\pi$  stacking.



Thus, there is a complex balance among intramolecular energy, hydrogen bond, dispersion terms in determining the relative stability of the co-crystals. A CSP for 2CBN:TMP is a big challenge, and the CSP was very limited CBN...TMP...CBN may be a good hydrogen bond structure that could have a low lattice energy if this motif can pack densely. There is not any CBN...TMP...CBN structure in the search. It could be that the assumptions in the CSP prevent such a structure being generated, or it may be impossible for such a structure to form a dense crystal.

Table 7 The most stable CSP generated co-crystal structures of 2CBN:TMP are named as the rank after point charge + FIT optimized by CrystalPredictor, for example, CCT1 is the global minimum structure after point charge+ FIT optimization. Reference energy is:  $2E(\text{TMP, matching exp,rig}) + E(\text{CBN,exp,opt}) = -354.948 \text{ kJ/mol}$ . Co-crystallization energy ( $\Delta E_{cc}$ ) is calculated by  $\frac{E_{latt}(\text{co-crysta}) - E_{latt}(\text{TMP})}{2} - E_{latt}(\text{CBN})$ . To compare stabilization of co-crystal with different stichometry, the co-crystallization energy is calculated with respect to 1 mol CBN.

\*Hydrogen bond graph sets: Ga,b,(c) G: C means chain, D means noncyclic dimers and other finite hydrogen bonds sets R means ring. a, b are number of acceptors. and donors.

Label	Hydrogen bond graph sets*	Space Group	a /Å	b /Å	c /Å	$\alpha /^\circ$	$\beta /^\circ$	$\gamma /^\circ$	Density /gcm <sup>-3</sup>	Intra energy mol1/ kJmol <sup>-1</sup>	Intra energy mol2/ kJmol <sup>-1</sup>	Intermolecular Lattice Energy /kJmol <sup>-1</sup>	Repulsion Dispersion Energy /kJmol <sup>-1</sup>	Lattice Energy /kJmol <sup>-1</sup>	co-crystallization energy/kJmol <sup>-1</sup>
CCT166	D1,1(2) a	P-1	17.7708	11.1877	12.326	80.528	82.874	71.826	1.0981	5.90548	9.13789	-351.09	-308.7	-336.046	9.451
CCT4	no	P-1	26.7141	14.1672	6.1314	96.676	84.632	78.812	1.12	9.02982	6.88602	-346.57	-324.96	-330.653	12.1475
CCT496	D1,1(2) a	P-1	10.6962	12.4294	18.4566	92.995	74.404	84.079	1.0733	3.40723	2.75947	-336.12	-300.57	-329.952	12.498
CCT383	D1,1(2) a	P212121	22.0799	22.5574	9.3079	90	90	90	1.0847	7.54903	4.47466	-341.06	-285.38	-329.035	12.9565
CCT399	D1,1(2) a	I2/a	29.3132	12.164	26.4437	90	82.925	90	1.0748	8.92431	15.40568	-352.73	-289.68	-328.399	13.2745
CCT88	no	P-1	16.899	22.1484	6.1459	94.443	91.004	77.27	1.1239	7.88806	8.5437	-342.77	-330.38	-326.337	14.3055
CCT74	D1,1(2) a	P-1	10.687	15.6755	14.4577	87.869	84.862	71.448	1.0994	6.26021	8.67324	-340.25	-306.97	-325.316	14.816
CCT569	D1,1(2) a	P-1	20.1587	12.4868	11.7929	102.709	72.705	63.079	1.0944	5.2218	8.48173	-337.46	-297.67	-323.756	15.596
CCT221	no	P21/a	17.8262	40.0502	6.3714	90	83.366	90	1.1129	7.44776	7.63528	-338.53	-323.91	-323.446	15.751
CCT125	no	P-1	15.6765	16.7139	10.2489	99.023	91.234	59.367	1.1044	5.53356	8.11866	-336.51	-322.96	-322.857	16.0455
CCT60	D1,1(2) a	P21/c	23.1693	23.313	8.6093	90	85.223	90	1.0851	3.36401	10.91374	-337.13	-295.27	-322.851	16.0485
CCT352	D1,1(2) a	P-1	14.3591	16.5379	11.1922	97.756	86.142	63.945	1.0736	10.09964	7.30713	-340.04	-286.63	-322.632	16.158
CCT143	D1,1(2) a	P-1	16.7624	12.1909	13.74	65.37	114.497	101.684	1.0832	4.96675	4.57113	-332.1	-282.5	-322.561	16.1935
CCT30	D1,1(2) a	P-1	29.1483	7.0182	11.2824	80.354	90.078	92.933	1.1065	6.77786	4.88607	-334.04	-319.44	-322.375	16.2865

### 5.3 Conclusion

There are 11 CBN:TMP and 7 CBN:2TMP structures generated whose lattice energies are more stable than those of the components. GM and 2<sup>nd</sup> and 3<sup>rd</sup> lowest lattice energy 1CBN:2TMP co-crystal structures have lower co-crystallization energy and higher density than 1CBN:1TMP GM. There is no 2CBN:TMP structure whose co-crystallization energy is lower than 8 kJ/mol. In the limited search, after ranking with co-crystallization energy, 2CBN:TMP GM structure is not in the 100 lowest energy structures. Thus, calculations in this thesis predict that the CBN:TMP ratios 1:1 and 1:2 are the experimental stoichiometries.

#### 5.3.1 Blind Test submission of structures in order of $\Delta E_{cc}$ .

The requirements of the Blind Test were to predict the stoichiometries of the two observed co-crystals which had different stoichiometries (with fewer than 4 molecules in the formula unit) and submit two ranked lists of 100 and 1500 structures. All generated co-crystal structures (including 1CBN:1TMP, 1CBN:2TMP, and 2CBN:1TMP) are ranked by co-crystallization energy using equation Equation 4 to get normalized energy respecting CBN. Only one 2CBN:1TMP structure was submitted, because the lowest 2CBN:1TMP structure is ranked 114. In the lowest 99 structures, there are 65 1CBN:1TMP co-crystals and 34 1CBN:2TMP co-crystals. The lowest 3 structures are 1CBN:2TMP co-crystals with co-crystallization = -8.204 kJ/mol, -5.701 kJ/mol, -5.054 kJ/mol, but the rest of the list is dominated by 1CBN:1TMP. The co-crystallization energies of 100 lowest structures are lower than 9 kJ/mol, based on the study<sup>24</sup>, all 100 structures are potentially stable.

#### 5.3.2 Limitations of methods and confidence.

Firstly, the biggest concern is that the CBN experimental structure (CANNOL) was not found in the CBN search, so it is necessary to assess whether the CSP search in this thesis can generate all stable structures and not ignore them after point charge + FIT optimization in the limited time although the CrystalOptimizer<sup>20</sup> computational method has been proved suitable. Based on the polar hydrogen bond analysis and the balance between the intermolecular and intramolecular energy, the crystal structure and the lattice energy after CrystalPredictor<sup>18,19</sup> will

have high error because of the balance hydroxyl group position in the co-crystal and CBN search.

### 5.3.3 Completeness of search and energy rankings

These searches were not completed as the low energy structures were not found lots of times. It is usually hoped to have all the lowest energy structures generated more than 20 times in the search, which is not feasible for multicomponent systems. But each co-crystal search have generated 1,000,000 unique structures after point charge + FIT potential<sup>85,86,87</sup> optimization. Similarly, the CBN search is not enough. All those uncompleted searches are caused by the flexible CBN, and more details will be discussed in the next chapter.

In the 6<sup>th</sup> Blind Test, the co-crystal experimental structure was in the second batch of 1,000,000 structures generated in the search<sup>37</sup> with fewer degree freedoms. As shown in the CBN conformational analysis, there are two so different conformations in the experimental asymmetric unit that those degrees of freedom cannot be ignored. Because of the time limitation (the deadline for submission is June 2021) and many degrees of freedom, the co-crystal searches are not completed (the lowest structure just has been found once).

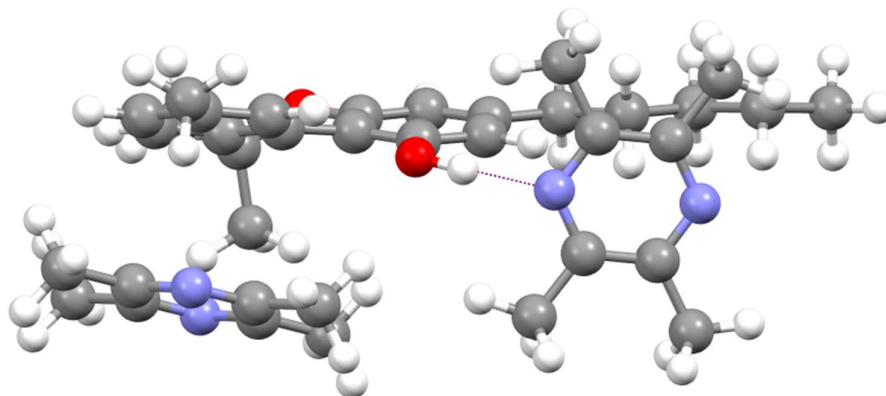
In this search, the lattice energy was calculated without free energy. To accurate calculation of  $E_{latt}$  model, the free energies calculation or PBE-TS/MBD\* refinements should be done.

### 5.3.4 What are submitted

The 1500 structures landscape and 100 most likely structures by the ranking of those 3 searches were submitted. Confidence level for this submission is poor, as there is no possible CBN...TMP...CBN with good hydrogen bonds and the CBN tails in different conformations that nicely wrapped around TMP. But in the co-crystal searches, there are hydrogen bonds between CBN and TMP in the lowest energy structures for 1CBN:2TMP and 1CBN:1TMP and the lowest 2CBN:1TMP structure. Some structures where CrystalOptimizer<sup>20</sup> makes hydrogen bonds were generated, but far too few for confidence that the correct structure is generated. The inability of the CBN search to find the experimental structure CANNOL supports this hypothesis that this search is not sufficiently complete to generate a hydrogen bonded structure.

### 5.3.5 Co-crystal confidence

As to 1CBN:1TMP and 2TMP:1CBN searches, most low energy 1CBN:1TMP structures are formed by CBN...TMP dimer as shown in Figure 27. Their co-crystallization energies are lower than or close to 0 calculated using the CBN experimental energy shown in Chapter 4. The lowest co-crystallization energy with a CBN...CBN hydrogen bond for CCT is 9.451 kJ/mol shown in Table 7.



*Figure 27 1CBN:2TMP GM structure with the hydrogen bond and  $\pi$ - $\pi$  stacking. The distance of the hydrogen bond is 2.8 Å, hydroxyl group torsion angle is  $-6.99^\circ$ , which is near  $0^\circ$  in the ring plane.*

The lowest 5 co-crystallization energy structures (CTT32, CTT332, CTT514, CT279, CT93) have good hydrogen bonds and  $\pi$  stacking and close packing. Comparing the CBN:TMP co-crystals, CBN:2TMP co-crystals usually have closer packing and  $\pi$  stacking, which lead to the 3 lowest crystallization energy structures are CBN:2TMP co-crystal.

But shown in Figure 27 most of the submitted structures do not have the gas phase CBN conformation caused by the hydroxyl group or the hydrogen carbon pentane tail. So, there is a balance between the CBN conformation (intramolecular energy) and packing (intermolecular energy). It seems difficult to generate structures for CBN or its co-crystal that have a good packing coefficient and strong hydrogen bond because it is an awkward shape. This may be why CANNOL is  $Z'=2$  with a high energy conformation with hydroxyl group out of the ring plane to form a hydrogen bond. This search method is inadequate for finding the CBN experimental structure CANNOL, which leads to low confidence in the

co-crystal results. It is hoped to find the CBN...TMP...CBN structure, but there is no this structure. Because of the time limitation, after submission of the Blind Test, the assumptions in those searches were analysed to improve confidence.

Appendix Submission to CCDC for the co-crystal challenge

The most likely stoichiometries for two co-crystals of cannabiniol (CBN) and tetramethyl pyrazine (TMP) are CBN:2TMP and CBN:TMP. The list of up to 1500 co-crystal structures submitted is stored on the Crystal Navigator Database and is available from Dr. Louise Price. The list of 100 structures, ranked in the likelihood of observation is in Table 8.

Table 8 The submitted 100 co-crystal structures including the label, space group, density and co-crystallization energy.

Label	Density /gcm <sup>-3</sup>	co-crystallization energy kJ/mol	Label	Density /gcm <sup>-3</sup>	co-crystallization energy kJ/mol	Label	Density /gcm <sup>-3</sup>	co-crystallization energy kJ/mol
dfCTT32	1.129	-8.204	dfCTT22	1.1209	2.226	dfCT230	1.1072	6.323
dfCTT332	1.1139	-5.701	dfCT601	1.0915	2.308	dfCT793	1.1047	6.331
dfCTT514	1.0961	-5.054	dfCTT754	1.0911	2.32	dfCTT634	1.0791	6.523
dfCT279	1.0961	-3.918	dfCTT643	1.0806	2.426	dfCT50	1.0982	6.527
dfCT93	1.0837	-3.666	dfCTT262	1.0819	2.473	dfCT325	1.1187	6.671
dfCT5	1.1025	-2.24	dfCT105	1.0866	2.712	dfCT472	1.091	6.71
dfCT59	1.1033	-2.17	dfCTT569	1.1083	2.727	dfCTT793	1.0675	6.761
dfCT60	1.0999	-2.043	dfCTT544	1.069	3.533	dfCT920	1.0915	6.786
dfCT814	1.088	-1.884	dfCT61	1.1013	3.737	dfCT1	1.1281	6.867
dfCTT130	1.0901	-1.662	dfCT711	1.0985	4.04	dfCTT539	1.0714	6.904
dfCT459	1.0908	-1.632	dfCTT122	1.0522	4.086	dfCT697	1.0942	6.909
dfCTT374	1.0798	-1.389	dfCT374	1.1193	4.264	dfCT751	1.0829	6.929
dfCT229	1.0949	-1.099	dfCT741	1.1006	4.317	dfCTT91	1.0684	7.163

dfCT119	1.0818	-1.043	dfCTT760	1.081	4.36	dfCTT138	1.0506	7.233
dfCT67	1.1258	-0.241	dfCT168	1.0964	4.52	dfCT719	1.0938	7.346
dfCT674	1.0971	-0.194	dfCT104	1.1296	4.607	dfCT151	1.0777	7.358
dfCTT572	1.104	-0.122	dfCT880	1.1224	4.615	dfCT370	1.0851	7.506
dfCT316	1.0921	0.07	dfCT807	1.0764	5.011	dfCTT405	1.0743	7.632
dfCT595	1.061	0.09	dfCTT640	1.0672	5.24	dfCT78	1.0941	7.775
dfCTT370	1.0727	0.274	dfCTT404	1.0765	5.321	dfCTT355	1.0879	7.806
dfCT317	1.1151	0.306	dfCT28	1.1054	5.338	dfCT667	1.0946	7.854
dfCT4	1.1182	0.655	dfCT106	1.0643	5.402	dfCT287	1.1017	7.886
dfCT113	1.0529	1.099	dfCTT165	1.12	5.474	dfCT74	1.0867	8.152
dfCTT752	1.0885	1.278	dfCT155	1.1177	5.477	dfCT23	1.1015	8.209
dfCT135	1.1132	1.504	dfCT273	1.088	5.538	dfCT545	1.0823	8.24
dfCT871	1.0783	1.556	dfCT175	1.1157	5.637	dfCT397	1.1043	8.241
dfCT8	1.12	1.631	dfCT26	1.102	5.655	dfCTT783	1.076	8.284
dfCT344	1.0716	1.705	dfCT353	1.0836	5.749	dfCT192	1.1332	8.501
dfCTT157	1.0938	1.747	dfCT68	1.0896	5.773	dfCT187	1.0853	8.683
dfCT929	1.0883	2.025	dfCT236	1.081	5.884	dfCTT222	1.062	8.781
dfCT52	1.1105	2.049	dfCTT856	1.0849	6.185	dfCT165	1.0714	8.797
dfCTT349	1.0595	2.171	dfCT126	1.0815	6.206	dfCTT532	1.058	8.831
						dfCCT166	1.0981	9.451



## 6 Analysis of limitations of the CSP method used for CBN and TMP and co-crystals

Firstly, in Chapter 3 and Chapter 5, the rigid molecule assumption was used for TMP CSP. This chapter assess whether this assumption is reasonable. Secondly, in Chapter 4 CBN experimental structure (CANNOL) had not been found and so in Chapter 5 the co-crystallization energy is calculated with the lattice energy of the experimental structure optimized by the same model as used in the CSP (Chapter 4). This chapter investigates why the CSP search method did not find the CBN experimental structure (CANNOL), to suggest improvements in the CSP methodology that will provide more confidence in the prediction of any crystal structure containing CBN.

### 6.1 Rigid methyl group assumption for TMP

To assess whether the rigid assumption is suitable for TMP, the relaxed conformational energy has been scanned for TMP versus 1 methyl group torsion angle.

#### 6.1.1 A flexible CSP for TMP

As shown in Figure 28, the rotation of one methyl group will not lead to much intramolecular energy change (less than 3.5 kJ/mol).

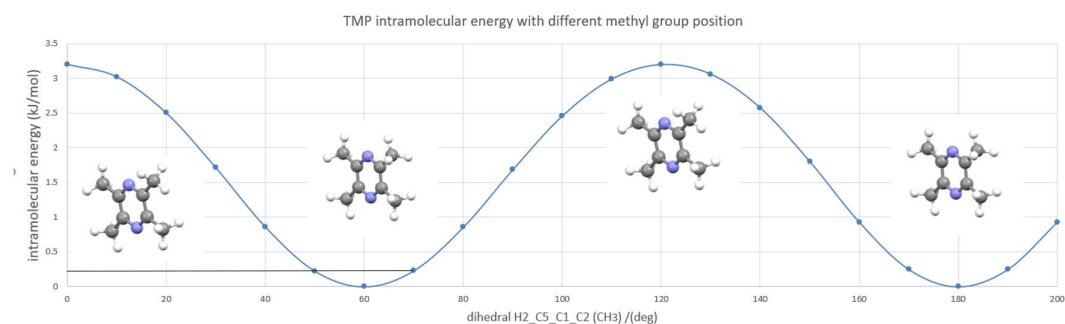


Figure 28 Relaxed conformational energy scan of one methyl group in TMP ( $H2\_C5\_C2\_C1$ ) at PBE0 6–31G(d,p). The black line shows that there is very little molecular energy difference (no more than 0.5 kJ/mol) between the conformations with the methyl dihedral between  $50^\circ \sim 70^\circ$ . The neighbouring methyl group shows a small change during the scan showed that the interaction among neighbour methyl group is weak.

The neighbouring methyl groups will affect each other. Hence TMP LAM grid of energies is formed by those 4 methyl groups. These methyl groups are ranging over 120° because of the symmetry of the methyl group.

Table 9 Torsion group ranges in CrystalPredictor LAM generation for TMP flexible search

Torsion angle	Start point(°)	interval (°)	Finish point(°)	range
H1_C5_C2_C1	15.0°	30.0°	105.0°	0°–120°
H4_C6_C1_C2	15.0°	30.0°	105.0°	0°–120°
H7_C7_C3_C4	15.0°	30.0°	105.0°	0°–120°
H10_C8_C4_C3	15.0°	30.0°	105.0°	0°–120°

\*Each LAM point is valid over a 30° range, so the point calculated at 15° covers 0°–30° degrees and so on. So in this case, the range of LAM points is 15°–105°, but these methyl groups are ranging over 120°.

There are two steps in CSP process (section 2.4), so whether the rigid assumption is reasonable will be assessed for those two steps separately.

### 6.1.2 Comparing TMP rigid search and flexible search generated in CrystalPredictor

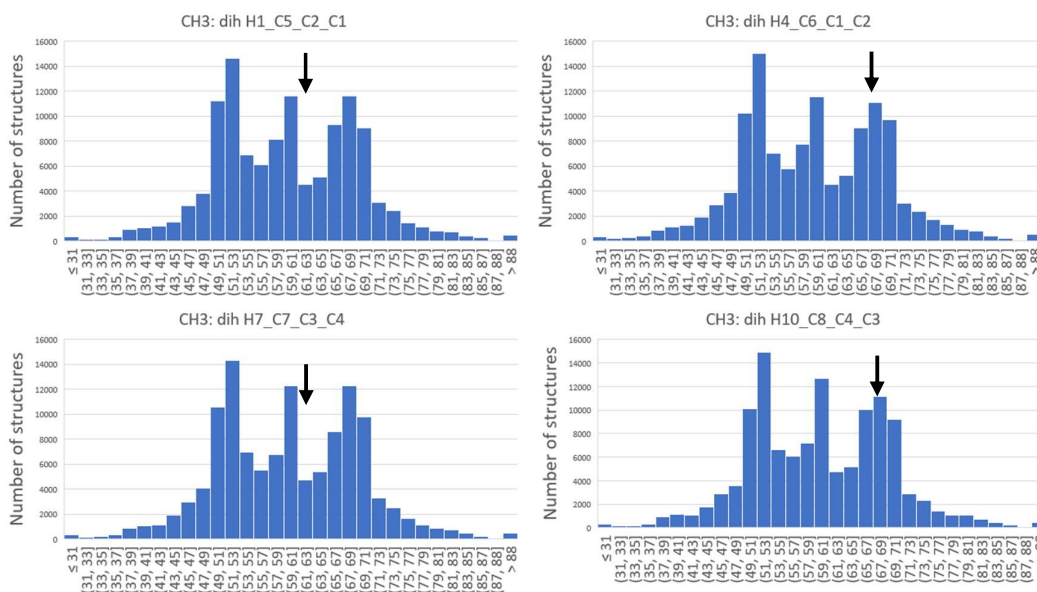


Figure 29 The torsion angles of the TMP CH3 groups (H7\_C7\_C3\_C4, H10\_C8\_C4\_C3 H1\_C5\_C2\_C1, H4\_C6\_C1\_C2,) for generated structures after the CrystalPredictor flexible search and are optimized using point charge + FIT. The black arrows denote the conformation in the rigid search.

Firstly, whether the rigid assumption is reasonable for the generation and rough optimization step in CrystalPredictor<sup>18,19</sup> will be assessed. CrystalPredictor2.4.3.2 was used to generate both rigid and flexible CSP searches, and optimize them using the FIT potential<sup>185,86,87</sup> model and point charge model.

The distribution of angles for each methyl group in the flexible search after optimization is shown in Figure 29, which shows the CH<sub>3</sub> torsion angles of the flexible search are around 50° ~ 70°. In the rigid search, the methyl groups were kept at the TMP gas phase conformation. Figure 28 shows little molecular energy difference (less than 0.5 kJ/mol) when a methyl group is between the 50° ~ 70° relaxed optimization. Therefore, keeping methyl groups rigid during the generation step in CrystalPredictor is reasonable and will not lead to too many errors.

#### 6.1.3 Comparing TMP rigid search optimized by CrystalOptimizer for flexible methyl groups and DMACRYS with rigid methyl groups

Next, whether the rigid assumption is reasonable for the accurate optimization step in CrystalOptimizer<sup>20</sup> will be assessed.

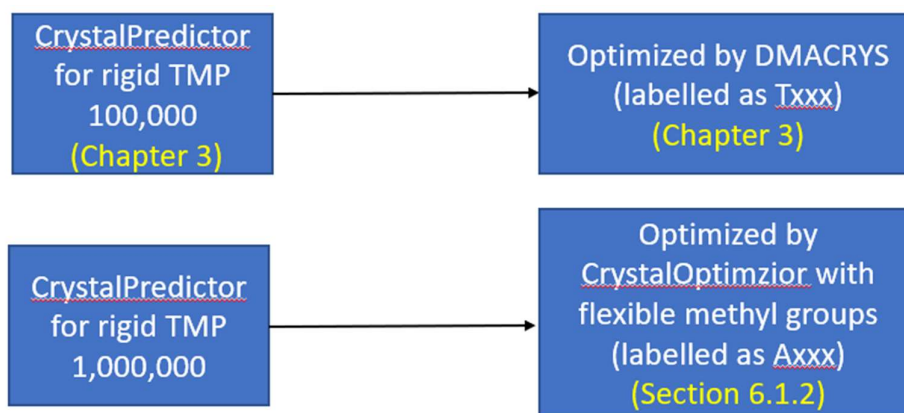


Figure 30 Workflow for comparing TMP rigid search (chapter 3) optimized by CrystalOptimizer for flexible methyl groups and DMACRYS with rigid methyl groups.

Based on the conclusion of the last generation part in CrystalPredictor, the rigid model is reasonable for the first global generation. So, in the accurate but expensive optimization stage, crystal structures generated in the rigid search will be optimized by CrystalOptimizer (marked as Axxx) with flexible methyl groups and DMACRYS<sup>39</sup> with rigid methyl groups (marked as Txxx, in chapter 3).

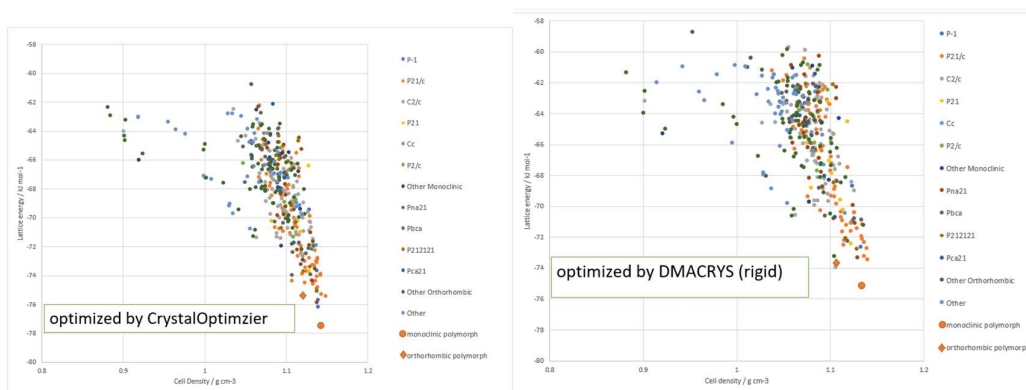


Figure 31 Lattice energy landscape summary plot for TMP from  $Z'=1$  rigid molecule search after CrystalOptimizer (left) and after DMACRYS rigid refinement (right). The orthorhombic polymorph is depicted by the orange diamond, which is lattice energy minimum structure from MPYRAZ02, and the GM structure with monoclinic polymorph is depicted by orange round, which is approximately a supercell of MPYRAZ03.

In Figure 31, rigid searches with flexible and rigid optimization find the two experimental structures in the CSD<sup>98</sup> database separately, as shown in Figure 31. In the rigid search with flexible optimization, the GM (A1193) matches the cell monoclinic polymorph (MPYRAZ03 ( $P2_1/c$ )) in CSD<sup>98</sup> with  $-77.458$  kJ/mol, and the 6th lowest energy structure (A362) matches the cell of orthorhombic polymorph (MPYRAZ02) in CSD<sup>98</sup> with  $-75.402$  kJ/mol. In the rigid search with rigid optimization (Chapter 3), the GM (T11) matches cell MPYRAZ03 ( $P2_1/c$ ) in CSD<sup>98</sup> with  $-75.06$  kJ/mol and 2nd lowest energy structure (T131) matches cell MPYRAZ, MPYRAZ01, MPYRAZ02 (Pbcu) in CSD<sup>98</sup> with  $-73.95$  kJ/. Between rigid and flexible optimization results there is about a  $1.5\sim 2.5$  kJ/mol difference. There is a bit of reranking between flexible and rigid optimization. But the energy differences between those two experimental structures in those two optimization results are similar (2 kJ/mol and 1 kJ/mol) and small. The difference between flexible search and rigid search is so small that the use of rigid TMP in the co-crystal CSP is a reasonable approximation.

To compare the rigid optimization result and flexible optimization result, the  $RMSD_{15}$  and  $RMSD_1$  had been calculated in the Mercury for the generated monoclinic polymorph (A1193 and T11) and orthorhombic polymorph (A362 and T131) in those flexible and rigid optimization results.

In Figure 32, the monoclinic and orthorhombic polymorphs in both searches are similar to each other ( $RMSD_{15}$  and  $RMSD_1$  are no more than  $0.11$  Å), and the

difference between TMP conformations is the hydrogen atoms (methyl groups position), but this difference is also small (no more than  $10^\circ$ , which will lead to the energy difference lower than 1 kJ/mol for each methyl group shown in Figure 28. To TMP CSP search, the rigid optimization and flexible optimization will not lead too many differences and the rigid assumption in the optimization stage is reasonable.

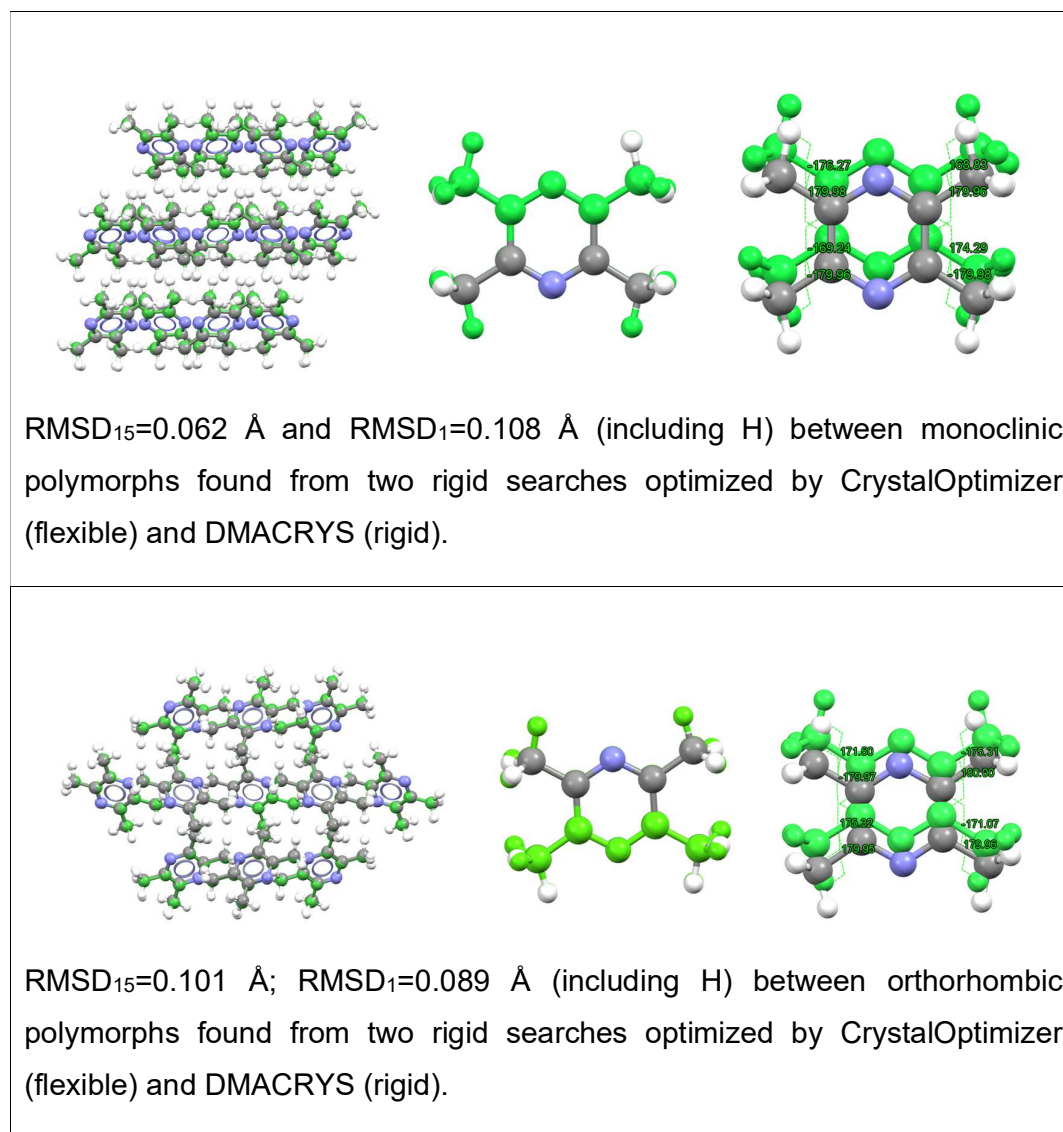


Figure 32 RMSD<sub>15</sub> and RMSD<sub>1</sub> between the generated structures matching the experimental structures optimized by CrystalOptimizer with 4 methyl groups and DMACRYS.

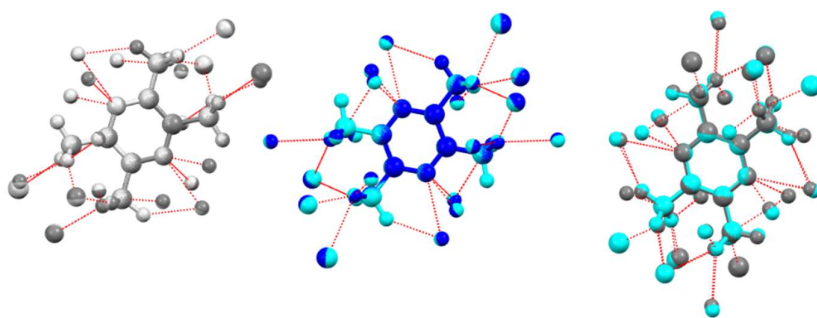
#### 6.1.3.1 PBE-TS and single point optimization

The experimental structure and the predicted structures are also optimized by the periodic electronic structure method using CASTEP with the PBE + TS. Periodic electronic structure method ( $\Psi_{\text{cry}}$ ) relaxes all atoms to optimize the crystal

structure (section 2.3), which is different from two level approximation in CrystalOptimizer<sup>20</sup> (section 2.4.2.1).  $\Psi_{\text{cry}}$  is more accurate than  $\Psi_{\text{mol}}$  method, thus, periodic electronic structure method can be used to analyze if there is something missed in the molecular conformation optimization in crystal.

The RMSD<sub>1</sub> and neighbor atoms have been calculated and shown in the Mercury to compare the generated structure matching orthorhombic polymorph (A362) and experimental structure (MPYRAZ02) optimized by CASTEP and CrystalOptimizer to access if the rigid assumption is reasonable for the pyrazine ring, which was used in the rigid searches and both rigid and flexible optimization.

GM structure in both searches is close to the (2,1,1) supercell of MPYRAZ03. MPYRAZ03 has no coordinates in CSD<sup>98</sup>, which has been discussed in chapter 3. Thus, there is no comparison for this polymorph.



Structure	A362	Optimized experimental structure (MPYRAZ02)	RMSD <sub>1</sub> / Å
Optimized by:			
CASTEP PBE-TS	in Light grey	grey	0.00577
CrystalOptimizer	Blue	Cyan	0.00228
RMSD <sub>1</sub> / Å	0.0131	0.0132	

Figure 33 The comparison among A362 and experimental structure optimized by CASTEP with PBE-TS and CrystalOptimizer with PBE0/6-31G(d,p) +FIT potential.

As can be shown from the table in Figure 33, the RMSD<sub>1</sub> between the experimental structure and A362 optimized by CASTEP in PBE-TS is 0.006 Å and optimized by CrystalOptimizer in PBE0 6-31G(d,p) is 0.002 Å. This small

RMSD<sub>1</sub> shows that A362 and T131 match the experimental structure (MPYRAZ02). Furthermore, the RMSD<sub>1</sub> between the experimental structure optimized by CASTEP and CrystalOptimizer is 0.0132 Å, and the RMSD<sub>1</sub> between A362 optimized by CASTEP and CrystalOptimizer is 0.0131 Å. This small RMSD<sub>1</sub> shows that there is not much difference between the  $\Psi_{\text{cry}}$  and  $\Psi_{\text{mol}}$  optimization methods for TMP, and  $\Psi_{\text{mol}}$  method is good enough for TMP.

Table 10 Cell parameters and details of A1193, A362 and MPYRAZ02 optimized by PBE-TS or single point optimization.

	Space group	Z'	a/Å	b/Å	c/Å	$\beta^\circ$	Lattice energy/ (kJ/mol)	RMSD <sub>1</sub> with exp/Å
MPYRAZ03	P2 <sub>1</sub> /c	0.5	5.481(2)	6.936(3)	10.302(3)	99.90(3)		
T11	P2 <sub>1</sub> /c	1	11.3075	6.7404	10.5452	82.881	-75.06	
T11 + PLATON	P2 <sub>1</sub> /c	0.5	5.6537	6.7404	10.5452	98.0346		
A1193	P2 <sub>1</sub> /c	1	11.2077	6.8098	10.4740	98.0346	-77.46	
A1193 + PLATON	P2 <sub>1</sub> /c	0.5	5.604	6.8098	10.4740	98.0346		
A1193+CASTEP	P1	4	11.092	6.78	10.07	99.98	0	
A1193+CASTEP+PLATON	P2 <sub>1</sub> /c	0.5	5.546	6.78	10.07	99.98	0**	
MPYRAZ02	Pbca	0.5	8.331(6)	9.225(5)	10.148(9)	90		
MPYRAZ02 +DMACRYS	Pbca	0.5	8.5816	9.1758	10.2507	90	-77.21 *	
MPYRAZ02+castep	P1	4	8.237	9.148	9.953	90		
MPYRAZ02+CASTP+PLATON	Pbca	0.5	8.237	9.148	9.953	90	3.60**	
T131	Pca2 <sub>1</sub>	1	10.107	9.2944	8.7085	90	-73.95	0.172
T131 after PLATON	Pbca	0.5	8.7085	9.2944	10.107	90		0.172
A362	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	1	9.2125	10.2681	8.5455	90	-75.40	0.112
A362 after PLATON	Pbca	0.5	8.5455	9.2125	10.2681	90		0.112
A362+CASTEP	P1	4	9.153	9.947	8.235	90	3.61**	0.093
A362+CASTEP+PLATON	Pbca	0.5	8.235	9.153	9.947	90		0.093

\*Minimum with the experimental conformation does not include the conformational energy penalty. In other words, in this case, the TMP molecule is not in the gas phase conformation.

\*\* The lattice energies of crystal structures optimized by the CASTEP are relative to the lattice energy of A1193 optimized by CASTEP in PBE-TS.

More details have been listed in Table 10 including cell parameters, lattice energy and RMSD<sub>15</sub> compared to the experimental structure. As shown in Table 10, the rigid and flexible optimization will not lead to too many errors in the crystal structure cell parameters. The intramolecular energy differences for both polymorphs are no more than 1 kJ/mol, and the intermolecular energy differences are no more than 4 kJ/mol. Furthermore, the energy difference between the global minimum (monoclinic polymorph shown in chapter 3) and orthorhombic polymorph is 3.6 kJ/mol by  $\Psi_{\text{crys}}$ (PBE-TS) and 2.2 kJ/mol (rigid optimization) or 2.0 by the  $\Psi_{\text{mol}}$  method, with both giving a reasonable reproduction of the crystal structures. Thus, optimized by CASTEP and CrystalOptimizer will not have too much difference for experimental structure in the landscape.

When the generated structure is optimized with a different method, the cell parameters changes. As shown in Table 10, generated structure optimized by CrystalOptimizer and CASTEP is similar to the experimental structure optimized by CrystalOptimizer and CASTEP respectively. But after DMACRYS<sup>39</sup> optimization for generated structure and experimental structure, there is an obvious difference for cell parameters.

## 6.2 CSP for TMP experimental structure without atomic coordinate in CSD

Table 11 TMP monoclinic polymorphs crystal parameters. MPYRAZ03 is the experimental crystal structure in CSD. A1193 is the structure generated by CrystalPredictor with rigid search and optimized by CrystalOptimizer with PBE0/6-31G(d,p). T11 is the structure generated by CrystalPredictor with rigid search and optimized by DMACRYS with PBE0/6-31G(d,p). A1193 opt by CASTEP is the structure optimized by CASTEP with PBE-TS starting from A1193.

	a	b	c	$\beta$	Standard deviation (%)*
MPYRAZ03	5.48 Å	6.94 Å	10.3 Å	99.9°	
A 1193 opt by CASTEP	+4.7%**	-2.3%	-2.2%	+0.08%	5.73
A1193	+9.04%	-1.87%	+1.69%	-1.87%	9.57
T11	+12.68%	-2.88%	+2.38%	-1.87%	13.35

\*It is calculated by  $\sqrt{\left(\frac{\Delta a}{a}\right)^2 + \left(\frac{\Delta b}{b}\right)^2 + \left(\frac{\Delta c}{c}\right)^2 + \left(\frac{\Delta \beta}{\beta}\right)^2}$

\*\* These data are relevant to the experimental data.

The cell parameters of the monoclinic structure optimized by a variety of methods different from those of the experimental structures for monoclinic polymorph as



shown in Table 11. Although the CASTEP optimization gives the best overall reproduction, the effect of these small errors on the powder patterns needs to be considered.

The pXRD results of generated structures optimized by different methods and experimental structures for monoclinic polymorph and orthorhombic polymorph have been shown in Figure 34 and Figure 35. As introduced in chapter 2, pXRD is the fingerprint for crystal which could be used to identify a new sample of the orthorhombic form. So finding the accurate pXRD pattern for crystal is important.

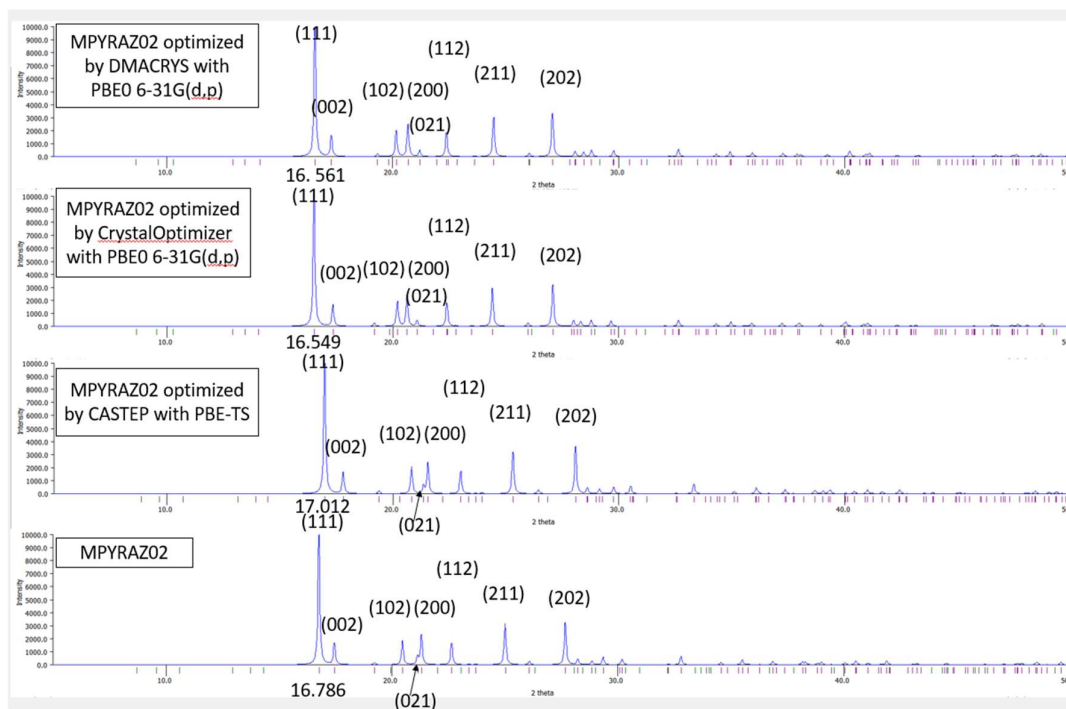


Figure 34 pXRD of MPYRAZ02 (which has atomic coordinates) and MPYRAZ02 optimized by different methods. The major peaks are labelled by the lattice planes.

For the orthorhombic polymorph with known atomic coordinates, the CASTEP optimised structure gives a power pattern (Figure 34), that is closest to the experimental structure. The (021) peak is overlapping with the (200) peak in the simulated experimental structure and the CASTEP structure. There is not too much difference between the rigid TMP structure (optimized by DMACRYS<sup>39</sup>) and the structure where the methyl groups can rotate (CrystalOptimizer<sup>20</sup> results), but the peaks are at a smaller  $2\theta$  than the experimental results of MPYRAZ02.

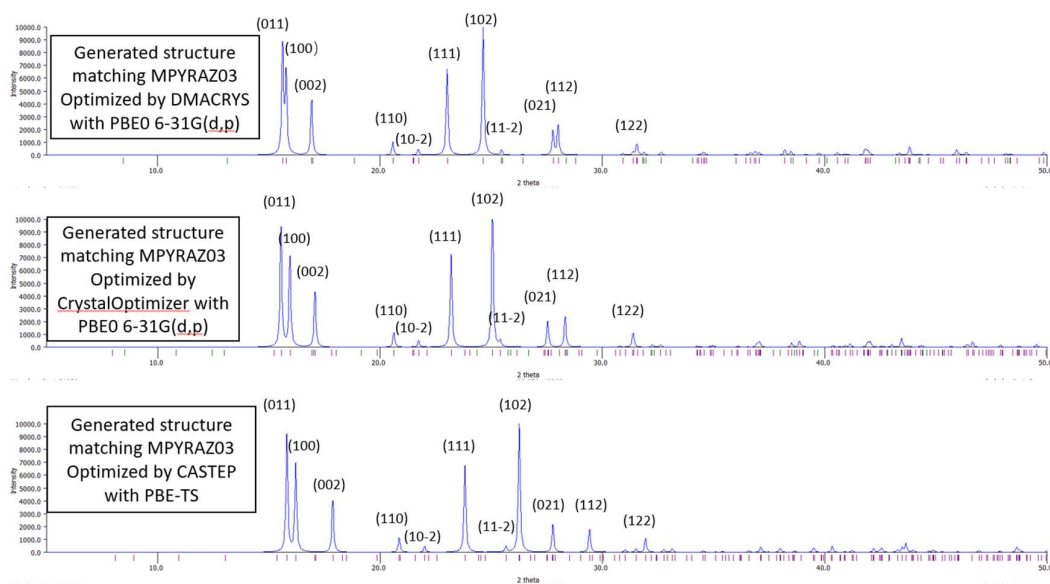


Figure 35 pXRD of generated structure matching the monoclinic MPYRAZ03 optimized by different methods. The major peaks are labelled by the lattice planes

As shown in Figure 35, the generated structure of the monoclinic polymorph optimized with different methods will have different pXRD results. In the rigid molecule DMACRYS optimization, the (011) and (100) peaks cannot be separated; the (021) and (112) peaks cannot be separated; (11-2) peak is so weak and near to the strong peak (102) that it is difficult to measure. As Table 11 suggests that the closest cell match is that given by CASTEP, and this provides the best reproduction of the known orthorhombic pXRD (Figure 34), it seems probable that the CASTEP optimization is providing the best estimate of the crystal structure of the monoclinic form. Thus, the “.res” file is in the Appendix.

### 6.2.1 Conclusion

Since the observed structures have been found near the global minimum in lattice energy and these minima reproduce the experimental structures well in both rigid and flexible searches, the energy model (DMACRYS<sup>39</sup> with PBE0/6-31G(d,p)) is adequate for the interactions between TMP molecules for the co-crystal studies.

A model for atomic coordinates of the monoclinic form has been put in the Appendix. Should an experimental sample of the monoclinic polymorph be obtained, its pXRD should be compared with Figure 35. If there is a match, within the variations of the different computational models, the CSP generated .res file

could be used as a starting point for refining the TMP monoclinic polymorph structure against the experimental pXRD pattern.

### 6.3 Limitations of CrystalPredictor in CSP of CBN

In chapter 4, the experimental CBN structure (CANNOL) was not found in the limited CBN  $Z'=2$  search and generated structures in this search were less dense and less stable (Figure 16) than the CBN experimental structure (CANNOL) optimized with the same model.

Whether the CrystalOptimizer<sup>20</sup> is suitable for the experimental structure CANNOL has been proved in Chapter 4, but as discussed in chapter 5, if the reason why there is no experimental structure in the search used in chapter 4 can be found, confidence in co-crystal searches will be improved.

Thus, whether the CrystalPredictor<sup>18,19</sup> is suitable for the CBN will be assessed in this section. As shown in 2.4.1.1, there are two steps in CrystalPredictor including the generation step using Sobol sequence and the rough optimization and ranking step using the point charge + FIT potential<sup>85,86,87</sup> model. Thus, in this section, those two steps will be assessed separately.

#### 6.3.1 Limitation of Sobol sequence (generation step)

Firstly, whether the Sobol sequence is suitable for the CBN will be assessed in this section. To assess it, as shown in Figure 36, the structures will be generated with rigid experimental conformations and rigidly optimised using point charge+FIT potential model<sup>85,86,87</sup> in CrystalPredictor<sup>18,19</sup>. Those results will be compared to the experimental structure optimized using point charge+FIT potential model to assess if the experimental structure can be generated in Sobol sequence. The optimization in CrystalPredictor using point charge + FIT potential model in this section is different from chapter 4. In chapter 4, there is the flexible optimization with flexible hydrogen carbon pentane tail and hydroxyl group, but in this section, there is the rigid optimization with rigid CBN conformation. Optimization with a flexible hydroxyl group using point charge model is not suitable for the CBN, which will be discussed in the next section.

In Figure 10, the hydroxyl group torsion angle of gas phase CBN conformation molecule is  $0^\circ$ , which is described as the hydroxyl group is “in the ring” in this thesis. The hydroxyl group torsion angle of the conformation of molecule 1 in the experimental structure’s is  $49.58^\circ$ , which is defined as “out of the plane of the ring” in this thesis.

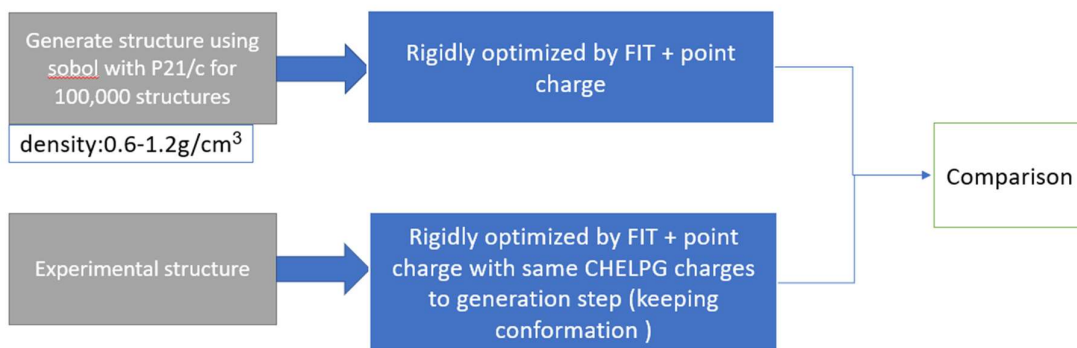


Figure 36 Workflow for rigid CBN search with experimental conformations and experimental structure.

### 6.3.1.1 Testing structure generation with rigid experimental conformations

The small search with only  $P2_1/c$  space group for  $Z'=2$  with rigid CBN experimental conformations is tested in CrystalPredictor2.4.3.2<sup>18,19</sup> with 100,000 structures and 0.6–1.2 g/cm<sup>3</sup> optimized by the FIT potential<sup>85,86,87</sup> + point charge. Then they will be compared to the experimental structure rigidly optimized by the FIT potential + point charge. The optimizing experimental structure with the same model was 11 kJ/mol more stable than the GM structure in this rigid search.

The CBN experimental structure (CANNOL) cannot be found in this rigid search with experimental molecular conformation and experimental space group ( $P2_1/c$ ). Compared to the rigid search, flexible search would require more degrees of freedom. Furthermore, in this rigid search, just the experimental space group has been considered, but in the search 61 common space groups are required at least. Thus, for CBN in  $Z'=2$  would need to be extremely exhaustive before it could generate the experimental structures for enough space group.

### 6.3.2 Limitation of point charge + FIT model

The wrong hydrogen atomic exp-6 potential parameter is used in chapter 4. So, it is necessary to make sure whether FIT potential<sup>85,86,87</sup> + point charge causes the error in the original search use. The repulsion–dispersion interaction is calculated using Equation 3. In Equation 3, the  $A$ ,  $B$ ,  $C$  parameters of hydrogen atoms are dependent on the atom types that are bonding to the hydrogen atom (such as C–H, N–H or O–H), because the hydrogen atom only has one electronic and one positive charge in the nucleus. Coombes polar hydrogen atomic potential parameter<sup>87</sup> ( $A_{HH}=5029.68$  kJ/mol,  $B_{HH}=4.66$  Å<sup>-1</sup>,  $C_{HH}= 21.50$  kJmol<sup>-1</sup>Å) are fitted to the hydrogen atom in N–H; Beyer polar hydrogen atomic potential parameter<sup>86</sup>

( $A_{HH}=2263.3$  kJ/mol,  $B_{HH}=4.66$  Å<sup>-1</sup>,  $C_{HH}= 21.50$  kJmol<sup>-1</sup>Å) are fitted to the hydrogen atom in O–H; and the standard non-polar hydrogen atomic potential parameter ( $A_{HH}=11971$  kJ/mol,  $B_{HH}=3.74$  Å<sup>-1</sup>,  $C_{HH}= 136.4$  kJmol<sup>-1</sup>Å) are fitted to the hydrogen atom in C–H (non-polar H). In chapter 4, the standard non-polar hydrogen atomic potential parameters (C–H) were used. As shown in Figure 40, there is a big difference compared to the polar hydrogen model, but these non exp-6 parameters were only used in CrystalPredictor<sup>18,19</sup>, which will be corrected in the CrystalOptimizer<sup>20</sup>. More details are shown later in Figure 40.

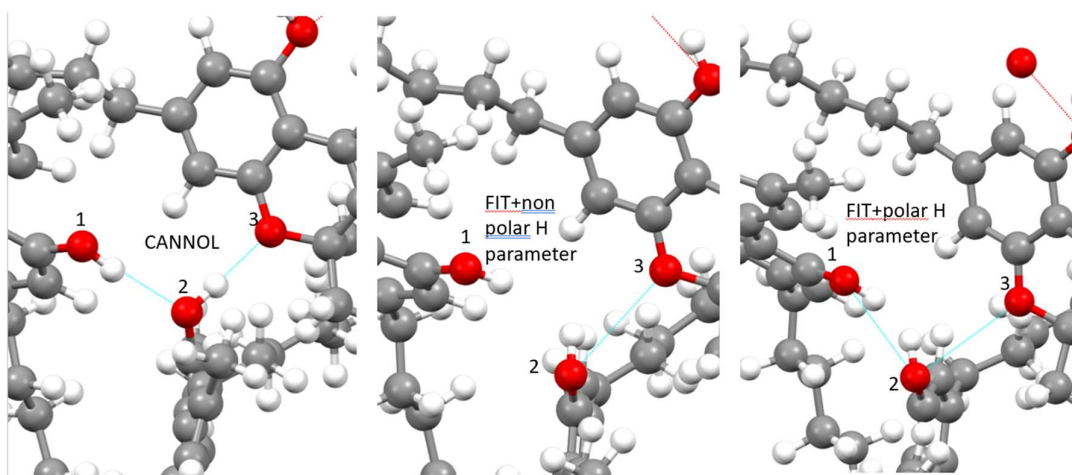


Figure 37 Figures from left to right are CANNOL, CANNOL optimized by FIT + point charge with non-polar hydrogen atomic potential parameters, CANNOL optimized by FIT + point charge with Beyer polar hydrogen atomic potential parameters respectively. The hydrogen bond contact distance is defined as smaller than the vdW radii (defaults in Mercury).

The experimental CANNOL structure is used as the starting point for MINIMISE with flexible hydroxyl group and hydrogen carbon pentane tail (which is the optimization program in CrystalPredictor with point charge + FIT potential<sup>85,86,87</sup> model used in chapter 4) to access whether the FIT + point charge model is suitable for CANNOL experimental structure. MINIMISE optimization results with Beyer polar hydrogen atomic potential parameters and non-polar hydrogen atomic potential parameters have been shown in Figure 37 and compared to the experimental structure. To easily compared optimization results, there are some renumbers for OH...O hydrogen bond.

After the optimization using point charge + FIT<sup>86,87</sup> model, the O2\_H2...O3 is optimized to the middle of the O1\_H1 and O3. Hence, the second reason why the experimental structure CANNOL was not found in chapter 4 is that FIT + point

charge model does not suit for hydroxyl group in CANNOL. This could be because the point charge model does not represent the electrostatic effects of the lone pairs on the oxygen atoms. This will be corrected using the atomic multipoles in CrystalOptimzier or  $\Psi_{\text{cry}}$  method.

Table 12 Testing different optimization models starting from different structures to find the suitable model that can form CANNOL.

No.	Method	Starting structures **			End point			
		Starting structures **	Are there O1_H2...O2_H2...O3****	OH torsion angle (°)	Are there O1_H2...O2_H2...O3****	OH torsion angle (°)	O1_O2, O2_O3 Distance (hydrogen bond length) Å***	RMSD <sub>15</sub> with experimental structure excludes hydrogen atoms
		Starting point			End point			
1	Nonpolar hydrogen* FIT potential + point charge (CrystalPredictor step in chapter 4)	Exp**	YES	49.58	NO	1.83	3.476 3.272	0.516
2	Polar hydrogen* FIT + point charge	Exp	YES	49.58	NO	16.82	3.015 3.260	0.525
3	Polar hydrogen* William potential +point charge	Exp	YES	49.58	NO	-30.2 4	2.889 3.320	0.606
4	Polar hydrogen FIT potential + DMA (CrystalOptimier step in chapter 4)	Exp	YES	49.58	YES	51.62	3.078 2.769	0.17
5	Polar hydrogen FIT potential + DMA CSP in Chapter 4	End2* *	NO	1.83	YES	45.41	3.148 2.737	0.238
6	DFTB3-D3	Exp	YES	49.58	YES	44.85	2.891 2.732	0.314
7	DFTB3-D3	End2	NO	16.82	YES	22.37	3.308 2.774	0.63

\* Nonpolar hydrogen means using the standard non-polar hydrogen (C-H) in point charge optimization, polar hydrogen means using the Coombes polar hydrogen parameters (N-H) in point charge optimization.

\*\* As to starting structures, "Exp" means the experimental structure with added hydrogen atoms in section 4.1 collected from CSD using Mercury; "End2" means the end structure in No.2 of Table 12, which is the experimental structure with added hydrogen atoms optimized by the Minimise using polar hydrogen FIT potential + point charge.

\*\*\* In this table, atoms are numbered as in Figure 37. And the experimental data are 2.921 Å and 2.734 Å

\*\*\*\* Hydrogen bond is defined by the Mercury default parameters.

In Table 12, CBN experimental structures (CANNOL exp) or CANNOL optimized with point charge model (end2) are optimized using different models to compare whether different models are suitable for CANNOL and to investigate whether the models can retrieve the correct structure starting from a minimum obtained with the point charge model, as would be generated by CrystalPredictor.

After the FIT + point charge optimization, the O1\_H2...O2\_H2...O3 hydrogen bond in experimental structure (CANNOL) cannot be formed, but using CrystalOptimizer or DFTB3-D3 optimization<sup>67</sup>, the O1\_H2...O2\_H2...O3 hydrogen bond will be formed. Results of CrystalOptimizer and DFTB3-D3 show that the O1\_H2...O2\_H2...O3 hydrogen bond structure is the lowest energy structure. It means that in chapter 4 the hydrogen atoms are added to the experimental structure correctly. But the results of FIT + point charge shows that point charge model is not suitable for this search regardless of exp-6 model used for the hydroxyl group.

As to optimization No. 5 and 7 (the last DFTB3-D3 result) in Table 12, the starting structure is the minimise structure optimized by point charge + FIT. After CrystalOptimizer optimization, the O1\_H2...O2\_H2...O3 is corrected and the RMSD<sub>15</sub> compared to the experimental structure is small. But after the DFTB3-D3, O1\_H2...O2\_H2...O3 is corrected to some extent but the hydroxyl group (OH) torsion angle is not good enough, and RMSD<sub>15</sub> to the experimental structure is also poor. Furthermore, compared with the distance between oxygen atoms, CrystalOptimizer results agree with the experimental structure well whatever the starting point is; the DFTB3-D3 result agree with the experimental structure only when the starting point is the experimental structure, as to start with the wrong structure optimized by point charge model, the mol2 molecule will be optimized far away from the mol1.

### 6.3.3 Conclusion

1, In section 6.3.1, a rigid search with 100,000 structures using experimental conformations for space group  $P2_1/c$  cannot find the CBN experimental structure (CANNOL). As to completed flexible search, more degrees of freedom and space groups should be considered than the rigid search, thus, Sobol sequence requires millions of generated structures for flexible CBN search to generate the



experimental structure. In other words, it is difficult to use Sobol sequence to generate CANNOL experimental structure.

2, The hydroxyl group hydrogen bond system with lone pairs is not well modelled with the used in CrystalPredictor<sup>18,19</sup> point charge model. When CANNOL is optimized with point charge model, the hydroxyl group that was out of the plane of the attached ring will switch to being in the plane of the ring forming an unrealistic structure with the two hydroxyl groups pointing toward each other. Checking the log file, the hydrogen bonds fluctuate a bit during the optimization, and then the hydrogen bonding region expands so there is space for the hydroxyl group to swing around. Then they close back up to make the "incorrect" hydrogen bond. This incorrect O-H...H-O hydrogen bond was seen a few times in the CBN limited searches, so it is favourable on the point charge + FIT potential<sup>85,86,87</sup> energy surface, which shows point charge model is not suitable for the lone pair hydrogen bond.

There is very little movement from cell parameters or atoms apart from the hydroxyl group dihedral. The problem with the point charge model in CrystalPredictor was found in rigid coronene. The point charge model cannot give a lattice energy minimum for one of the coronene experimental polymorphs<sup>1</sup>. Coronene experimental polymorphs will be optimized to the same structure when those two coronene experimental structures are optimized by point charge+FIT model. Two structures keep distinct after optimization with DMACRYS<sup>39</sup> using a distributed multipole electrostatic model, which shows point charge model is poor. The anisotropic nature of the delocalized  $\pi$  system is needed to define separate minima for the two polymorphs in a way that point charges cannot. Coronene have delocalized  $\pi$  bonds, and experimental CBN has the lone pair in the hydroxyl group, their delocalized  $\pi$  bond or lone pair cannot be presented by point charge model accurately. Coronene experimental structure was found in Molpak<sup>99</sup> search and ranked as low energy because the multipoles can express their electronic structure accurately<sup>2</sup>. MOLPAK cannot be used for flexible

---

<sup>1</sup> Dr. Rui Guo did the coronene CSP search using CrystalPredictor, but only one experimental structure has been found.

<sup>2</sup> Dr. Louise Price did this Molpak search.

molecules like CBN, however it appears that point charge model is not suitable for delocalized  $\pi$  bonds and multipoles model is suitable for delocalized  $\pi$  bonds.

So, the point charge model cannot present the  $\pi$  stacking and lone pairs where CrystalPredictor2.4.3.2 is failing. CrystalOptimizer<sup>20</sup> with atomic multipoles can present the lone pair and  $\pi$ -bond with anisotropy. DFTB3-D3<sup>67</sup> can present the lone pair to some extent.

The group of Pantelides and Adjiman at Imperial College is trying to adapt CrystalPredictor to use the multipole model instead the point charge model to deal with this problem, however the code is not yet available.

#### 6.4 Testing alternative methods

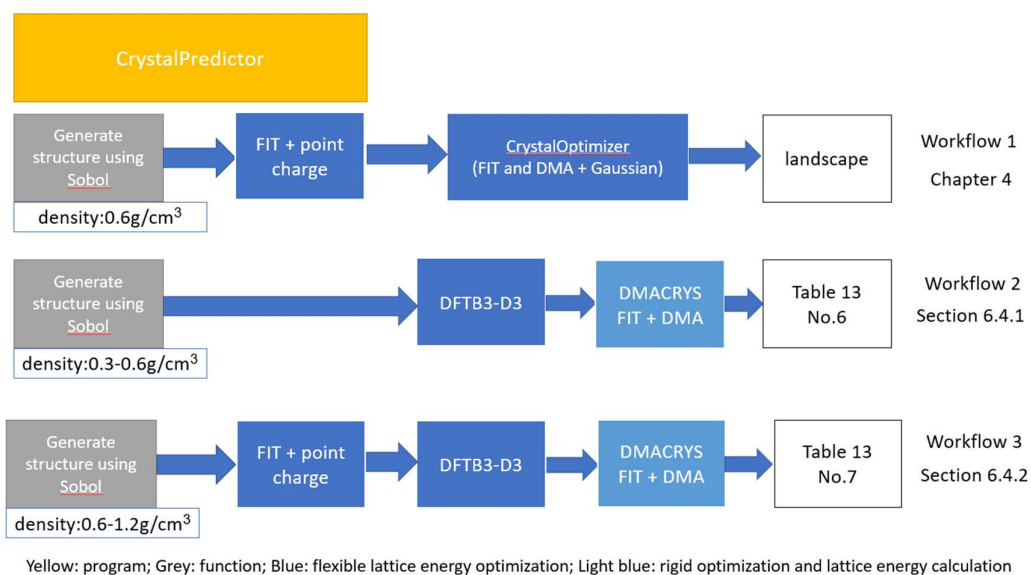


Figure 38 3 workflows for a CBN CSP search. Workflow 1 is the original search, and it is used in the co-crystal search. Workflow 2 and Workflow 3 are investigated new CSP workflows using DFTB3-D3 based on workflow1 and aim to correct the problem caused by the point charge model.

Two new workflows have been investigated (Figure 38) to test with limited searches. The Beyer polar hydrogen atomic FIT potential<sup>85,86,87</sup> was used for all repulsive and dispersion models. Workflow 1 in Figure 38 was used for the first CBN flexible search (chapter 4) and co-crystal searches (chapter 5).

##### 6.4.1 New search using Workflow 2

As shown in Table 12, after optimizations by DFTB3-D3<sup>67</sup>, CBN experimental structure can keep the hydroxyl group out of the plane, so this DFTB3-D3  $\Psi_{\text{cr}}$

relaxed all atoms optimization model is suitable for CBN lone pair with the limited computational resource.

The DFTB3–D3 method is used instead of the FIT potential<sup>85,86,87</sup> + point charge optimization model in the first rough lattice energy calculation and ranking stage. The lattice energies were calculated using the DMACRYS<sup>39</sup> shown in Figure 38 workflow 2.

1,000 structures are generated in the small  $Z'=2$  search with only  $P2_1/c$  space group (CBN experimental structure space group) from 0.3 to 0.6 g/cm<sup>3</sup> density. Then generated structures are optimized using DFTB3–D3. Finally, DMACRYS is used to calculate lattice energies of those optimized structures.

The experimental structure in this small search still was not found. The GM lattice energy after DFTB3–D3 and DMACRYS optimization is –135 kJ/mol which is different from the CANNOL lattice energy after DFTB3–D3 and DMACRYS with –150 kJ/mol. There is still a large gap between the GM and experimental structure, but it is a small search for this very flexible molecule. GM forms the hydrogen bond between the hydroxyl group and the oxygen atom in the ring by the same symmetry equivalent molecules, which is similar to experimental structure, but other symmetry equivalent molecule is still formed with different ways from the experimental structure.

In theory, if the search is large enough and completed, this workflow could find the experimental structure because the DFTB3–D3 is suitable for modelling the lone pair in the hydroxyl group of CBN. Although the DFTB3–D3 model is cheaper than CrystalOptimizer<sup>20</sup> in PBE0/6–31G(d,p) and FIT potential, it is still more expensive than point charge model too much. It is difficult to optimize more than 100,000 structures using DFTB3–D3 model in small computer and as shown in section 6.3.1, 100,000 structures search is far away from the completed Sobol sequence search.

#### 6.4.2 New search using workflow 3

To assess the larger search than the search in section 6.4.1, workflow 3 is investigated. As shown in Table 12, after the poor point charge model optimization and DFTB–D3<sup>67</sup> optimization, the hydroxyl group in CBN molecule of CANNOL can be corrected to some extent. So Sobol sequence generated CBN

crystal structures are roughly optimized and ranked using point charges+FIT potential<sup>85,86,87</sup> model firstly, then DFTB3-D3 is used to optimize the lowest 300 structures and their lattice energies are calculated by DMACRYS<sup>39</sup> shown in Figure 38 workflow 3.

There is a balance between the inter- and intra- molecular energy in the lattice energy calculation. When the conformation is not the gas phase, but the hydrogen bond can be formed, there is more intermolecular energy reduction than the intramolecular increase and the total lattice energy will reduce. there is a balance between the intermolecular interaction and the intramolecular interaction.

After the generation step if the density of a structure is low and molecules in this structure are far away with each other, during the optimization, the intramolecular energy usually will be optimized to the gas phase conformation firstly. The intermolecular energy is negligible as the molecules are too far apart, so those molecules are difficult to be changed from the gas phase to get the lower intermolecular interaction in the Minimize step. So, increasing the density in the generation step can reduce the distance among molecules and make the intermolecular and intramolecular interactions are optimized together. So, the density limitation is increased in workflow 3 of Figure 38 from 0.3–0.6 g/cm<sup>3</sup> to 0.6–1.2 g/cm<sup>3</sup> to improve the accuracy.

There is a significant reranking after the DFTB3-D3 optimization. GM structure in workflow 3 of Figure 38 is formed with the OH...O ring hydrogen bond, but this hydrogen bond is formed between the symmetry equivalent molecules which is different from the experimental structure CANNOL. The lattice energy of the GM is 8 kJ/mol higher than the experimental structure using the same optimization model.

In Table 12, after FIT and point charge optimization and DFTB3-D3<sup>67</sup> optimization, the hydroxyl group is optimized to the correct position to some extent, but there is a still large error compared to the experimental structure. So, there are some better structures in workflow 3 than workflow 1 (which will be chosen and briefly introduced in section 6.5), but it is still an unsuccessful improvement on the search, this DFTB3-D3 corrects the hydrogen bonding conformation but it is not sufficient.

### 6.4.3 Conclusion

As shown in Table 12, FIT potential<sup>85,86,87</sup> + point charge model forces the hydroxyl group into the plane of the ring where it cannot form the hydrogen bond in CANNOL. DFTB3-D3<sup>67</sup> can keep the hydroxyl group out of the ring and form the hydrogen bond, but it cannot correct the hydrogen bonding conformation sufficiently, and it is expensive.

In this section, two new workflows have been investigated aiming to deal with the limitations in sections 6.3.1 and 6.3.2. Although the experimental structure was still not be found, it can be believed that the experimental CBN structure should be found in a workflow 2 search that was extensive enough. However, DFTB3-D3 is more expensive than point charge model, so new workflow 2 is too expensive. Workflow 3 is investigated to reduce the cost of workflow 2, but it is unsuccessful because DFTB3-D3 cannot correct the hydrogen bonding conformation sufficiently.

## 6.5 Sensitivity to potential model

This section is used to test the relative energy for CBN crystal intermolecular energy surface to make sure that the polar hydrogen atomic FIT potential<sup>85,86,87</sup> model used in chapter 5 co-crystal CSP in final energy evaluation, ie using atomic multiple, is reasonable.

### 6.5.1 Methodology

To test the effect of different potential, basis sets and charge density, CSP generated structures are selected which are low in energy after the DFTB3-D3<sup>67</sup> and DMACRYS<sup>39</sup> optimization, but differ in hydrogen bonds.

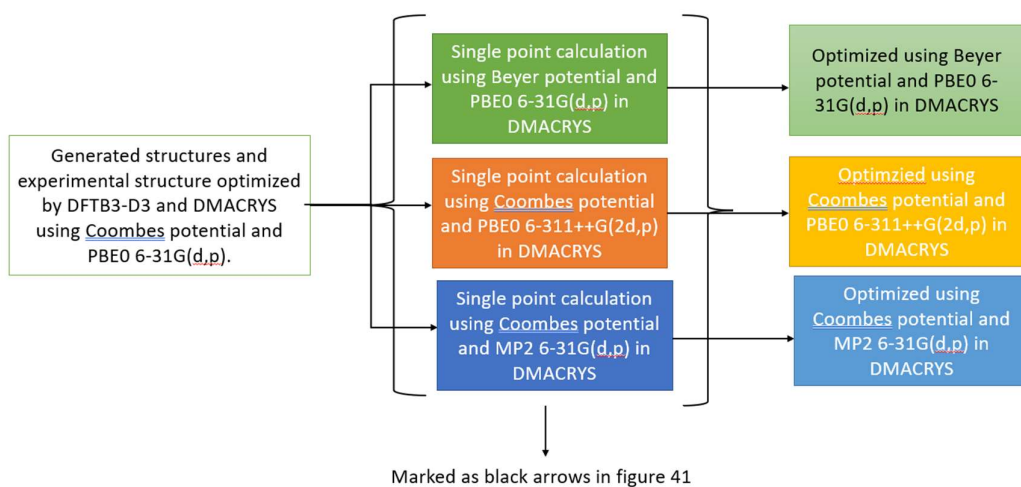


Figure 39 Workflow of testing sensitivity to potential model in DMACRYS with different basis sets, functional, and potential parameters. The colours of the rectangles match the colours of the arrows in Figure 41.

GM structure in the workflow 3 of Figure 38 is formed with the OH...O ring hydrogen bond, but this hydrogen bond is formed between the symmetry equivalent molecules which is different from the experimental CANNOL structure. 8<sup>th</sup> and 9<sup>th</sup> lowest structures form the chain with the same symmetry equivalence molecule which is similar to the experimental structure. However, another symmetry equivalence molecule does not form a hydrogen bond with others, which disagree with the experimental structure. The 14<sup>th</sup> lowest structure forms the chain by the same symmetry equivalence molecule.

The single point energy calculation and DMACRYS optimization will be calculated by different models including changing the hydrogen atomic exp-6 FIT potential<sup>85,86,87</sup> and using a different charge density for the atomic multipoles, ie MP2, PBE rather than PBE0 and 6-311++G(2d,p) rather than 6-31G(d,p) basis set.

## 6.5.2 Results

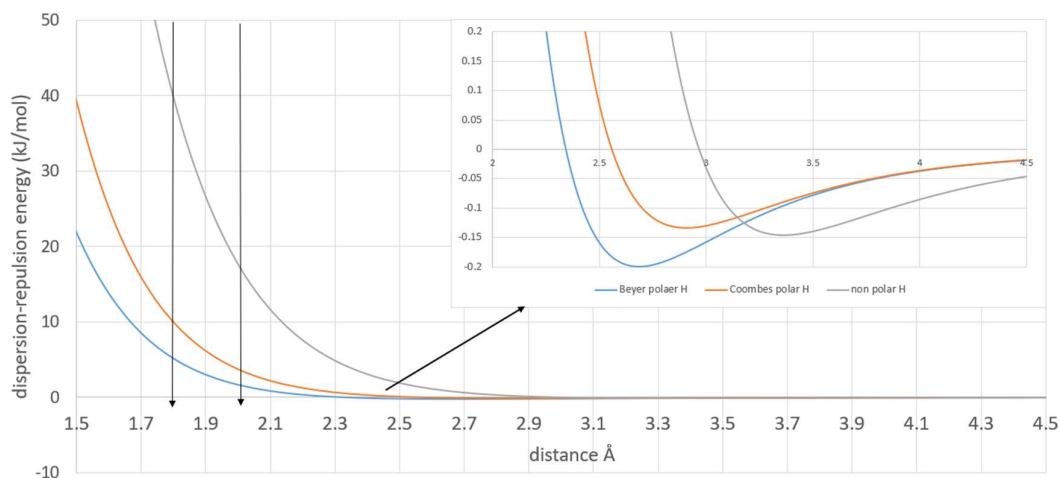


Figure 40 Dispersion–repulsion potential for hydroxyl group along non–polar hydrogen (grey) ( $A_{HH}=11971$  kJ/mol,  $B_{HH}=3.74$  Å<sup>-1</sup>,  $C_{HH}= 136.4$  kJmol<sup>-1</sup>Å), compared to that along Coombes polar H (orange) ( $A_{HH}=5029.68$  kJ/mol,  $B_{HH}=4.66$  Å<sup>-1</sup>,  $C_{HH}= 21.50$  kJmol<sup>-1</sup>Å) and Beyer polar hydrogen model (blue) ( $A_{HH}=2263.3$  kJ/mol,  $B_{HH}=4.66$  Å<sup>-1</sup>,  $C_{HH}= 21.50$  kJmol<sup>-1</sup>Å). Black arrows highlight the O...H bond distance in the hydrogen bonds in the experimental structure CANNOL. This shows the experimental hydrogen bond repulsion and dispersion energy is sensitive to the different hydrogen parameters.

Coombes potential is the most popular potential. This polar hydrogen atomic potential model is usually used in the repulsion–dispersion interaction calculation using Equation 3. This Coombes potential model is based on the hydrogen bond formed by N–H, but the O–H...O hydrogen bond is different from N–H...O hydrogen bond. Thus, there is another Beyer potential that is used to present the hydrogen bond formed by the donor O–H. As shown in Figure 40, when the distance between the acceptor and hydrogen atom is longer than 3.5 Å, there is not too much difference between Coombes and Beyer polar hydrogen atomic potential, but if the distance between the acceptor and hydrogen atom is smaller than 3.0 Å, the difference between those two polar hydrogen atomic models cannot be ignored, for example, in the experimental structure distance about 1.8 Å, those two models will lead to about 5 kJ/mol difference. As to the non–polar hydrogen model, there is a big difference compared to the polar hydrogen model, but these non–polar exp–6 parameters were only used in CrystalPredictor<sup>18,19</sup>, which will be corrected in the CrystalOptimizer<sup>20</sup>.

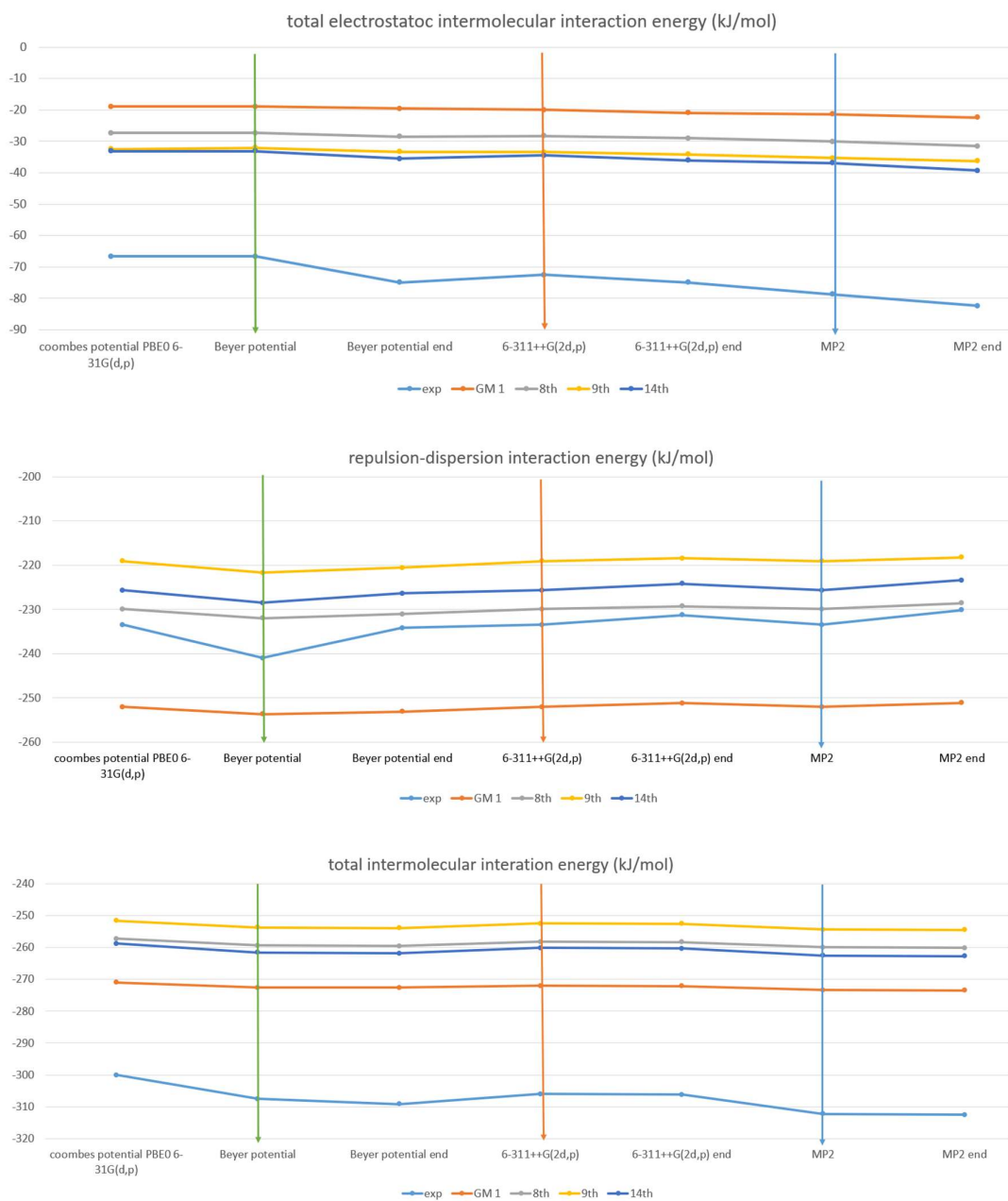


Figure 41 Energy comparison among 5 structures using different methods. Green arrow means the single point calculation using Beyer potential shown in Figure 39. Red arrow means the single point calculation using 6-31G(d,p) shown in Figure 39. Blue arrow means the single point calculation using MP2 shown in Figure 39.

The experimental structure is more sensitive to different polar hydrogen potentials, and charge densities for multipoles than the other CSP generated structures. Compared to the model used in CSP chapter, different polar hydrogen potentials, DFT method and basis set will lead to a large energy difference between the crystal structures after optimization.



The other CSP generated structures are not sensitive to the different polar hydrogen potentials, because their hydrogen bond is longer than the experimental structure. After the Coombes potential is changed to Beyer potential and the structure optimization, the final repulsion–dispersion energies are close to each other with small crystal structure changes (RMSD<sub>15</sub> is less than 0.1). In the model that is not accurate enough, the experimental structure has the lowest lattice energy as well. Other generated structures are not sensitive to the different models, whose packing coefficients are lower than the experimental structure and distances between hydrogen are longer than the experimental structure. Hence the shorter hydrogen bond in the experimental structure means that the relative lattice energy is sensitive to the charge densities used for the electrostatics model and the polar hydrogen parameters.

## 6.6 Conclusion

1. CSP approach used for co–crystal is adequate for TMP rigid assumption and has been used to propose a structure for the monoclinic polymorph shown in section 6.1.3.1. The results are only slightly affected by keeping TMP rigid. However, it is not adequate for CBN.
2. Sobol sequence is difficult to be used to generate the CBN experimental structure CANNOL, because the rigid 100,000 structures search with the experimental space group did not find the experimental structure CANNOL. The flexible CBN search with 61 space groups will be expensive because of the big search.
3. Point charge model cannot present the lone pairs in hydroxyl groups, so point charge model optimization method will lead to large errors for CBN experimental structure (CANNOL).
4. There is a balance between the intra– and inter– molecular energy in CBN experimental structure (CANNOL). When the hydroxyl group is not in the gas phase conformation and the intramolecular energy will increase, but two hydrogen bonds can be formed to reduce the intermolecular energy in CANNOL. Furthermore, the CBN hydrogen carbon pentane tail is so flexible that it affects the packing. When two symmetry equivalences are near to each other, hydrogen carbon pentane tail will cause steric hindrance in the search. In Figure 9, CSP generated structures are likely

to have molecules that are parallel to each other, but in the experimental structure, the molecules' tails are perpendicular to each other.

5. Using DFTB3-D3<sup>67</sup> for the initial lattice energy optimisation in CrystalPredictor<sup>18,19</sup> instead of a point charge model is in principle an improvement that should lead to finding the experimental structure CANNOL. However, it is too expensive to test within a sufficiently large CSP search. Adapting CrystalPredictor to take multipoles rather than just the atomic charge in the initial lattice optimisation may be more cost-effective than using DFTB3-D3 method to represent lone pair and  $\pi$  bond that stabilises the out of plane hydroxyl group.
6. The density in the generation step was increased to ensure that the intermolecular and intramolecular interaction are optimized together. Although the experimental structure of CBN still was not found, this emphasises that all the parameters in CrystalPredictor need to be considered for large systems.

## 6.7 Appendix: monoclinic TMP suggested .res file

TITL monoclinic TMP suggested .res file.									
CELL 1.54180 11.09186 6.782278 10.06723 90 99.98487 90									
LATT -1									
SFAC H C N									
H	1	0.4800833151621	0.1884921912568	0.4550810054986	H	1	0.0199466869342	0.8113889743922	0.5449460863210
H	1	0.5199166848379	0.8115078087432	0.5449189945014	H	1	0.9800533130658	0.1886110256078	0.4550539136790
H	1	0.4800833151621	0.3115078087432	0.9550810054986	H	1	0.0199466869342	0.6886110256078	0.0449460863210
H	1	0.5199166848379	0.6884921912568	0.0449189945014	H	1	0.9800533130658	0.3113889743922	0.9550539136790
H	1	0.3479380696649	0.0362280990852	0.4488500743783	C	2	0.3204127719251	0.5370378162941	0.4056865832517
H	1	0.6520619303351	0.9637719009148	0.5511499256217	C	2	0.6795872280749	0.4629621837059	0.5943134167483
H	1	0.3479380696649	0.4637719009148	0.9488500743783	C	2	0.3204127719251	0.9629621837059	0.9056865832517
H	1	0.6520619303351	0.5362280990852	0.0511499256217	C	2	0.6795872280749	0.0370378162941	0.0943134167483
H	1	0.3726484801162	0.1720292253847	0.3045558172615	C	2	0.3121789008966	0.3460243885466	0.4588046703312
H	1	0.6273515198838	0.8279707746153	0.6954441827385	C	2	0.6878210991034	0.6539756114534	0.5411953296688
H	1	0.3726484801162	0.3279707746153	0.8045558172615	C	2	0.3121789008966	0.1539756114534	0.9588046703312
H	1	0.6273515198838	0.6720292253847	0.1954441827385	C	2	0.6878210991034	0.8460243885466	0.0411953296688
H	1	0.3668230105287	0.4935249456820	0.2087002879952	C	2	0.1795818943524	0.4627745424175	0.5943561343612
H	1	0.6331769894713	0.5064750543180	0.7912997120048	C	2	0.8204181056476	0.5372254575825	0.4056438656388
H	1	0.3668230105287	1.0064750543180	0.7087002879952	C	2	0.1795818943524	0.0372254575825	0.0943561343612
H	1	0.6331769894713	-0.0064750543180	0.2912997120048	C	2	0.8204181056476	0.9627745424175	0.9056438656388
H	1	0.3937110664057	0.7363596136773	0.2757732554603	C	2	0.1878319703725	0.6537983329615	0.5412534177224
H	1	0.6062889335943	0.2636403863227	0.7242267445397	C	2	0.8121680296275	0.3462016670385	0.4587465822776
H	1	0.3937110664057	0.7636403863227	0.7757732554603	C	2	0.1878319703725	0.8462016670385	0.0412534177224
H	1	0.6062889335943	0.2363596136773	0.2242267445397	C	2	0.8121680296275	0.1537983329615	0.9587465822776
H	1	0.4940606853525	0.5386201937647	0.3367425627866	C	2	0.3816626772575	0.1759760427082	0.4147827830931
H	1	0.5059393146475	0.4613798062353	0.6632574372133	C	2	0.6183373227425	0.8240239572918	0.5852172169069
H	1	0.4940606853525	0.9613798062353	0.8367425627866	C	2	0.3816626772575	0.3240239572918	0.9147827830931
H	1	0.5059393146475	0.0386201937647	0.1632574372134	C	2	0.6183373227425	0.6759760427082	0.0852172169069
H	1	0.0059451145798	0.4614313964807	0.6633227101787	C	2	0.3979636889847	0.5792230665630	0.3010390606076
H	1	0.9940548854202	0.5385686035193	0.3366772898213	C	2	0.6020363110153	0.4207769334370	0.6989609393924
H	1	0.0059451145798	0.0385686035193	0.1633227101787	C	2	0.3979636889847	0.9207769334370	0.8010390606076
H	1	0.9940548854202	0.9614313964807	0.8366772898213	C	2	0.6020363110153	0.0792230665630	0.1989609393924
H	1	0.1061065531250	0.2633615346973	0.7240801063508	C	2	0.1020015060922	0.4205545208000	0.6989713092037
H	1	0.8938934468750	0.7366384653027	0.2759198936492	C	2	0.8979984939078	0.5794454792000	0.3010286907963
H	1	0.1061065531250	0.2366384653027	0.2240801063508	C	2	0.1020015060922	0.0794454792000	0.1989713092037
H	1	0.8938934468750	0.7633615346973	0.7759198936492	C	2	0.8979984939078	0.9205545208000	0.8010286907963
H	1	0.1332322806378	0.5060070113593	0.7913998260735	C	2	0.1183624022712	0.8238588645172	0.5852841083532
H	1	0.8667677193622	0.4939929886407	0.2086001739265	C	2	0.8816375977288	0.1761411354828	0.4147158916468
H	1	0.1332322806378	-0.0060070113593	0.2913998260735	C	2	0.1183624022712	0.6761411354828	0.0852841083532
H	1	0.8667677193622	1.0060070113593	0.7086001739265	C	2	0.8816375977288	0.3238588645172	0.9147158916468
H	1	0.1273511842519	0.8277609459304	0.6955089968864	N	3	0.2582317513727	0.6877467871141	0.4478112987475
H	1	0.8726488157481	0.1722390540696	0.3044910031136	N	3	0.7417682486273	0.3122532128859	0.5521887012525
H	1	0.1273511842519	0.6722390540696	0.1955089968864	N	3	0.2582317513727	0.8122532128859	0.9478112987475
H	1	0.8726488157481	0.3277609459304	0.8044910031136	N	3	0.7417682486273	0.1877467871141	0.0521887012525
H	1	0.1521192366524	0.9636112562190	0.5512599606558	N	3	0.2417733482344	0.3120715532969	0.5522401821741
H	1	0.8478807633476	0.0363887437810	0.4487400393442	N	3	0.7582266517656	0.6879284467031	0.4477598178259
H	1	0.1521192366524	0.5363887437810	0.0512599606558	N	3	0.2417733482344	0.1879284467031	0.0522401821741
H	1	0.8478807633476	0.4636112562190	0.9487400393442	N	3	0.7582266517656	0.8120715532969	0.9477598178259
END									

## 7 Conclusions, before and after receiving experimental structures for co-crystals

In this thesis, CSP studies on TMP and CBN and their co-crystals are reported in chapter 3,4,5, which was the entry for the 7th Blind Test target XXX.

Predicted co-crystal structures of target XXX in this thesis were submitted to CCDC, the required lists of generated 1500 structures and co-crystallization energy ranked the most stable 100 structures shown in Table 8 on 29th June 2021. Further work on CSP CBN revealed that the method used in Chapters 4, and 5 was inadequate for CBN, because the experimental conformation with hydroxyl group (OH) out of the plane was too energetically unfavourable.

Thus, this thesis suggested an alternative strategy for CBN CSP, but it is too expensive to be carried out for CBN  $Z'=2$ . This thesis also proposed a structure for TMP monoclinic form. In the CSD, there are no atomic coordinates for this polymorph.

On 15<sup>th</sup> July 2022, a summary of blind test results was provided from CCDC. For target XXX, there are 2 stoichiometries of the experimentally observed forms, which are CBN:TMP=1:1 and 2:1. 1CBN:1TMP and 1CBN:2TMP stoichiometries are predicted in chapter 5, which does not agree with the experimental results. In the generated and optimized 572 2CBN:1TMP co-crystal structures, co-crystallization energy  $\Delta E_{cc}$  of GM is 9.451 kJ/mol, which means that no generated 2CBN:1TMP co-crystal structures are more stable than the components, whereas this stoichiometry co-crystal had been found.

The CCDC compared every group that submitted 1500 generated structures to the experimentally observed forms with the Crystal Packing Similarity tool, using a cluster of 20 molecules with 25% distance tolerance and 25 degrees for angle tolerance. They reported that the list of 1500 structures had not included either observed crystal structure. No group found the experimental form B (2CBN:1TMP), but as to the form A (1CBN:1TMP), two groups has generated experimental structures with the best match being  $RMSD_{20}=0.08 \text{ \AA}$ .

At the end of August, the experimental structures were provided from CCDC, whose CBN conformations have a folded hydrocarbon pentane tail and in-plane hydroxyl group (OH). More detail is in section 7.2.

### 7.1 Conclusions – towards a CSP method for CBN structures

In CBN searches of 6.3, it is noticed that the point charge model is not suitable for CBN molecules:

- Minimising the experimental structure starting with a point charge model gives a distorted structure shown in section 6.3.2 Table 12 and Figure 37. The local energy minimum with a point charge model is so poor that it is unlikely to be found in the method used in chapters 4, 5.
- Relative energies for experimental structures optimized by point charge model are very poor in comparison with the better methods (such as multipoles model, DFTB3–D3). In the small flexible CBN search after the point charge model optimization, CANNOL is ranked as the 6<sup>th</sup> lowest structure, but the CrystalOptimizer<sup>20</sup> and DFTB–D3<sup>67</sup> models calculate the CANNOL lattice energy to be much lower (~ 20 kJ/mol) than GM in the CBN search (chapter 4).
- The Z'=2 CANNOL structure was not found in the rigid CBN search of with experimental conformation and  $P2_1/c$  space group, this search used 100,000 structures, so a full search considering the flexibility and in all space groups will be very expensive for Sobol sequence approach.

The cause of this problem appears to be the hydroxyl group needing to be out of the plane in the experimental CANNOL structure. The intramolecular energy penalty for distorting the hydroxyl group out of the plane of the ring was not compensated for by the intermolecular energy from forming the hydrogen bonds with the atomic charge model that was the basis of the initial structure generation method. Modelling CBN needs a more realistic description of the hydrogen bonding, either using distributed multipoles to represent the lone pair electrons, or a  $\Psi_{\text{cry}}$  method.

The CrystalPredictor<sup>18,19</sup> methodology used in the Blind Test submission for XXX used the atomic point charge model as the first crude lattice energy minimisation. The lattice energy minimisation using DFTB+<sup>77</sup> for structures generated by the

Sobol sequence should prove a more effective way of producing realistic structures (6.4.3).

## 7.2 The experimental information on XXX, the co-crystals of CBN and TMP.

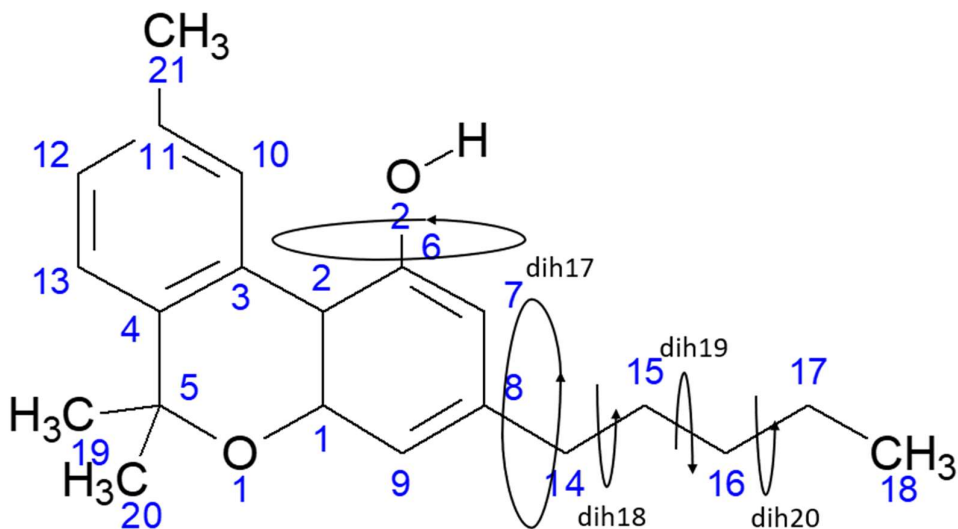


Figure 42 The CBN molecule with black curved arrows qualitatively indicates the range of the flexible group torsions used in the CSP searches (Table 3).

At the end of August 2022, the CCDC provided the two experimental crystal structures of the co-crystals of XXX, on a strictly confidential basis, because there is a patent application associated with this system which the providers are waiting to be granted, expected Q3–Q4 2022.

The structure of the 1CBN:1TMP co-crystal has a hydrogen bond between the TMP and CBN. The 2CBN:1TMP structure has the CBN...TMP...CBN structure. It does have the hydrogen bonding that has been discussed in chapter 5 might be stabilising the 2CBN:1TMP co-crystal, but this structure had not been generated in the search used in this thesis.

In the co-crystal experimental structures, the hydrocarbon tails of CBN are folded, and in conformations that were not sampled in the search. In the 1CBN:1TMP experimental structure, dih18 is 68.95°, whereas the analysis of structures in the CSD done by Dr. Louise Price, showed that 88.5% of structures had this angle close to 180° (Figure 14 in chapter 4). In the CCT experimental structure, one molecule has dih20=66.90° and the other is disordered with

dih<sub>20</sub>=-68.78° in the major component and -173.04° in the minor. This suggests that a long straight hydrocarbon pentane tail will lead to a packing problem in structures with hydrogen bonding because the TMP is so small.

The experimental structures are well reproduced after they have been lattice energy minimised (Table 13) showing that computational method used in this thesis is suitable.

Table 13 RMSD<sub>15</sub> comparing experimental co-crystal structures and those optimised by CrystalOptimier with PBE0/6-31G(d,p)

	CT	CCT <sub>maj</sub>	CCT <sub>min</sub>
RMSD <sub>15</sub> /Å	0.248	0.265	0.308

As shown in Table 14, the folded tails, although not in the gas phase minima, do not lead to a high intramolecular energy penalty, (no more than 5 kJ/mol) consistent with the CBN intramolecular energy scans (Figure 12 in chapter 4).

Table 14 The lattice energy optimized experimental structures of XXX compared with the global minimum in each search (Table 5 Table 6 Table 7 Table 8).

CCT	CBN1 ΔE <sub>intra</sub> kJ/mol	CBN2 ΔE <sub>intra</sub> kJ/mol	Intermolecular energy kJ/mol	Total lattice energy kJ/mol	Density g/cm <sup>3</sup>	CBN1 dih C18_C17 _C16_C15	CBN2 dih C18_C17 _C16_C15	Co-crystal lization energy kJ/mol
Exp <sub>Major</sub> *	5.32	6.86	-375.04	-362.87	1.0977	-67.67	62.17	-3.959
Exp <sub>Minor</sub> *	6.14	4.14	-371.25	-360.97	1.1005	-169.82	77.40	-3.009
GM	5.91	9.14	-351.09	-336.05	1.0981	177.35	178.02	9.451
CT						CBN dih C16_C15 _C14_C8	CBN dih C15_C14 _C8_C7	
Exp*	3.96		-226.38	-222.42	1.0983	68.95	89.14	-7.972
GM	6.88		-225.25	-218.37	1.0961	179.89	-9.45	-3.918
#2	7.20		-225.31	-218.12	1.0837	-173.49	-22.35	-3.666
#3	9.52		-226.21	-216.69	1.1025	-173.58	-12.68	-2.24
CTT								
GM	7.34		-303.94	-296.60	1.1290	179.44	-8.93	-8.204
#2	7.29		-301.39	-294.10	1.1139	179.94	-10.19	-5.701

\*" Exp" means the experimental structure. "Major" means the major structure and "Minor" means the minor structure.

The intramolecular energies for CBN in the experimental co-crystal structures are more than compensated for by the intermolecular lattice energies. Although for 2CBN:1TMP the conformations are similar in energy to those in the GM structure, the intermolecular lattice energy with two hydrogen bonds is so much more stabilising that the co-crystallization energy is about 12 kJ/mol more stabilising. This is consistent with the 2CBN:1TMP search did not include the most obvious motif with a CBN molecule hydrogen bonded to the two nitrogen atoms of TMP. The structures that were generated did not have many hydrogen bonds (shown in Table 7), so it is not surprising that they were not low enough in energy to be plausible co-crystals. This explains why a 2CBN:1TMP co-crystal could not be formed in the process in chapter 5.

As the 1CBN:1TMP co-crystal structure, the three lowest energy structures in the search have a similar intermolecular energy to the experimental structure, but all have a higher intramolecular energy of CBN than the experimental structure. This is caused by the torsion angle between the pentane tail and the aromatic ring being a high energy conformation (Figure 11 and Figure 42) in the CSP generated structures.

Comparing the co-crystallization energies between the stoichiometries, the GM and next most stable 1CBN:2TMP structures are competitive with the experimental structures. So, maybe there are other 1CBN:2TMP co-crystal structures that have not been found experimentally. However, the co-crystallisation energy differences between the observed 2CBN:1TMP, 1CBN:1TMP structures and the most stable 1CBN:2TMP hypothetical structures are small, considering the likely errors in the lattice energies (6.5) and that these errors do not cancel when comparing stoichiometries. The disorder in the pentane tail of 2CBN:1TMP suggests that thermal motion and dynamic disorder could be stabilising the 2CBN:1TMP structure. Indeed, given the closeness of the lattice energies in (Table 14) the relative stability of the co-crystals at ambient could favour the two observed structures.

The reason why the experimental structures cannot be found in chapter 5 is that the CSP did not treat the hydrocarbon pentane tail as sufficiently flexible. In chapter 4, combining the fragment search (Figure 14), CBN intramolecular



energy scans (Figure 12) and the computational limitation, it is decided to generate the LAM only between 140° and 220° for torsion angle (C16\_C15\_C14\_C8) (C17\_C16\_C15\_C14) (C18\_C17\_C16\_C15). This led to the search only covering structures with limited flexibility in the pentane tail as shown in (Figure 42) which could not generate the experimental structures. It seems there are different challenges in the CANNOL CSP and co-crystal CSP, as the analysis in Chapter 6 suggested that it was the hydroxyl group conformation that caused the problems with finding CANNOL.

### 7.3 Outlook for CSP

The ability to predict co-crystal formation and structures is important as co-crystals can be useful pharmaceutical materials (chapter 1). Previous work has shown that it is challenging. However, groups participating in the 7<sup>th</sup> Blind Test are testing whether recent CSP methods are suitable for co-crystals containing an API (CBN) and a much smaller co-former (TMP). Only 2 submissions from the companies that provide CSPs to industries had successfully predicted the 2CBN:1TMP co-crystal. These large teams of professionals used millions of CPU hours for the blind test. Hence CSP for pharmaceutical co-crystals is still a challenge and requires more work to make it accessible for academic research.

One proposal from this work is that CrystalPredictor<sup>18,19</sup> needs to be adapted to use distributed multipoles, particularly for molecules like coronene and CBN where there is reason to believe that the point charge model will not even produce a lattice energy minimum corresponding to the structure.

A model has been generated for atomic coordinates for MPYRAZ03. This might help solve the crystal structure from the powder pattern if someone did the experimental work on it.

Hence in this research project, this thesis has been contributing to the development of CSP methods for organic co-crystals.

## 8 References

- (1) Price, S. L. Is zeroth order crystal structure prediction (CSP<sub>0</sub>) coming to maturity? What should we aim for in an ideal crystal structure prediction code? *Faraday Discussions* **2018**, *211*, 9–30, 10.1039/C8FD00121A. DOI: 10.1039/C8FD00121A.
- (2) Day, G. M.; Cooper, A. I. Energy–Structure–Function Maps: Cartography for Materials Discovery. *Advanced Materials* **2018**, *30* (37), Article. DOI: 10.1002/adma.201704944.
- (3) Nyman, J.; Reutzel–Edens, S. M. Crystal structure prediction is changing from basic science to applied technology. *Faraday Discussions* **2018**, *211*, 459–476, Article. DOI: 10.1039/c8fd00033f.
- (4) McCrone, W. C. Polymorphism. In *Physics and Chemistry of the Organic Solid State*, Fox, D., Labes, M. M., Weissberger, A. Eds.; Vol. II; Wiley Interscience, 1965; pp 725–767.
- (5) *Polymorphism (materials science) – Wikipedia*. (accessed 2022 2nd Aug).
- (6) Iuzzolino, L. Survey of Crystallographic Data and Thermodynamic Stabilities of Pharmaceutical Solvates: A Step toward Predicting the Formation of Drug Solvent Adducts. *Crystal Growth & Design* **2021**, *21* (8), 4362–4371, Article. DOI: 10.1021/acs.cgd.1c00265.
- (7) Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. Ritonavir: An extraordinary example of conformational polymorphism. *Pharmaceutical Research* **2001**, *18* (6), 859–866.
- (8) Rietveld, I. B.; Ceolin, R. Rotigotine: Unexpected Polymorphism with Predictable Overall Monotropic Behavior. *Journal of Pharmaceutical Sciences* **2015**, *104* (12), 4117–4122, Article. DOI: 10.1002/jps.24626.
- (9) Threlfall, T. Structural and thermodynamic explanations of Ostwald's rule. *Organic Process Research & Development* **2003**, *7* (6), 1017–1027.
- (10) ten Wolde, P. R.; Frenkel, D. Homogeneous nucleation and the Ostwald step rule. *Physical Chemistry Chemical Physics* **1999**, *1* (9), 2191–2196.
- (11) Neumann, M. A.; van de Streek, J. How many ritonavir cases are there still out there? *Faraday Discussions* **2018**, *211*, 441–458, Article. DOI: 10.1039/c8fd00069g.
- (12) Wöhler; Liebig. Untersuchungen über das Radikal der Benzoesäure. *Annalen der Pharmacie* **1832**, *3* (3), 249–282. DOI: 10.1002/jlac.18320030302.
- (13) Thun, J.; Seyfarth, L.; Senker, J.; Dinnebier, R. E.; Breu, J. Polymorphism in Benzamide: Solving a 175–Year–Old Riddle. *Angewandte Chemie (International ed.)* **2007**, *46* (35), 6729–6731. DOI: 10.1002/anie.200701383.
- (14) Nyman, J.; Day, G. M. Static and lattice vibrational energy differences between polymorphs. *CrystEngComm* **2015**, *17* (28), 5154–5165, Article. DOI: 10.1039/c5ce00045a.
- (15) Abramov, Y.; Sun, G.; Zeng, Q. Emerging Landscape of Computational Modeling in Pharmaceutical Development. *Journal of Chemical Information and Modeling* **2022**, *62* (5), 1160–1171, Review. DOI: 10.1021/acs.jcim.1c01580.
- (16) Yuan, J.; Liu, X.; Wang, S.; Chang, C.; Zeng, Q.; Song, Z.; Jin, Y.; Zeng, Q.; Sun, G.; Ruan, S.; et al. Virtual coformer screening by a combined machine learning and physics–based approach. *CrystEngComm* **2021**, *23* (35), 639–644. DOI: 10.1039/d1ce00587a.
- (17) Sugden, I.; Braun, D.; Bowskill, D.; Adjiman, C.; Pantelides, C. Efficient Screening of Cofomers for Active Pharmaceutical Ingredient Cocrystallization. *Crystal Growth & Design* **2022**, Article|Early Access. DOI: 10.1021/acs.cgd.2c00433.
- (18) Sugden, I.; Adjiman, C. S.; Pantelides, C. C. Accurate and efficient representation of intramolecular energy in ab initio generation of crystal structures. I. Adaptive local approximate models. *Acta Crystallographica Section B–Structural Science Crystal Engineering and Materials* **2016**, *72*, 864–874, Article. DOI: 10.1107/S2052520616015122.
- (19) Sugden, I. J.; Adjiman, C. S.; Pantelides, C. C. Accurate and efficient representation of intramolecular energy in ab initio generation of crystal structures. II. Smoothed intramolecular

- potentials. *Acta Crystallographica Section B–Structural Science Crystal Engineering and Materials* **2019**, *75*, 423–433, Article. DOI: 10.1107/S2052520619005778.
- (20) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C. CrystalOptimizer. An efficient Algorithm for Lattice Energy Minimisation of Organic Crystal using Isolated–Molecule Quantum Mechanical Calculations. In *Molecular System Engineering*, Adjiman, C. S., Galindo, A. Eds.; Process Systems Engineering, Vol. 6; WILEY–VCH Verlag GmbH & Co., 2010; pp 1–42.
- (21) *The 7th CSP Blind Test – 2020 to 2022 – The Cambridge Crystallographic Data Centre (CCDC)*. (accessed 2022 2nd AUG).
- (22) Aitipamula, S.; Banerjee, R.; Bansal, A. K.; Biradha, K.; Cheney, M. L.; Choudhury, A. R.; Desiraju, G. R.; Dikundwar, A. G.; Dubey, R.; Duggirala, N.; et al. Polymorphs, Salts, and Cocrystals: What's in a Name? *Crystal Growth & Design* **2012**, *12* (5), 2147–2152. DOI: 10.1021/cg3002948.
- (23) *Cocrystal – Wikipedia*. (accessed 2022 2nd AUG).
- (24) Sun, G.; Jin, Y.; Li, S.; Yang, Z.; Shi, B.; Chang, C.; Abramov, Y. Virtual Coformer Screening by Crystal Structure Predictions: Crucial Role of Crystallinity in Pharmaceutical Cocrystallization. *Journal of Physical Chemistry Letters* **2020**, *11* (20), 8832–8838, Article. DOI: 10.1021/acs.jpcllett.0c02371.
- (25) Williams, H. D.; Trevaskis, N. L.; Charman, S. A.; Shanker, R. M.; Charman, W. N.; Pouton, C. W.; Porter, C. J. H. Strategies to Address Low Drug Solubility in Discovery and Development. *Pharmacological Reviews* **2013**, *65* (1), 315–499, Review. DOI: 10.1124/pr.112.005660.
- (26) Censi, R.; Di Martino, P. Polymorph Impact on the Bioavailability and Stability of Poorly Soluble Drugs. *Molecules* **2015**, *20* (10), 18759–18776, Review. DOI: 10.3390/molecules201018759.
- (27) Newman, A. Specialized Solid Form Screening Techniques. *Organic Process Research & Development* **2013**, *17* (3), 457–471.
- (28) McNamara, D. P.; Childs, S. L.; Giordano, J.; Iarriccio, A.; Cassidy, J.; Shet, M. S.; Mannion, R.; O'Donnell, E.; Park, A. Use of a Glutaric Acid Cocrystal to Improve Oral Bioavailability of a Low Solubility API. *Pharmaceutical Research* **2006**, *23*, 1888–1897.
- (29) Blagden, N.; de Matas, M.; Gavan, P.; York, P. Crystal engineering of active pharmaceutical ingredients to improve solubility and dissolution rates. *Advanced Drug Delivery Reviews* **2007**, *59* (7), 617–630.
- (30) Schultheiss, N.; Newman, A. Pharmaceutical Cocrystals and Their Physicochemical Properties. *Crystal Growth & Design* **2009**, *9* (6), 2950–2967.
- (31) Friscic, T.; Jones, W. Benefits of cocrystallisation in pharmaceutical materials science: an update. *Journal of Pharmacy and Pharmacology* **2010**, *62* (11), 1547–1559.
- (32) Allen, F. H. The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallographica Section B – Structural Science* **2002**, *58* (3), 380–388.
- (33) *CSP Blind Tests – The Cambridge Crystallographic Data Centre (CCDC)*. (accessed 2022 2nd AUG).
- (34) Zhang, P. Y.; Wood, G. P. F.; Ma, J.; Yang, M. J.; Liu, Y.; Sun, G. X.; Jiang, Y. A.; Hancock, B. C.; Wen, S. H. Harnessing Cloud Architecture for Crystal Structure Prediction Calculations. *Crystal Growth & Design* **2018**, *18* (11), 6891–6900, Article. DOI: 10.1021/acs.cgd.8b01098.
- (35) *GRACE (the Generation, Ranking and Characterisation Engine)*; Avant-garde Materials Simulation Deutschland GmbH: 2007. [www.avmatsim.eu](http://www.avmatsim.eu) (accessed).
- (36) Elking, D. M. Torque and atomic forces for Cartesian tensor atomic multipoles with an application to crystal unit cell optimization. *Journal of Computational Chemistry* **2016**, *37* (22), 2067–2080, Article. DOI: 10.1002/jcc.24427.
- (37) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B* **2016**, *72* (4), 439–459. DOI: doi:10.1107/S2052520616007447.

- (38) Williams, D. E. Improved intermolecular force field for molecules containing H, C, N, and O atoms, with application to nucleoside and peptide crystals. *Journal of Computational Chemistry* **2001**, *22* (11), 1154–1166.
- (39) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling Organic Crystal Structures using Distributed Multipole and Polarizability-Based Model Intermolecular Potentials. *Physical Chemistry Chemical Physics* **2010**, *12* (30), 8478–8490.
- (40) Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J.; Della Valle, R. G.; Venuti, E.; Jose, J.; et al. Significant progress in predicting the crystal structures of small organic molecules – a report on the fourth blind test. *Acta Crystallographica Section B – Structural Science* **2009**, *65* (2), 107–125.
- (41) Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S. X. M.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; Della Valle, R. G.; Desiraju, G. R.; et al. Towards crystal structure prediction of complex organic compounds – a report on the fifth blind test. *Acta Crystallographica Section B–Structural Science* **2011**, *67*, 535–551. DOI: 10.1107/s0108768111042868.
- (42) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L.; Galek, P. T. A.; Day, G. M.; Cruz-Cabeza, A. J. Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction. *International Journal of Pharmaceutics* **2011**, *418* (2), 168–178. DOI: 10.1016/j.ijpharm.2011.03.058.
- (43) Zhizhong, L. D. J. P. X. *solid state physics*; 2010.
- (44) *Crystal structure* – Wikipedia. (accessed 2022 2nd AUG).
- (45) *Crystallographic defect* – Wikipedia. (accessed 2022 2nd AUG).
- (46) *X-ray* – Wikipedia. (accessed 2022 2nd AUG).
- (47) Huang, S. *Solid Structural Analysis*; 1985.
- (48) Macrae, C. F.; Sovago, I.; Cottrell, S. J.; Galek, P. T. A.; McCabe, P.; Pidcock, E.; Platings, M.; Shields, G. P.; Stevens, J. S.; Towler, M.; et al. Mercury 4.0: from visualization to analysis, design and prediction. *Journal of Applied Crystallography* **2020**, *53*, 226–235. DOI: 10.1107/s1600576719014092.
- (49) Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. Mercury CSD 2.0 – new features for the visualization and investigation of crystal structures. *Journal of Applied Crystallography* **2008**, *41*, 466–470.
- (50) Etter, M. C.; MacDonald, J. C.; Bernstein, J. Graph-Set Analysis of Hydrogen-Bond Patterns in Organic Crystals. *Acta Crystallographica Section B – Structural Science* **1990**, *46*, 256–262.
- (51) *PLATON, A Multipurpose Crystallographic Tool*; Utrecht University: Utrecht, The Netherlands, 2003. (accessed).
- (52) *Packing Coefficients (ucl.ac.uk)*. (accessed 2022 2nd AUG).
- (53) Vandersluis, P.; Spek, A. L. Bypass – An Effective Method for the Refinement of Crystal-Structures Containing Disordered Solvent Regions. *Acta Crystallographica Section A – Foundations of Crystallography* **1990**, *46*, 194–201.
- (54) Gavezzotti, A. The Calculation of Molecular Volumes and the Use of Volume Analysis in the Investigation of Structured Media and of Solid-State Organic Reactivity. *Journal of the American Chemical Society* **1983**, *105* (16), 5220–5225.
- (55) Bondi, A. Van der Waals Volumes + Radii. *Journal of Physical Chemistry* **1964**, *68* (3), 441–451.
- (56) Ahmadi, S.; Mondal, P.; Wu, Y.; Gong, W.; Mirmehrabi, M.; Rohani, S. Virtual Multicomponent Crystal Screening: Hydrogen Bonding Revisited. *Crystal Growth & Design* **2021**, *21* (10), 5862–5872, Article. DOI: 10.1021/acs.cgd.1c00737.
- (57) McQuarrie, D. A. *Quantum chemistry / Donald A. McQuarrie*; University Science Books, 2008.

- (58) Lowe, J. P. *Quantum chemistry / John P. Lowe, Kirk A. Peterson*; Burlington, MA ; London : Elsevier Academic Press, 2006.
- (59) Moller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Physical Review* **1934**, *46*, 618–622.
- (60) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chemical physics letters* **1988**, *153* (6), 503–506. DOI: 10.1016/0009-2614(88)85250-3.
- (61) *TYC Materials Modelling course*. (accessed 2020 Oct).
- (62) *Density functional theory – Wikipedia*. (accessed 2020 Oct).
- (63) *Hohenberg–Kohn–Theorem – Wikipedia*. (accessed 2020 Oct).
- (64) Nityananda, R.; Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Resonance* **2017**, *22* (8), 809–811. DOI: 10.1007/s12045-017-0529-3.
- (65) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters* **1996**, *77* (18), 3865–3868.
- (66) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for mixing exact exchange with density functional approximations. *Journal of Chemical Physics* **1996**, *105* (22), 9982–9985, Article. DOI: 10.1063/1.472933.
- (67) Iuzzolino, L.; McCabe, P.; Price, S. L.; Brandenburg, J. G. Crystal structure prediction of flexible pharmaceutical-like molecules: density functional tight-binding as an intermediate optimisation method and for free energy estimation. *Faraday discussions* **2018**, *211*, 275–296. DOI: 10.1039/c8fd00010g.
- (68) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *PHYS REV B* **1995**, *51* (19), 12947–12957. DOI: 10.1103/PhysRevB.51.12947.
- (69) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *Journal of Chemical Theory and Computation* **2013**, *9* (1), 338–354. DOI: 10.1021/ct300849w.
- (70) Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method. *The Journal of Physical Chemistry A* **2007**, *111* (26), 5678–5684. DOI: 10.1021/jp070186p.
- (71) Elstner, M. SCC-DFTB: What Is the Proper Degree of Self-Consistency? *The Journal of Physical Chemistry A* **2007**, *111* (26), 5614–5621. DOI: 10.1021/jp071338j.
- (72) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J Chem Phys* **2010**, *132* (15), 154104. DOI: 10.1063/1.3382344.
- (73) Neumann, M. A.; Perrin, M. A. Energy ranking of molecular crystals using density functional theory calculations and an empirical van der Waals correction. *Journal of Physical Chemistry B* **2005**, *109* (32), 15531–15541.
- (74) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Physical Review Letters* **2009**, *102* (7), 073005.
- (75) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry* **2006**, *27* (15), 1787–1799.
- (76) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. J.; Refson, K.; Payne, M. C. First principles methods using CASTEP. *Zeitschrift fur Kristallographie* **2005**, *220* (5–6), 567–570.
- (77) Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshayé, M. Y.; Dumitrică, T.; Dominguez, A.; et al. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *The Journal of chemical physics* **2020**, *152* (12), 124101–124101. DOI: 10.1063/1.5143190.
- (78) Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalleri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D'Arco, P.; et al. CRYSTAL14: A program for the ab initio

- investigation of crystalline solids. *International Journal of Quantum Chemistry* **2014**, *114* (19), 1287–1317. DOI: 10.1002/qua.24658.
- (79) *Gaussian 09, Revision D.01*; 2009. (accessed).
- (80) Stone, A. J.; Alderton, M. Distributed Multipole Analysis – Methods and Applications. *Molecular Physics* **1985**, *56* (5), 1047–1064, Article. DOI: 10.1080/00268978500102891.
- (81) Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *Journal of Chemical Theory and Computation* **2005**, *1* (6), 1128–1132.
- (82) Day, G. M.; Motherwell, W. D. S.; Jones, W. Beyond the isotropic atom model in crystal structure prediction of rigid molecules: Atomic multipoles versus point charges. *Crystal Growth & Design* **2005**, *5* (3), 1023–1033.
- (83) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press, 1996.
- (84) Stone, A. J.; Misquitta, A. J. Atom–atom potentials from ab initio calculations. *International Reviews in Physical Chemistry* **2007**, *26* (1), 193–222.
- (85) Williams, D. E.; Cox, S. R. Nonbonded Potentials For Azahydrocarbons: the Importance of the Coulombic Interaction. *Acta Crystallographica Section B – Structural Science* **1984**, *40* (8), 404–417.
- (86) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *Journal of Physical Chemistry* **1996**, *100* (18), 7352–7360.
- (87) Beyer, T.; Price, S. L. Dimer or catemer? Low–energy crystal packings for small carboxylic acids. *Journal of Physical Chemistry B* **2000**, *104* (12), 2647–2655.
- (88) *NEIGHCRY*; In house built program: University College London, 2010. (accessed).
- (89) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C. Efficient Handling of Molecular Flexibility in Lattice Energy Minimization of Organic Crystals. *Journal of Chemical Theory and Computation* **2011**, *7* (6), 1998–2016.
- (90) Hoja, J.; Tkatchenko, A. First–principles stability ranking of molecular crystal polymorphs with the DFT plus MBD approach. *Faraday Discussions* **2018**, *211*, 253–274, Article. DOI: 10.1039/c8fd00066b.
- (91) Cervinka, C.; Beran, G. J. O. Ab initio prediction of the polymorph phase diagram for crystalline methanol. *Chemical Science* **2018**, *9*, 4622–4629, 10.1039/C8SC01237G. DOI: 10.1039/C8SC01237G.
- (92) Beran, G. J. O.; Nanda, K. Predicting organic crystal lattice energies with chemical accuracy. *Journal of Physical Chemistry Letters* **2010**, *1* (24), 3480–3487.
- (93) Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schutz, M.; Chan, H. Ab initio determination of the lattice energy in crystalline benzene to sub–kilojoule per mole accuracy. *Science* **2014**, *345* (6197), 640–643.
- (94) Cromer, D. T.; Ihde, A. J.; Ritter, H. L. The Crystal Structure of Tetramethylpyrazine. *Journal of the American Chemical Society* **1951**, *73* (12), 5587–5590. DOI: 10.1021/ja01156a025.
- (95) Braam, A. W. M.; Eshuis, A.; Vos, A. Tetramethylpyrazine at 300 and 100 K. *Acta Crystallographica Section B* **1981**, *37* (3), 730–732. DOI: 10.1107/S056774088100410X.
- (96) Duchateau, R.; Williams, A. J.; Gambarotta, S.; Chiang, M. Y. Carbon–carbon double–bond formation in the intermolecular acetonitrile reductive coupling promoted by a mononuclear titanium(II) compound. Preparation and characterization of two titanium(IV) imido derivatives. *Inorganic chemistry* **1991**, *30* (25), 4863–4866. DOI: 10.1021/ic00025a036.
- (97) Iuzzolino, L.; Reilly, A. M.; McCabe, P.; Price, S. L. Use of Crystal Structure Informatics for Defining the Conformational Space Needed for Predicting Crystal Structures of Pharmaceutical Molecules. *Journal of Chemical Theory and Computation* **2017**, *13* (10), 5163–5171. DOI: 10.1021/acs.jctc.7b00623.
- (98) *Mercury CSD 2.2 (Build RC5)*; 2009. <http://www.ccdc.cam.ac.uk/mercury/> (accessed).

(99) Holden, J. R.; Du, Z. Y.; Ammon, H. L. Prediction of Possible Crystal-Structures For C-, H-, N-, O- and F-Containing Organic Compounds. *Journal of Computational Chemistry* **1993**, *14* (4), 422-437.