



# **Experimental and Computational Approaches for Control and Prediction of Solid-State of Pharmaceuticals**

A thesis presented for the degree of

Doctor of Philosophy

in the Faculty of Science of the University of Strathclyde

by

Rajni Miglani Bhardwaj

Strathclyde Institute of Pharmacy and Biomedical Sciences

December, 2013

This thesis is a result of author's original research. It has been composed by the author and has not been previously submitted for an examination that has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.49. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

## Abstract

This thesis illustrates techniques for discovery of solid-state forms and probing the relationship between molecular structure and crystallisability. Also, the value of combined experimental and computational approaches to provide better understanding of the key factors underpinning the structural diversity in two groups each comprising of two structurally related pharmaceutical compounds is demonstrated.

An effective methodology of high throughput crystallisation and analysis for polymorph, solvate and salt screening using quartz 96/48 multi well plate with an automated system for collecting high quality Raman spectra was developed and validated. Using this efficient technique, 10 novel salts of amoxapine, 3 novel physical forms of clozapine and 16 novel solid forms of olanzapine were obtained by utilising a total of only ~640 mgs of API and ~65 ml of solvents.

A statistical model with ~70% prediction accuracy has been built for predicting the crystallisability of small organic molecules. This model is first of its type and provides an opportunity to identify problematic systems at early stages and would allow early targeting for improvements.

Structurally related molecules within each group were found to have markedly different experimental solid-state diversity after comprehensive physical form screening using multiple crystallisation techniques selected to maximise the crystallisation search space. Crystal structure prediction studies have been proved to be an important tool in rationalisation of the observed solid-state diversity. PIXEL calculations revealed that the largest contribution to crystal stabilisation comes from dispersion energy and enabled the identification of dominant intermolecular interactions in the crystal structures. Structural packing analysis using XPac and Mercury has enabled the structural relationship amongst all the crystal structures to be investigated. In case of olanzapine solvates XPac analysis provides a rationale of desolvation products by highlighting the close relationships between the forms and desolvated 'end product'. Statistical modelling analysis revealed that the physicochemical properties of the solvents were directing the crystal packing in olanzapine solvates.

## **Acknowledgements**

First and foremost, I would like to thank to my supervisor, Professor Alastair Florence for his invaluable guidance, encouragement, support and advice throughout the course of PhD. I would like to express my gratitude to Commonwealth Scholarship Commission in the United Kingdom (UK) and British Council for sponsoring my studies and all the support during my stay.

I also thank my co-supervisors Dr. Iain Oswald and Dr. Blair Johnston for their guidance and support throughout my research. The contribution of Professor Sarah Price and Louise Price for crystal structure prediction studies is also acknowledged. Numerous advices from Susan Reutzler-Edens during studies on olanzapine are appreciated. Many thanks to Dr. Andrea Johnston for her support and advice during statistical modelling and experimental work.

Many thanks to Dr. Gary Miller for his help with XPac analysis. Thanks to Mr. Vishal Raval for his timely help with the experimental work. Thanks are also due to Dr Julie Bardin with CASTEP calculations, Dr. Catriona Morrison regarding the use of PIXEL, Dr. Dimitrois Lamprou for his help during preparation of various surfaces, Master students Eugene and Paula for their help with the experimental work. The use of National Grid service facilities, UK for CASTEP calculations is also acknowledged.

Thanks are due to everyone from the solid-state research group, and CMAC particularly Dr. Jean-Baptiste Arlin, Dr. Scott Mckellar, Dr. Ryan Taylor, Naomi Briggs, Dr. Thomas Mcglone, Dr. Lihua Zhao, Dr. Anna Jawor-Baczynska and Lorna.

Finally, I would like to thank my family and friends for their support throughout my PhD.

# Table of Contents

Abstract.....	i
Acknowledgements.....	ii
Table of Contents .....	iii
List of Figures.....	ix
List of Tables.....	xx
List of Abbreviations .....	xxiii
<b>1 Introduction.....</b>	<b>1</b>
1.1 Background.....	2
1.2 Polymorphism.....	5
1.3 Multi-Component Crystalline Systems: Solvates, Salts and Co-crystals .....	7
1.4 Amorphous Solids .....	10
1.5 Importance of Solid-State Form for Pharmaceutical Industry .....	11
1.6 Crystallisation .....	12
1.7 Crystallisation Techniques .....	15
1.8 Control of Solid Form.....	17
1.9 Computational Methods .....	23
1.9.1 Crystal Structure Prediction.....	23
1.9.1.1 Monomorphic Crystal Energy Landscape.....	24
1.9.1.2 Predictive Crystal Energy Landscape.....	25
1.9.1.3 Complex Crystal Energy Landscape.....	26
1.9.1.4 Interchangeable Crystal Energy Landscape and Disorder .....	27
1.9.2 Energetics of Non-bonded Interactions: The PIXEL/Semi-classical Density Sums Methods.....	29
1.10 Experimental and Computational Studies on Structurally Related Compounds .....	33
<b>2 Aims and Objectives .....</b>	<b>41</b>
2.1 Aims .....	42
2.2 Objectives.....	44
<b>3 Material and Methods.....</b>	<b>45</b>
3.1 Material.....	46
3.2 Methods.....	46
3.2.1 Crystallisation Techniques.....	46

3.2.1.1	Slow and Fast Solvent Evaporation.....	47
3.2.1.2	Semiautomatic Cooling Crystallisation on React Array Platform.....	47
3.2.1.3	Mechanical Grinding.....	47
3.2.1.4	Crystallisation from Vapour Phase.....	47
3.2.1.5	Melt Quenching.....	48
3.2.1.6	Freeze Drying.....	48
3.2.2	X-ray Crystallography .....	49
3.2.2.1	X-ray Powder Diffraction .....	49
3.2.2.2	Single Crystal X-ray Diffraction Analysis .....	51
3.2.3	Thermal Analysis.....	52
3.2.4	Raman and Infrared Spectroscopy .....	52
3.2.5	PIXEL Calculations.....	53
3.2.6	Crystal Packing Analysis .....	54
<b>4</b>	<b>Development and Validation of High-Throughput Crystallisation and Analysis (HTCAA) Methodology for Physical Form Screening .....</b>	<b>55</b>
4.1	Introduction.....	56
4.2	Sample Preparation and Methodology Development .....	63
4.2.1	96/48 Quartz Multi-well Plate.....	63
4.2.2	Preparation of 96-well Plate for Salt Screening of Amoxapine.....	64
4.2.3	Preparation of 48-well Plate for Physical Form Screening of Clozapine .....	66
4.2.4	Preparation of 96-well Plate for Physical Form Screening of Olanzapine .....	68
4.2.5	Raman Microscopy .....	69
4.2.6	Chemometric Analysis .....	71
4.2.7	Scale-up and Characterisation of Novel Forms .....	73
4.3	Results and Discussion.....	74
4.3.1	Salt Screening of Amoxapine.....	74
4.3.1.1	Scale-up and Characterisation of Novel Salts of Amoxapine.....	79
4.3.2	Physical Form Screening of Clozapine .....	87
4.3.2.1	Experiment 1, PS-1.....	87
4.3.2.2	Experiment 2, PS-2.....	89
4.3.2.3	Characterisation of Novel Forms of Clozapine after Scale-up, PS-1 .....	91
4.3.3	Physical Form Screening of Olanzapine .....	94

4.3.3.1	Scale –up of Novel Physical Forms of Olanzapine.....	97
4.4	Key Findings of the Developed HTCAA Methodology .....	102
4.5	Summary .....	108
<b>5</b>	<b>Predicting Crystallisability of Organic Molecules using Statistical Modelling Techniques.....</b>	<b>110</b>
5.1	Statistical Modelling Techniques .....	111
5.1.1	Introduction .....	111
5.1.2	Principal Component Analysis.....	112
5.1.3	Criteria for Deciding the Number of Principal Components.....	114
5.1.4	Random Forests Classification Method .....	116
5.1.5	Applications of Statistical Modelling Techniques in Pharmaceutical Industry.....	119
5.1.5.1	Solvent Classification/Selection.....	120
5.1.5.2	Design of Experiment.....	122
5.1.5.3	Assessment of Completeness of Experimental Screens.....	122
5.1.5.4	Revealing Trends in Larger Datasets .....	123
5.1.5.5	Predicting Crystallisation of Small Molecules from the Melt .....	123
5.1.5.6	Predicting Crystallisability of Proteins.....	124
5.2	Descriptor Calculations, Model Building and Validation .....	126
5.2.1	Training Dataset and 2- and 3-Dimensional Descriptors Calculations.....	126
5.2.2	Training the Statistical Model.....	129
5.3	Results and Discussion.....	131
5.3.1	Principal Component Analysis.....	131
5.3.2	Random Forests Classification Model .....	132
5.3.3	Model Optimisation Attempts.....	137
5.3.4	Important Descriptors Assessment.....	137
5.3.5	Limitations of Random Forests Classification Model.....	140
5.4	Summary .....	142
<b>6</b>	<b>Exploring the Crystal Structure Landscape of Olanzapine .....</b>	<b>143</b>
6.1	Introduction.....	144
6.2	Experimental Procedures .....	149
6.2.1	Crystallisation.....	149
6.2.1.1	Cooling Crystallisation of Olanzapine using Automated Platform.....	150
6.2.1.2	Antisolvent Crystallisation of Olanzapine .....	151

6.2.1.3	Crystallisation of Olanzapine as Function of pH.....	151
6.2.1.4	Crystallisation of Olanzapine on Functionalised Surfaces.....	151
6.2.1.5	Spray Drying of Olanzapine.....	152
6.2.1.6	Recrystallisation from Amorphous Olanzapine.....	152
6.2.2	Variable Temperature-X-ray Powder Diffraction.....	153
6.2.2.1	Desolvation of Olanzapine Solvates.....	153
6.2.3	XPac Analysis of Crystal Structures of Olanzapine.....	153
6.2.4	Random Forests Classification Model of Olanzapine Solvates.....	154
6.3	Calculation and Analysis of the Crystal Energy Landscape (Crystal Structure Prediction) of Olanzapine.....	155
6.4	Results and Discussion.....	159
6.4.1	Forms I, II and III of Olanzapine.....	159
6.4.2	Solution Crystallisation of Olanzapine.....	163
6.4.3	Neat and Liquid-Assisted Grinding of Olanzapine.....	172
6.4.4	Desolvation of Olanzapine Solvates.....	173
6.4.5	Melt Quenching and Recrystallisation of Amorphous Olanzapine.....	174
6.4.6	Spray Drying and Freeze Drying of Olanzapine.....	177
6.5	Molecular Packing Analysis of Olanzapine Crystal Structures using XPac.....	178
6.6	Hydrogen-Bonding Analysis of Olanzapine Crystal Structures.....	183
6.6.1	Hydrogen Bonding in Olanzapine Solvates Based on Packing Type SC <sub>31</sub> ....	186
6.6.2	Hydrogen Bonding in Olanzapine Solvates Based on Packing Type SC <sub>32</sub> ....	187
6.6.3	Hydrogen Bonding in Olanzapine Solvates Based on Packing Type, SC <sub>33</sub> ...	188
6.6.4	Hydrogen Bonding in Other Olanzapine Solvates Based on SC <sub>0</sub> , SC <sub>11</sub> and SC <sub>22</sub> .....	189
6.7	Prediction of Olanzapine Solvate Formation using Random Forests Classification Model.....	190
6.7.1	Random Forests Classification Model – Prediction Results.....	191
6.7.2	Important Solvent Physicochemical Descriptors for Random Forests Classification Model of Olanzapine Solvates.....	196
6.8	Crystal Energy Landscape of Olanzapine.....	198
6.9	PIXEL Calculations on Olanzapine.....	203
6.10	Concomitant Appearance of Form III with other Polymorphs of Olanzapine.....	208
6.11	Prolific Solvate Formation of Olanzapine.....	209



6.12	Challenges with Crystal Structure Prediction of Olanzapine and Unobserved Calculated Structures.....	212
6.13	Summary.....	214
<b>7</b>	<b>Exploring the Physical Form Landscape of Clozapine, Amoxapine and Loxapine.....</b>	<b>217</b>
7.1	Introduction.....	218
7.1.1	Background of Molecule in Group 1-Clozapine.....	219
7.1.2	Background of Molecules in Group 2-Amoxapine and Loxapine.....	220
7.2	Experimental Details.....	221
7.2.1	Principal Component Analysis of Solvent Properties.....	221
7.2.2	Crystallisation Experiments of Clozapine.....	222
7.2.3	Crystallisation Experiments of Amoxapine.....	223
7.2.4	Crystallisation Experiments of Loxapine.....	224
7.2.5	Preliminary Crystal Structure Prediction Studies for Clozapine, Amoxapine and Loxapine.....	225
7.2.6	Solid-State Calculations using CASTEP.....	226
7.2.7	Structure Analysis.....	227
7.3	Results and Discussion.....	228
7.3.1	Physical Form Screening of Clozapine.....	228
7.3.1.1	Molecular Packing Analysis of Physical Forms of Clozapine.....	233
7.3.1.2	Preliminary Crystal Structure Prediction of Clozapine.....	240
7.3.1.3	PIXEL Calculations on Clozapine.....	241
7.3.2	Physical Form Screening Results of Amoxapine.....	248
7.3.2.1	Molecular Packing Analysis of Crystal Structure of Amoxapine.....	249
7.3.2.2	Preliminary Crystal Structure Prediction of Amoxapine.....	251
7.3.2.3	PIXEL Calculations on Amoxapine.....	252
7.3.3	Physical Form Screening Results of Loxapine.....	255
7.3.3.1	Purity of Supplied Material of Loxapine.....	255
7.3.3.2	Crystal Structure Determination of Anhydrous Loxapine Succinate using XRPD.....	256
7.3.3.3	Physical Form Screening Results on Loxapine Free Base.....	260
7.3.3.4	Molecular Packing Analysis of Form I and II of Loxapine Free Base.....	264
7.3.3.5	Preliminary Crystal Structure Prediction of Loxapine Free Base.....	267
7.3.3.6	PIXEL Calculations on Form I and II of Loxapine Free Base.....	268

7.4 Summary.....	273
<b>8 Conclusions and Further Work .....</b>	<b>275</b>
8.1 Conclusions and Further Work.....	276
8.1.1 Development and Validation of High Throughput Crystallisation and Analysis Methodology for Physical Form Screening .....	280
8.1.2 Predicting Crystallisability of Organic Molecules using Statistical Modelling Techniques .....	281
8.1.3 Exploring the Physical Form Landscape of Structurally Related Pharmaceutical Molecules in Group 1 (Olanzapine and Clozapine) and 2 (Amoxapine and Loxapine).....	283
8.2 Further Work .....	287
8.2.1 Development and Validation of High Throughput Crystallisation and Analysis Methodology for Physical Form Screening .....	287
8.2.2 Predicting Crystallisability of Organic Molecules using Statistical Modelling Techniques .....	289
8.2.3 Exploring the Physical Form Landscape of Structurally Related Pharmaceutical Molecules in Group 1 (Olanzapine and Clozapine) and 2 (Amoxapine and Loxapine).....	290
<b>9 References .....</b>	<b>293</b>
Appendix .....	317
List of Publications.....	318

## List of Figures

Figure 1.1. Schematic illustration of molecular arrangements of different solid forms showing ordered arrangement in crystalline forms and molecule disorder in amorphous form. Reproduced from reference (Florence, 2009a). .....	5
Figure 1.2. (a) Molecular structure of carbamazepine, (b) dimer motif of carbamazepine present in form I-IV, and (c) catameric motif of carbamazepine present in form V. Adapted from references (Florence et al., 2006; Arlin et al., 2011). .....	6
Figure 1.3. (a) Extended conformation of L-glutamic acid which leads to the $\beta$ -polymorph (CSD refcode: LGLUAC11), and (b) twisted conformation of L-glutamic acid which leads to $\alpha$ -polymorph (CSD refcode: LGLUAC03).....	7
Figure 1.4. Multi-component systems and their overlap with each other. Reproduced from reference (Aitipamula et al., 2012). .....	8
Figure 1.5. Plausible scenarios of crystallisation outcome in a dimorphic system under different circumstances. Polymorph A has higher free energy barrier ( $\Delta G$ ) to nucleation than polymorph B and is the stable form. Adapted from reference (Aaltonen et al., 2009). .....	15
Figure 1.6. Timescale of various crystallisation techniques, which is also highlighting the techniques which can be used to obtain stable and metastable crystal forms. Adapted from reference (Anderton, 2007; Llinàs and Goodman, 2008). .....	16
Figure 1.7. Schematic illustration of factors controlling the polymorphic outcome in crystallisation processes. Reproduced from reference (Kitamura, 2003). .....	19
Figure 1.8. Schematic examples of the crystal energy landscapes; a, b, c and d denote monomorphic, predictive, complex and interchangeable crystal energy landscapes respectively. Each point denotes a crystal structure that is a local energy minimum, with the symbols representing significantly different type of packing such as different hydrogen bonding motifs. The experimentally known structures are denoted by empty red symbols. Reproduced from reference (Price, 2009).....	25
Figure 1.9. Chemical Structures of (a) hydrochlorothiazide and (b) chlorothiazide. ....	34
Figure 1.10. Polymorphs and hydrogen bonding motifs observed in carbamazepine and its analogues. The number of polymorphs having particular motif are shown in parentheses. *For oxycarbamazepine, only crystal structures of two polymorphs are known.....	36
Figure 1.11. Chemical structures of (a) fenamic acid (monomorphic) and (b) tolfenamic acid (5 polymorphs).....	38
Figure 1.12. Molecular structures of fenamic acid analogues. The number of structurally characterised polymorphs for each compound is written in parentheses. Reproduced from reference (López-Mejías et al., 2012). .....	39

Figure 2.1. The four compounds studied in this thesis. The differences in the chemical structures within a group are highlighted in red.....	43
Figure 4.1. Illustration of experimental workflow for high-throughput crystallisation and analysis methodology.....	59
Figure 4.2. Calculated pKa of nitrogen atoms in amoxapine.....	62
Figure 4.3. Chemical structures of olanzapine and clozapine.....	63
Figure 4.4. Layout of (a) salt screening in the 96-well plate, and (b) anti-solvent screening in the 48-well plate. A polymorph screen could be carried out using a different solvent in each well for example.....	64
Figure 4.5. DXR Raman microscope with motorised x, y, z stage and multi-well plate reader.....	69
Figure 4.6. A screenshot of the data acquisition interface set-up for use with the bespoke 96-well quartz plate. Each circle in the 8 x 12 grid (top left) represents a well on the plate and each square within a well represents a point of data collection (i.e. a Raman spectrum). The point (spot size-1 $\mu\text{m}$ ) from where the spectrum is being collected is shown by a plus sign on the image (top right) with the respective Raman spectrum shown (bottom). Spectra in the grid are coloured based on the magnitude of the measured Raman intensity. Red and violet colours indicate the highest and lowest intensity respectively. ....	71
Figure 4.7. Salt screening results of amoxapine after group analysis on a 96-well plate. Each square (spectrum) from the wells is colour coded according to classification group. Group 1 to 11 are shown by blue, fluorescent green, dark orange, cyan, green, yellow, orange, dark yellow, light green, pink-red and violet colour respectively. All other spectra shown by violet colour are outliers.....	75
Figure 4.8. Characteristic Raman spectra for 10 novel salts observed during salt screening of amoxapine. Each spectrum is labelled according to the well number from which they have been obtained on the 96-well plate.....	76
Figure 4.9. Stack plots of Raman spectra collected from samples of salts of (a) amoxapine and succinic acid, (b) amoxapine and acetic acid, at 96-well plate and on 150 mg scales. The Raman spectra of salts obtained from 96-well plate and on 150 mg scales are labelled in black and blue colour respectively.....	80
Figure 4.10. Stack plot of XRPD patterns of the 10 novel amoxapine salts obtained on 150 mg scales in the range of 4-35° 2 $\theta$ . ....	81
Figure 4.11. Microscope images (20x) of the M2 crystals observed in wells C8, D4 and D12 of 96-well plate.....	84
Figure 4.12. Stack plots of Raman spectra collected from samples of salts of (a) amoxapine and maleic acid and (b) amoxapine and para-toluene sulfonic acid at 96-well plate and on 150 mg scales. The Raman spectra of salts obtained from 96-well plate and on 150 mg scales are labelled in black and blue respectively. ....	85

Figure 4.13. Physical form screening (PS-1) results of clozapine after group analysis on the 48-well plate using the conditions shown in Table 4.12. The wells are colour coded according to classification group. Group 1, 2, 3, 4 and 5 are shown by green, blue, orange, red and yellow colour, respectively. All other spectra shown by magenta colour are outliers.....	88
Figure 4.14. Overlay of Raman spectra in the region of 1450-1700 $\text{cm}^{-1}$ of Raman shift for four physical forms (A, B, C and D) observed during physical form screening of clozapine.....	89
Figure 4.15. Physical form screening (PS-2) results of clozapine after group analysis on the 48-well plate using the conditions shown in Table 4.13. The wells are colour coded according to classification group. Group 1, 2, 3 and 4 are shown by blue, green, orange and yellow colour, respectively. All other spectra shown by magenta colour are outliers.....	90
Figure 4.16. Stack plot of the Raman spectra collected from samples of physical forms of clozapine observed on the 48-well plate and on 150 mg scales. The Raman spectra are labelled according to the source of their origin.....	92
Figure 4.17. Stack plot of the characteristic XRPD patterns of the 3 physical forms of clozapine obtained on 150 mg scales in the range of $4\text{-}35^\circ 2\theta$ . CZPN_anhydrous, CZPN_amorphous and CZPN_monohydrate represent form A, C, and B observed on the 48-well plate.....	93
Figure 4.18. Characteristic Raman spectra in the $1480\text{-}1650 \text{ cm}^{-1}$ region for 7 novel physical forms and form I of olanzapine identified during HTCAA. Spectra are coloured and labelled according to the well from which they have been detected.....	96
Figure 4.19. Stack plot of Raman spectra collected from samples of physical forms of olanzapine at 96-well plate (black colour) and 150 mg scales (blue colour). .....	98
Figure 4.20. Raman spectra of isostructural methanol and ethanol solvates of olanzapine. ....	99
Figure 4.21. Raman spectra of isostructural hydrated solvates of olanzapine.....	100
Figure 4.22. Stack plot of XRPD patterns of the 16 novel solvates and form I of olanzapine obtained on 150 mg scales. Abbreviations used in the figure are: THF-tetrahydrofuran, $\text{CH}_3\text{CN}$ -acetonitrile, MXE-2-methoxyethanol, DMSO-dimethylsulfoxide, TFE-2,2,2-trifluoroethanol and IAA-isoamyl alcohol. ....	101
Figure 4.23. Raman spectra of the sample exhibiting fluorescence (red) and after applying fluorescence correction (black).....	106
Figure 5.1. Schematic workflow of various stages of a typical statistical modelling process.....	111
Figure 5.2. Principal Component Analysis. Reproduced from reference (Wiklund, 2008). .....	112
Figure 5.3. Example of Score and loading plots obtained from Principal Component Analysis while investigating the patterns of food consumption of 20 food products in 16	

countries. Score and loading plots relate to observations (countries in this case) and variables (food products in this case) respectively. Reproduced from reference (Wiklund, 2008).....	114
Figure 5.4. Scree plot of the number of principal components vs. eigenvalues. Adapted from reference (Xu and Redman-Furey, 2007).....	115
Figure 5.5. Schematic workflow of building of Random Forests model.....	117
Figure 5.6. Basic skeleton of acylanilide molecules, R and X represent the two points of modifications. X includes H, CH <sub>3</sub> , C <sub>2</sub> H <sub>5</sub> , C <sub>3</sub> H <sub>7</sub> , C(CH <sub>3</sub> ) <sub>3</sub> , CF <sub>3</sub> , OCH <sub>3</sub> , OC <sub>2</sub> H <sub>5</sub> NH <sub>2</sub> and Cl. R includes H, CH <sub>3</sub> , C <sub>2</sub> H <sub>5</sub> , C(CH <sub>3</sub> ) <sub>3</sub> , OCH <sub>3</sub> , OC <sub>2</sub> H <sub>5</sub> , OCF <sub>3</sub> , F, Cl, Br, I, CF <sub>3</sub> , OH, NH <sub>2</sub> and COOH. ....	127
Figure 5.7. Scatter plot of properties of molecules, shown with an ellipse that represents the Hotelling T <sub>2</sub> with 95% confidence. Each of the 382 points on the plot represents a molecule from the dataset. Molecules are coloured according to the crystallisation outcome: class 1 (crystals, red) and class 2 (no crystals, black). ....	131
Figure 5.8. Error plot for the Random Forests classification model trained using 2-D and 3-D molecular descriptors of the molecules present in the dataset. Blue line shows the evolvement of overall OOB error of prediction with the addition of number of trees. ....	132
Figure 5.9. The multidimensional scaling plot of the scaling coordinates (1,2) obtained from the Random Forests classification proximity matrix. Each of the 382 points on the multidimensional scaling plot represents a molecule from the dataset and is coloured according to crystallisation outcome: class 1 (red) and class 2 (blue). 1 and 2 denote the two separate zones correspond to molecules from class1 and class 2. 3 (encircled in green) denotes the overlapped zone with molecules from both classes. ....	134
Figure 5.10. Convex hull plot of scaling coordinates obtained from Random Forests proximity matrix. Molecules in class 1 and 2 are represented by red and green points respectively. The cross sign is the mean of MDS1 and MDS2 for each group. ....	135
Figure 5.11. Two different conformation of propyl chains in 4-methylbutylanilide. Reproduced from reference (Hursthouse et al., 2009).....	138
Figure 5.12. Alternate conformations of meta-substituted acylanilides. The rotation is around aromatic carbon to N bond. Reproduced from reference (Hursthouse et al., 2009). ....	139
Figure 6.1. Molecular structure of olanzapine.....	144
Figure 6.2. Centrosymmetric dimer observed in all the experimental crystal structures of olanzapine. ....	148
Figure 6.3. Common ordered sets of points selected for XPac analysis.....	154
Figure 6.4. The difference between the observed conformation (region A, coloured by element) and the alternative low energy conformation (region B, coloured by orange) depicted by overlaying (left) the benzodiazepine or (right) the piperazine rings, of the gas phase optimised conformations. ....	156

Figure 6.5. The CrystalPredictor grid of intramolecular energies $\Delta E_{\text{intra}}$ , in kJ mol <sup>-1</sup> relative to the lowest energy point in each region shown in Figure 6.4. ....	156
Figure 6.6. The molecular structure of olanzapine, showing all degrees of flexibility considered in final CrystalOptimizer refinements (C <sub>11</sub> -C <sub>10</sub> -N <sub>3</sub> -C <sub>12</sub> , C <sub>10</sub> -N <sub>3</sub> -C <sub>12</sub> -C <sub>13</sub> , H-C <sub>17</sub> -N <sub>4</sub> -C <sub>13</sub> , C <sub>17</sub> -N <sub>4</sub> -C <sub>13</sub> -C <sub>12</sub> and H-C <sub>16</sub> -C <sub>2</sub> -C <sub>1</sub> ). ....	158
Figure 6.7. Overlay of XRPD patterns of form I (experimental, black), form II (simulated from room temperature single crystal structure, blue) and III + II (experimental, green) of olanzapine. Characteristic peaks of form III are marked. ....	160
Figure 6.8. DSC thermograms from samples of olanzapine (a) form I, (b) form II, sample comprising of several single crystals, (c) form III + II. ....	161
Figure 6.9. Raman spectra of form I (blue), form II (red) and form III + II (green). ....	162
Figure 6.10. FT-IR spectra of form I, II (major) + III (minor), III (major) + II (minor) and amorphous phase of olanzapine. ....	163
Figure 6.11. XRPD patterns of the 18 novel isostructural solvates (belong to crystal structures in space group C2/c) observed during experimental screen and a reference pattern of the known olanzapine DMSO hydrate. ....	170
Figure 6.12. XRPD patterns of the four novel olanzapine solvates obtained from crystallisation on the automated platform. ....	171
Figure 6.13. Pawley fit in TOPAS Academic (version 4.1), observed profile (blue), calculated profile (red) and difference plot (black) of the Pawley fit for form II in the range of 6-39° 2 $\theta$ . R <sub>wp</sub> = 2.461. ....	172
Figure 6.14. 3-D overlay of the XRPD patterns during <i>in-situ</i> desolvation of olanzapine DCM solvate at 318 K, showing appearance of peaks of form III and II and disappearance of peaks of DCM solvate. ....	174
Figure 6.15. XRPD data of amorphous olanzapine in the region of 2-40° 2 $\theta$ . ....	175
Figure 6.16. DSC data of amorphous olanzapine. Data was collected from 263 K to 483 K at a heating rate of 20 K min <sup>-1</sup> . ....	175
Figure 6.17. XRPD data in the range 6-30° 2 $\theta$ collected from melt-quenched sample of olanzapine recrystallised within boro-silicate capillary or in a vial held at RT or at 313 K. ....	176
Figure 6.18. Diagram showing the similarity relationship between 35 structures of olanzapine with structures identified by the crystal structure numbers used in Table 6.7. ....	180
Figure 6.19. Packing diagrams highlighting the supramolecular constructs (SCs) and their relationships highlighted in Figure 6.18. The directions of the primary vectors are indicated by arrows and in the SC <sub>21</sub> layer the SC <sub>0</sub> pairs are differentiated by light or dark blue shading. As the three-dimensional arrays correspond to three distinct packing modes of the SC <sub>21</sub> layers these are all viewed along the direction of the $t_2$ vector with the constituent SC <sub>21</sub> layers differentiated by dark and light blue shading. The three	

one-dimensional SCs are coloured green (SC<sub>11</sub>), purple (SC<sub>12</sub>) and red (SC<sub>13</sub>) respectively and are viewed perpendicular to the axis of their translation vectors. In SC<sub>31</sub> and SC<sub>32</sub>, the SC<sub>21</sub> layers are related by the SC<sub>13</sub> and SC<sub>11</sub> constructs respectively and a single occurrence of each, viewed perpendicular to its translation vector, is highlighted with the corresponding colour. A single SC<sub>22</sub> layer is shown with the SC<sub>12</sub> component running approximately vertical, highlighted in purple, with three SC<sub>11</sub> constructs running horizontally and differentiated by dark and light green shading. 182

Figure 6.20. The five different packing arrangements adopted by the SC<sub>11</sub> stacks, these are viewed along the axis of the  $t_4$  vector (which run perpendicular to page) with each pair of face-to-face molecules (coloured yellow) representing a single SC<sub>11</sub> stack. In the dichloromethane solvate the dark red molecules represent SC<sub>11</sub> columns in which the  $t_4$  vector runs perpendicular to those in the neighbouring layers, in form I and dihydrate D, a SC<sub>22</sub> layer is shown by the green molecules. .... 183

Figure 6.21. Different arrangements of SC<sub>11</sub> in (a) form I and (b) II of olanzapine, which share the same H-bonding (shown by green dotted line). (c) A closer view at the olanzapine atoms involved in H-bonding . Other hydrogen atoms are omitted for clarity. .... 184

Figure 6.22. Three water H-bonding environments observed in family of olanzapine solvate crystal structures, (D-Donor, A-acceptor). Weak interactions such as C-H...O were not included in the analysis. Reproduced from reference (Clarke et al., 2010)... 185

Figure 6.23. (a) Hydrogen bonds (N1H...N4) between centrosymmetric dimers, conciliated by water molecules present within SC<sub>21</sub> sheet of olanzapine 1,4-dioxane solvate hydrate. H-bonds are shown by blue dotted lines (b) organic solvent molecule surrounded by 4 pairs of centrosymmetric dimers from adjacent SC<sub>21</sub> sheets (coloured by blue and cyan), arranged in a staggered manner. Other hydrogen atoms are omitted for clarity..... 186

Figure 6.24. (a) Olanzapine acetic acid solvate in which acetic acid links the centrosymmetric dimers through N1H...N4 H-bonds. (b) Olanzapine ethylene glycol solvate, where each olanzapine molecule is involved in three H-bonds (N1H, N2 and N4) with solvent molecules. H-bonds are shown by blue dotted lines. Other hydrogen atoms are omitted for clarity..... 187

Figure 6.25. (a) Catemer arrangement of water molecules in olanzapine t-butanol solvate hydrate. (b) Water molecules exist in type 2 environment in olanzapine TBMe solvate hydrate. H-bonds are shown by blue dotted lines. Other hydrogen atoms are omitted for clarity..... 188

Figure 6.26. H-bonding (N1H...N4) in (a) olanzapine 2-butoxyethanol solvate (b) olanzapine dihydrate D, where water is present as cyclic tetramer (c) cyclic tetramer, where water is present in type 2 and type 3 environments. H-bonds are shown by blue colour. Other hydrogen atoms are omitted for clarity..... 189

Figure 6.27. Three 3-dimensional supramolecular constructs, (a) SC<sub>31</sub>, (b) SC<sub>32</sub> and (c) SC<sub>33</sub>, formed by SC<sub>21</sub> sheets (coloured by cyan and blue colour) in olanzapine solvates. .... 190



Figure 6.28. Representative crystal structures of olanzapine solvates based on (a) SC <sub>31</sub> -olanzapine-1,4-dioxane solvate hydrate; (b) SC <sub>32</sub> -Olanzapine-acetic acid solvate and (c) SC <sub>33</sub> -Olanzapine-TBMe hydrate; showing the position of solvent molecules within the crystal lattice.....	191
Figure 6.29. Plot of Random Forest classification error for molecular packing types of olanzapine solvates. The black line represents the error associated with the overall model. Red, blue and green lines represent the error associated with the molecular packing type classes; SC <sub>31</sub> , SC <sub>32</sub> and SC <sub>33</sub> . .....	192
Figure 6.30. The MDS plot obtained from the Random Forest classification proximity matrix. Each of the 24 points represents a olanzapine solvate and is coloured according to crystal packing type: class 1 (red), class 2 (green) and class 3 (blue).....	193
Figure 6.31. Variable dependency plots (SMR, apol, VdW_vol and b_count) for molecular packing types, SC <sub>31</sub> , SC <sub>32</sub> and SC <sub>33</sub> . Numerical values for solvent descriptors are summarised in Table 6.9. ....	197
Figure 6.32. Summary of crystal structure search for olanzapine. Each point represents a crystal structure classified by the conformational region, A or B, as defined in Figure 6.4, and whether it contains the centrosymmetric dimer SC <sub>0</sub> . The open red diamond and square symbols denote lattice energy minima corresponding to experimental structures, forms I and II respectively. The crystal energy landscape is comprised of the structures below -125 kJ mol <sup>-1</sup> . A11 and A45 are the computed structures corresponding to form I and II respectively.....	199
Figure 6.33. Overlay of (a) experimental form I (colour by element) with A11 (green) and (b) experimental form II (colour by element) with A45 (green). The RMSD <sub>15</sub> of these overlays are (a) 0.170 Å and (b) 0.225 Å.....	200
Figure 6.34. Comparison of (a) form I (calculated structure A11) and (b) calculated structure A1392, showing the arrangements of the common 2-D SCs; SC <sub>23</sub> (based on stacking of SC <sub>12</sub> ). The constructs are viewed along the axis of the <i>t</i> <sub>6</sub> vector with the <i>t</i> <sub>8</sub> axis (red arrows) running left-right across the page for the central layers, which are shown in black. In form I the upper and lower layers (grey) stack in parallel to the middle layer (black) while in 1392 they are rotated such that the top and bottom layers pack with opposite direction to that of the central layer. SC <sub>23</sub> is not defined in Figure 6.18, as it only occurs in one of the experimental structures, i.e. form I.....	201
Figure 6.35. Comparison of (a) form II and (b) A162, showing the arrangement of the common 2-D SC, SC <sub>24</sub> . The central constructs (black) are viewed along the axis of the <i>t</i> <sub>11</sub> vector (shown by the green circles) with the <i>t</i> <sub>10</sub> axis running left-right across the page as indicated by the red arrows. The upper and lower layers in form II (grey) run parallel with <i>t</i> <sub>11</sub> axis of the central construct while in A162 alternate layers pack in opposite directions (- <i>t</i> <sub>11</sub> indicated by the blue circles). SC <sub>24</sub> , not shown in Figure 6.18 as this 2-D construct was only observed in form II of the experimental structures included in the XPac analysis. ....	202

Figure 6.36. Multi-phase Pawley fit of unit cell of A162 against XRPD data of mixture of form III (major) and II (minor), observed profile (blue), calculated profile (red) and difference plot (black) in the range of 6-39° 2θ. $R_{wp} = 6.152$ .....	203
Figure 6.37. Top three pairwise intermolecular interactions in form I and form II, quantified in Table 6.11. The central molecule (colour by atom type and labelled as 'C') interacts with neighbouring molecule to form a a) centrosymmetric dimer, SC <sub>0</sub> , b) molecular pair involving hydrogen-bond (N1-H...N2), and c) molecular pair involving weaker molecule-molecule interaction.....	206
Figure 6.38. Packing coefficient (%) for the non-solvated polymorphs (black), co-crystals (orange) and solvates of OZPN. Solvate structures are coloured according to structural relationships. Specifically, SC <sub>31</sub> , SC <sub>32</sub> and SC <sub>33</sub> are shown by red, blue and green bars.....	211
Figure 7.1. The four compounds studied in this thesis. Differences in the chemical structures within a group are highlighted in red.....	218
Figure 7.2. Principal component analysis score plot of solvent physicochemical properties, shown with an ellipse that represents the Hotelling T <sub>2</sub> with 95% confidence. 1-methylnaphthalene, dodecane, trichloroethylene and carbon tetrachloride were outliers from the 95% confidence limit. Each number in the plot represents one solvent as mentioned below the score plot. ....	222
Figure 7.3. Conformation 'A' of (a) clozapine, (b) amoxapine and (c) loxapine used for preliminary crystal structure prediction study and is also present in the experimental crystal structures (CSD refcodes: NDNHCL01, AMOXAP and NDOCLH01 for clozapine, amoxapine and loxapine respectively).....	226
Figure 7.4. Scatter plot of solvents with positions according to the value of t[1] and t[2] from Figure 7.2 and coloured according to crystallisation outcome: clozapine anhydrous form (blue), clozapine monohydrate (magenta), mixture of clozapine anhydrous form and clozapine monohydrate (green), amorphous (orange), no sample obtained (cyan) and solvated form (maroon). Solvents not used in crystallisation condition are coloured grey. Plots shown for crystallisation conditions listed in Table 7.1: (a) solvent evaporation in vials at RT and (b) solvent evaporation on watch glass at RT.....	229
Figure 7.5. Scatter plot of solvents with positions according to the value of t[1] and t[2] from Figure 7.2 and coloured according to crystallisation outcome: clozapine anhydrous form (blue), clozapine monohydrate (magenta), mixture of clozapine anhydrous form and clozapine monohydrate (green), amorphous phase (orange), no sample obtained (cyan) and solvated form (maroon). Solvents not used during crystallisation condition are coloured grey. Plots shown for crystallisation conditions listed in Table 7.1: (a) cooling crystallisation and (b) liquid-assisted grinding.....	230
Figure 7.6. Stack plot of XRPD patterns of physical forms of clozapine observed during experimental screening in the range of 5-35° 2θ.....	232

Figure 7.7. STA thermograms of (a) clozapine anhydrous form, (b) clozapine N,N-dimethylacetamide and (c) clozapine monohydrate. DSC and TG curves are shown by black and blue colour, respectively.....	233
Figure 7.8. Overlay of the molecular conformations of olanzapine and clozapine which are present in the experimental crystal structures (RMSD of non-hydrogen atoms = 0.21 Å).....	234
Figure 7.9. (a-b) Two views of the dimer formed by clozapine molecules, (c) 3-D arrangement of columns of clozapine dimers. Active H-bonds between dimers are shown by dotted black lines. Other hydrogen atoms are omitted for clarity.....	235
Figure 7.10. (a) 2-D sheet structure formed by the two antiparallel chains of clozapine molecules (shown by red and blue colour) and N,N-dimethylacetamide molecules. H-bonds are shown by black dotted lines. (b) 3-D arrangement of sheets. All the hydrogen atoms are omitted for clarity.....	236
Figure 7.11. 3-D arrangement formed by the columns of the centrosymmetric dimers of clozapine in (a) clozapine monohydrate and (b) clozapine ethylene glycol solvate. H-bonds between clozapine and solvent molecules are shown by the blue dotted line. Other hydrogen atoms are omitted for clarity.....	237
Figure 7.12. Packing coefficient (%) for the anhydrous form (red), solvates (orange), analogues (green) of clozapine and anhydrous polymorphs (blue) of olanzapine. ....	238
Figure 7.13. Chemical Structures of clozapine and its analogues. ....	239
Figure 7.14. Summary of the preliminary crystal structure search of clozapine. Each point represents a hypothetical crystal structure/lattice energy minimum. The cyan rectangle with open red diamond represents the global minimum in the lattice energy, which corresponds to the known experimental structure.....	240
Figure 7.15. The top three molecular pairs present in the crystal lattice of clozapine identified in Table 7.6. (a) dimer involving hydrogen-bond between clozapine molecules (shown by orange line), (b) side-ways dimer, and (c) dimer involving weaker molecule-molecule interactions. The central molecule is labelled as 'C'. ....	244
Figure 7.16. Top three pair-wise intermolecular interactions in (a) clozapine N,N-dimethylacetamide solvate and (b-c) clozapine monohydrate, quantified and labelled as in Table 7.7. The central molecule is labelled as 'C'. Other hydrogen atoms are omitted for clarity.....	246
Figure 7.17. Stack plot of XRPD patterns of amoxapine raw material, melt quenched amorphous phase and partially recrystallised sample of amorphous amoxapine in the range of 10-35° 2θ. ....	248
Figure 7.18. DSC curves of amoxapine (a) heat-cool-heat cycle of amoxapine. The heat-cool-heat segments are shown by blue, golden and black colour respectively. In heating curve, star denotes an instrument artefact. (b) amorphous amoxapine.....	249

Figure 7.19. a) Layer structure formed by stacking of two types of chains (highlighted in pink and green colour) of amoxapine molecules. (b) The stacking of layers (shown by red and blue colours). All the hydrogen atoms are omitted for clarity. ....	250
Figure 7.20. Summary of preliminary crystal structure search of amoxapine. Each point represents a hypothetical crystal structure/lattice energy minimum. The blue square with open red diamond represents the global minimum in lattice energy, which corresponds to the known experimental structure. ....	251
Figure 7.21. Top five pairwise intermolecular interactions in amoxapine, quantified in Table 7.9, (a-b) molecular pairs involve contribution from C-H...Cl interaction (black line) and C-H...O interaction (orange line) and (c) a molecule-molecule interaction. The central molecule is labelled as 'C'. ....	253
Figure 7.22. Overlay of the best simulated annealing solution with proton manually placed at succinic acid (co-crystal, grey) with the CASTEP geometry optimised structure (black) indicating proton transfer from succinic acid to loxapine. H-atom position before and the one obtained after geometry optimisation is denoted by H1 (SA) and H1 (DFT) respectively. Other hydrogen atoms are omitted for clarity. RMSD non-hydrogen atoms = 0.121Å. ....	258
Figure 7.23. Final fit showing the observed $y_{\text{obs}}$ (lines), calculated $y_{\text{calc}}$ (points) and difference $[(y_{\text{obs}} - y_{\text{calc}})/\sigma(y_{\text{obs}})]$ profiles for the restrained Rietveld refinement of the best simulated annealing solution of loxapine hydrogen succinate. ....	259
Figure 7.24. Scatter plot of solvents with positions according to the value of $t[1]$ and $t[2]$ from Figure 7.2. Solvents are coloured according to the crystallisation outcome: form I (blue), form II (magenta), mixture of forms I and II (green), novel physical forms, solvates and/or salts (red), no sample obtained (golden) and no spectral information (cyan). Solvents which were not used in crystallisation are coloured grey. Plots are shown for crystallisation conditions listed in Table 7.3 (a) solvent evaporation in vials and (b) solvent evaporation on watch glass. ....	261
Figure 7.25. Scatter plot of solvents with positions according to the value of $t[1]$ and $t[2]$ from Figure 7.2. Solvents are coloured according to the crystallisation outcome: form I (blue), form II (magenta), mixture of forms I and II (green), novel forms, solvates and/or salts (red) and no spectral information (cyan). Solvents which were not used in crystallisation are coloured grey. Plot is shown for cooling crystallisations, listed in Table 7.3. ....	262
Figure 7.26. Stack plot of XRPD patterns of physical forms of loxapine identified during experimental screen. ....	263
Figure 7.27. STA thermograms of loxapine form I (red) and form II (blue). ....	264
Figure 7.28. Overlay of the molecular conformations of amoxapine (green) and loxapine (blue) which are present in the experimental crystal structures (RMSD of non-hydrogen atoms = 0.08Å). ....	264

Figure 7.29. (a) 2-D packing of loxapine. (b) The staggered arrangement of sheets. Alternate sheets are shown by red and blue colour respectively. All the hydrogen atoms are omitted for clarity. ....	265
Figure 7.30. (a–b) 1-D chains of loxapine molecules along <i>b</i> -axis. (c) The stacking of chains of loxapine molecules along <i>c</i> -direction. All the hydrogen atoms are omitted for clarity.....	266
Figure 7.31. Summary of preliminary crystal structure search of loxapine. Each point represents a hypothetical crystal structure. The violet triangle with open red square represents one of the lowest energy structures, which corresponds to the known experimental structure. ....	268
Figure 7.32. Top three molecular pairs in form I of loxapine, quantified in Table 7.12, (a) centrosymmetric dimer, (b) molecule-molecule interaction and (c) molecular pair involves C-H...O interaction (shown in orange line). The central molecule is labelled as 'C'. Other hydrogen atoms are omitted for clarity. ....	270
Figure 7.33. Top three pairwise intermolecular interactions in form II of loxapine quantified in Table 7.12. (a) Molecular pair involving C-H...O (shown by orange line) and CH <sub>2</sub> ...π interactions and (b-c) molecular pairs involving contribution from CH...π interactions. The central molecule is labelled as 'C'. Other hydrogen atoms are omitted for clarity.....	271
Figure 8.1. The four compounds studied in this thesis. The differences in the chemical structures within a group are highlighted in red.....	276
Figure 8.2. Schematic workflow developed in this work.....	279

## List of Tables

Table 1.1. Different types of Intermolecular interactions involved in PIXEL calculations, their origin and their dependence on the intermolecular distance, r. ....	30
Table 3.1. Compounds studied in this work.....	46
Table 4.1. Examples of high-throughput crystallisation carried out for physical form screening of pharmaceuticals. Reproduced from reference (Aaltonen et al., 2009).....	58
Table 4.2. Counter ions used for salt screening of amoxapine .....	65
Table 4.3. Solvents used for physical form screening of clozapine in experiment 1 (PS-1) using 48-well plate.....	67
Table 4.4. Solvents used for physical form screening of clozapine in experiment 2 (PS-2) using 48-well plate.....	67
Table 4.5. Solvents used for physical form screening of olanzapine using 96-well plate .....	68
Table 4.6. Raman bands within the region of 1400-1700 cm <sup>-1</sup> of Raman shift .....	72
Table 4.7. Details of 11 groups observed on the 96-well plate and their respective colour shown in Figure 4.6.....	74
Table 4.8. Total number amoxapine salts identified from each counter acid during HTCAA.....	77
Table 4.9. Amoxapine salts classified by Raman spectra using group analysis on the results obtained in the screening experiment on the 96 well plate.....	78
Table 4.10. Lattice parameters of the amoxapine salts identified from the salt screen. All parameters were determined from single-crystal diffraction data collected at 123(2) K (see Section 3.2.2).....	82
Table 4.11. Content of asymmetric unit of six novel salts of amoxapine reported in Table 4.10.....	83
Table 4.12. Solvent systems used in physical form screening, PS-1 of clozapine using a 48-well plate. Wells where no crystals were observed are coloured green. Group analysis of all the Raman spectra is shown in Figure 4.13.....	88
Table 4.13. Solvent systems used in physical form screening, PS-2 of clozapine using a 48-well plate. Wells where no crystals were observed (based on visual inspection) are coloured pink. Group analysis of all the Raman spectra is shown in Figure 4.15.....	90
Table 4.14. Physical form screening results of olanzapine using 96-well plate after group analysis based on Raman spectra. The wells are colour coded according to the classification group. Wells with no fill represent cases where no crystals were observed.....	95
Table 4.15. Potential advantages and disadvantages of HTCAA methodology.....	103

Table 5.1. Confusion matrix generated by Random Forests for classification of dataset of 382 molecules. Class 1 represents the molecules which crystallised and class 2 represents the molecules which did not crystallise.....	133
Table 5.2. Ten most important descriptors selected by Random Forests for classification model. Reproduced from web reference .....	136
Table 6.1. Nomenclature for non-solvated polymorphs of olanzapine and their characteristic data [this work and from reference (Reutzel-Edens et al., 2003)].....	146
Table 6.2. Summary of manual crystallisation conditions used in the experimental search. ....	150
Table 6.3. The angles considered in the CrystalPredictor grid generation. ....	157
Table 6.4. Summary of crystallisation techniques used and summary of solid forms produced in the experimental search for physical forms of olanzapine. ....	164
Table 6.5. Lattice parameters for olanzapine crystal structures found in the experimental search (standard uncertainties in parentheses). All parameters were determined from single-crystal diffraction collected at 123(2) K. ....	167
Table 6.6. Results of spray drying experiments for olanzapine. ....	177
Table 6.7. Highest-dimensionality SCs observed in the 35 olanzapine structures and forms obtained after desolvation. The number in square brackets corresponds to each structure in Figure 6.18. ....	179
Table 6.8. Confusion matrix generated from Random Forest classification for the three molecular packing types of olanzapine solvates. ....	192
Table 6.9. Numerical values of solvent descriptors obtained from variable dependency plots generated during Random Forests classification of olanzapine solvates. ....	198
Table 6.10. Partitioned intermolecular interaction energies (kJ mol <sup>-1</sup> ) from the PIXEL calculations of form I and II structures at 298 K, 123 K and 0 K. ....	204
Table 6.11. Intermolecular interaction energies (kJ mol <sup>-1</sup> ) for the top three ranked pairwise interactions in experimental form I and II at 298 K, 123 K and in predicted form I and II at 0 K. ....	205
Table 6.12. Partitioned intermolecular interaction energies (kJ mol <sup>-1</sup> ) from the PIXEL calculations of hypothetical structures A1 and A30. ....	207
Table 6.13. Intermolecular interaction energies (kJ mol <sup>-1</sup> ) for the top two ranked dimer interactions in computed structures, A1 and A30 at 0 K. ....	207
Table 7.1. Summary of the manual crystallisation techniques used in the experimental search for physical forms of clozapine. ....	223
Table 7.2. Summary of the manual crystallisation techniques used in the experimental search for physical forms of amoxapine. ....	224
Table 7.3. Summary of the manual crystallisation conditions used in the experimental search for physical forms of loxapine. ....	225

Table 7.4. Lattice parameters for clozapine crystal structures obtained in the experimental search (standard uncertainties in parentheses). All parameters were determined from single-crystal diffraction data collected at 123(2) K.....	231
Table 7.5. Partitioned intermolecular interaction energies (kJ mol <sup>-1</sup> ) from the PIXEL calculations on clozapine anhydrous form, monohydrate and DMA solvate. ....	242
Table 7.6. Intermolecular interaction energies (kJ mol <sup>-1</sup> ) for the top three ranked molecular pairs in the anhydrous form of clozapine. These pairs are shown in Figure 7.15.....	243
Table 7.7. Intermolecular interaction energies (kJ mol <sup>-1</sup> ) for top three ranked dimer interactions in clozapine monohydrate and N,N-dimethylacetamide solvate. These pairs are shown in Figure 7.16.....	245
Table 7.8. Partitioned intermolecular interaction energies (kJ mol <sup>-1</sup> ) from the PIXEL crystal lattice calculations of amoxapine.....	252
Table 7.9. Intermolecular interaction energies (kJ mol <sup>-1</sup> ) for the top three ranked molecular pairs in amoxapine. These pairs are shown in Figure 7.21.....	253
Table 7.10. Lattice parameters for loxapine polymorph crystal structures (redeterminations) found in the experimental search (standard uncertainties in parentheses). All parameters were determined from single-crystal diffraction collected at 123(2) K. ....	263
Table 7.11. Partitioned intermolecular interaction energies (kJ mol <sup>-1</sup> ) from the PIXEL calculations of loxapine polymorphs. ....	269
Table 7.12. Intermolecular interaction energies (kJ mol <sup>-1</sup> ) for the top three ranked dimer interactions in loxapine polymorphs. These pairs are shown in Figure 7.32 and Figure 7.33. ....	269
Table 8.1. Recrystallisation methods used in this study to generate various solid forms .....	277
Table 8.2. List of the computational tools used in this study .....	278



## List of Abbreviations

Active Pharmaceutical Ingredient	API
Amoxapine	AXPN
Cambridge Structure Database	CSD
Carbamazepine	CBZ
Chlorothiazide	CT
Clozapine	CZPN
Coulombic Energy	$E_c$
Crystal Structure Prediction	CSP
Design of Experiment	DOE
Dispersion Energy	$E_d$
High Throughput Crystallisation	HTC
High Throughput Crystallisation and Analysis	HTCAA
Hydrochlorothiazide	HCT
Loxapine	LXPN
Multi-well Plate	MWP
Olanzapine	OZPN
Polarisation Energy	$E_p$
Principal Component Analysis	PCA
Quality by Design	QbD
Random Forests	RF
Relative Humidity	RH
Repulsive Energy	$E_r$
Single Crystal X-ray Diffraction	SXD
Structure Determination from Powder data	SDPD
Supramolecular Construct	SC
Total Lattice Energy	$E_{tot}$
Variable Temperature X-ray Powder Diffraction	VT-XRPD
X-ray Powder Diffraction	XRPD

# **1 Introduction**

## 1.1 Background

Pharmaceutical products frequently display multiple solid-state forms of the same chemical entity including polymorphs, salts, solvates, co-crystals and amorphous solids. According to one literature report based on 245 screens carried out by SSCI (a division of Aptuit Inc.), ~80% of the screened compounds exhibited multiple solid-state forms (Stahly, 2007). Sometimes, the properties of two solid forms of a drug do not show substantial differences however they can, and often do, vary substantially with potentially significant impact on important physicochemical properties including solubility that may impact on the ease of manufacturing, stability of the drug in a formulated dosage form as well as bioavailability (Huang and Tong, 2004; Llinàs and Goodman, 2008). A fundamental impact of a change in solid-state structure is on aqueous solubility, which is a prerequisite for the efficient absorption of drug (Tian et al., 2007; Koradia et al., 2011) from the gastrointestinal tract (GIT). Thus, establishing biological activity with minimal side effects is not enough to enable molecules to progress to the clinical development stage: biopharmaceutical properties are also vitally important and the solid-state form of the drug is a considerable part of this and must be characterised and controlled (Aaltonen et al., 2009).

The crystallisation behaviour of active pharmaceutical ingredients (APIs) needs to be fully explored and understood to enable control over purity, crystal form, particle size, particle size distribution, degree of crystallinity and crystal habit. All of these properties can affect bulk properties such as compaction, tableting,

flowability and mixing (Pfeiffer et al., 1970) and ultimately product performance.

For commercial pharmaceutical products, the most thermodynamically stable crystalline form under ambient conditions is typically the most desirable one (Miller et al., 2005). There are exceptions to this practice, e.g. some marketed formulations of norfloxacin contain a metastable form (R. Barbas et al., 2007). Whilst metastable polymorphs can offer advantages such as enhanced solubility, they also can revert to more stable less soluble forms. A well-documented example of catastrophic transition from a metastable to stable polymorphic form is that of Ritonavir (Norvir™, Abbot Laboratories). The formulation (Norvir™) consisted of the only known polymorph of ritonavir at that time and was marketed and distributed for 18 months. Later on, the appearance of a more thermodynamically stable and less soluble form during manufacturing resulted in reduced bioavailability of the final product. The product had to be reformulated with the new polymorph to achieve the desired bioavailability (Bauer et al., 2001). Had information on the various forms of ritonavir and their properties been known prior to marketing of the product, the market recall and subsequent expense to manufacturer and risk to patients could have been avoided. Therefore, in order to assure that the correct form with known defined properties is used, there is a need to explore all potential solid-state forms of a given molecule (Llinàs and Goodman, 2008). However, even after selecting the appropriate solid-state form, phase transformations (polymorphic changes or desolvation or amorphisation) that occur during processing or storage cannot be completely ruled out. Hence, it is necessary to identify and characterise the

phase changes to avoid the risk of transformation of active solid form to less/non-active solid form (Koradia et al., 2011).

In addition to physical changes, alternative solid-state forms can display different chemical stability (Byrn, 1982) with different rates of solid-state chemical reactions and consequent variations in the relative proportion of degradation product(s). For example, NCEX, a dual peroxisome proliferator-activated receptors (PPAR) agonist exists in two monotropically related polymorphs, form I and II with melting points of 160 °C and 140 °C, respectively. Form I and II showed 0.3 and 6.0 % total degradation respectively after 4 weeks of storage at 40 °C and 75% RH (accelerated stability storage conditions). The amorphous form showed much higher degradation (20%) under similar conditions. (Huang and Tong, 2004). Due to different microenvironmental pH and molecular arrangements in the crystal lattice, photochemical stability of the salt formulations may vary (Huang and Tong, 2004). For example, the hydrochloride salt of GG818 showed higher photostability as compared to the free base (Tong et al., 1999).

Polymorphs and solvates/hydrates are encountered frequently (Kitamura, 2005) during drug development and there is still a considerable challenge in control and prediction of the appearance of all possible solid-state forms. The various solid forms that may be encountered for a given organic solid include polymorphs, solvates, salts, co-crystals and non-crystalline or amorphous forms (Figure 1.1).

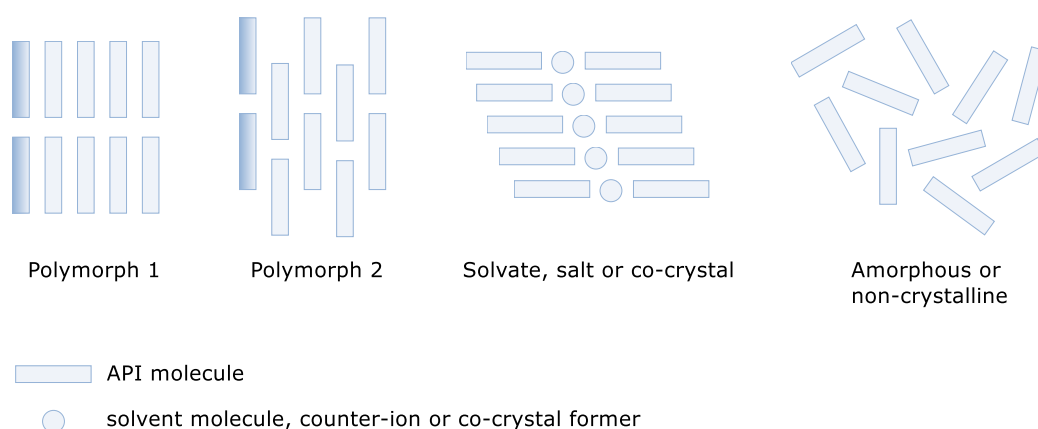


Figure 1.1. Schematic illustration of molecular arrangements of different solid forms showing ordered arrangement in crystalline forms and molecule disorder in amorphous form. Reproduced from reference (Florence, 2009a).

## 1.2 Polymorphism

Polymorphs are chemically identical crystalline forms of a chemical entity in which the constituent molecules have different packing arrangements and/or conformations in the crystal lattices (Vippagunta et al., 2001). Conformational flexibility and hydrogen bonding are the most important factors responsible for polymorphism (Yu et al., 2000). However, predicting the polymorphic behaviour or number of polymorphs of a compound is still a challenging task (Grzesiak et al., 2003; Price, 2008a; Lutker and Matzger, 2010; Bardwell et al., 2011).

Packing polymorphism arises when constituent molecules with specific rigid conformation are packed in different arrangements which may or may not have same intermolecular binding motifs. Packing polymorphism has been reported in five polymorphs of carbamazepine (CBZ) (Figure 1.2) in which forms I-IV are

based on amide-amide hydrogen bonded dimers (Grzesiak et al., 2003), whereas form V is based on N-H...O=C catemeric motif (Arlin et al., 2011).

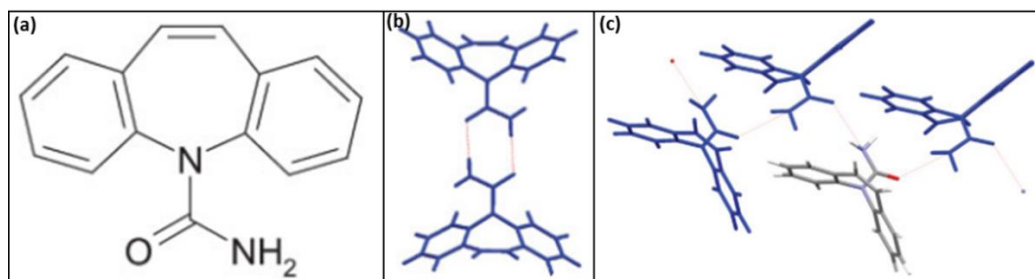


Figure 1.2. (a) Molecular structure of carbamazepine, (b) dimer motif of carbamazepine present in form I-IV, and (c) catameric motif of carbamazepine present in form V. Adapted from references (Florence et al., 2006; Arlin et al., 2011).

Conformational polymorphism is observed when constituent flexible molecules with different molecular conformations are packed in different arrangements (Nangia, 2008). L-glutamic acid for example exhibits conformational polymorphism, in which the  $\beta$ -polymorph is based on an extended conformation and  $\alpha$ -polymorph is based on a twisted conformation (Figure 1.3) (Davey et al., 1997).

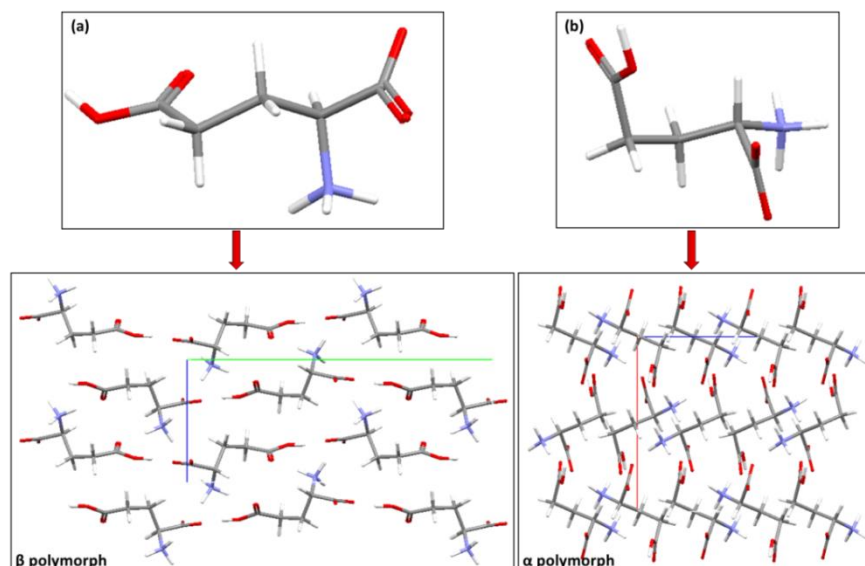


Figure 1.3. (a) Extended conformation of L-glutamic acid which leads to the  $\beta$ -polymorph (CSD refcode: LGLUAC11), and (b) twisted conformation of L-glutamic acid which leads to  $\alpha$ -polymorph (CSD refcode: LGLUAC03).

### 1.3 Multi-Component Crystalline Systems: Solvates, Salts and Co-crystals

Multi-component systems (solvates, salts and co-crystals) have more than one type of molecule in the crystal lattice (Figure 1.4) (Almarsson and Zaworotko, 2004). These multi-component systems can also display polymorphism, adding to the complexity of the experimental crystal energy landscape.

Solvates are crystalline solids in which the stoichiometry of the constituent chemical species is  $A_xB_y$ , where A and B are the higher and lower molecular weight components, respectively (Angelo, 2007). This definition coincides with the traditional definition of a solvate, where B is liquid at the room temperature (RT). X is a small integer and Y is any integer number greater than 0 (Angelo,



2007). Generally, solvent molecules from the solvent of crystallisation are incorporated in the crystal lattice with the organic solute or API. When the solvent is water the structure is known as a hydrate. Hydrates have particular pharmaceutical relevance and roughly 75% of the pharmaceutical compounds form hydrates (Infantes et al., 2003). Some examples of the marketed pharmaceutical hydrates are ciprofloxacin hexahydrate, morphine hydrochloride hydrate, naloxone hydrochloride dihydrate and rosiglitazone maleate hydrate.

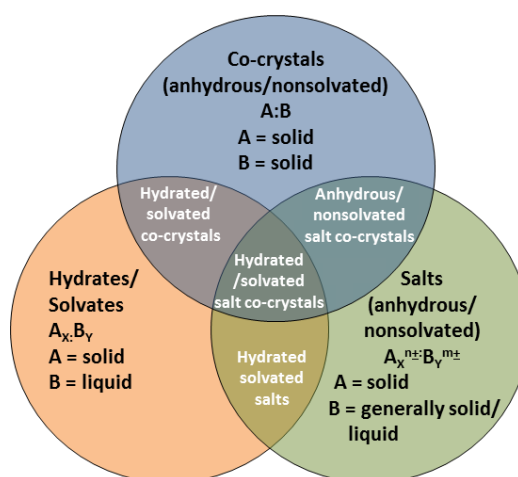


Figure 1.4. Multi-component systems and their overlap with each other. Reproduced from reference (Aitipamula et al., 2012).

Salts are  $A_x^{n\pm}B_y^{m\pm}$  solids, where A is the higher molecular weight component with an integer charge and B is a counterion (Angelo, 2007) (Figure 1.4). There is a proton transfer from acid to base resulting in neutralisation of charge. Approximately two-third of the APIs are ionisable (weakly acidic or basic entities) and therefore, salt formation provides a significant opportunity to

improve the physicochemical properties without altering the molecular structure and the mode of action of the API (Gould, 1986).

The introduction of a charge on the molecule enhances aqueous solubility, however other properties such as melting point as well as the polymorphic profile of the molecule may also change. Each new salt is a new chemical entity and therefore polymorph screening must be carried out on each salt form of interest. For acidic molecules, sodium, potassium and calcium are the most commonly used counter ions while for basic molecules chloride is the most prevalent counter ion (Bernstein, 2002). Also sodium salts can often form hydrates as a result of the sodium ion seeking to satisfy its coordination shell creating more complexity in the crystal structure landscape (Marcus, 1991). Some examples of drugs which are marketed as salts are naltrexone, diclofenac, flurbiprofen, cimetidine, and phenytoin (Bernstein, 2002).

Co-crystals are comprised of two or more species, similar to salts, except that the molecules are non-ionised. Both co-crystals and solvates are comprised of two or more neutral species; however co-crystals formation has been subjected increasingly to efforts to achieve an element of design (Bernstein, 2005), i.e. intended structural or physical properties as a result of rational design or crystal engineering approaches. Co-crystals can be used to alter the physicochemical properties like aqueous solubility or melting point and stability as compared to parent compound (Trask et al., 2006; Aakeröy et al., 2009).

There is some debate on the formal definition of a co-crystal in the scientific literature. Desiraju has referred to these systems as molecular complexes

(Desiraju, 2003), however solvates and hydrates can also be considered as co-crystals as per this definition. Later, according to Aakeroy and Salmon, all co-crystals components should be solid at RT (Aakeroy and Salmon, 2005). Although this definition excludes systems where one component is a liquid or gas by intent (Bond, 2007) and Stahly debated that the physicochemical properties of the co-crystal components should not be important (Stahly, 2007).

In December 2011, FDA proposed a definition of co-crystal in a draft guidance as “solids that are crystalline materials composed of two or more molecules in the same crystal lattice”. The guidance recommends that co-crystals should be classified within the Agency’s current regulatory framework as dissociable “API-excipient” molecular complexes (with the neutral guest compound being the excipient). Molecular association of API and the excipient(s) occurs within the crystal lattice and co-crystal may be treated as a drug product intermediate rather than as a drug (USFDA, 2011)”. This proposed definition has prompted further debate and a universal definition remains to be established (Aitipamula et al., 2012).

## **1.4 Amorphous Solids**

An amorphous material is a non-crystalline solid which does not possess long-range internal molecular order. In their diffraction patterns, they give only a diffuse scattering halo with no Bragg diffraction peaks. Crystallisation of amorphous form provides an alternate route of obtaining different polymorphs. For example different polymorphs of indomethacin were obtained by

crystallisation of amorphous solid at various temperatures. This study also showed that the polymorphic outcome was affected by the small variations in sample handling of generated amorphous form (Bhugra et al., 2008). Various polymer surfaces can also be used as a template to direct the crystallisation of amorphous form. For example,  $\alpha$ -polymorph of indomethacin was selectively obtained from recrystallisation of super cooled melt of indomethacin and polymer substrate (Grzesiak et al., 2006; McKellar et al., 2012).

## **1.5 Importance of Solid-State Form for Pharmaceutical Industry**

Physical form screening has become a standard procedure in the pharmaceutical industry. The phenomenon of polymorphism has been known since the early 19<sup>th</sup> century but study of polymorphism was not a standard procedure until the last decade of the 20<sup>th</sup> century. The production problems such as sudden appearance of thermodynamically more stable polymorph of Ritonavir (Norvir<sup>TM</sup>, Abbot Laboratories) with low solubility and bioavailability (Chemburkar et al., 2000), patent cases (Bernstein, 2002) and enormously expensive Hatch-Waxman agreement between innovator and generic pharmaceutical companies (Bulow, 2004) acted as a motivation for the pharmaceutical industry to take polymorphism seriously (Aaltonen et al., 2009). The famous patent litigation case between Glaxo and Genpharm/Novopharm for Ranitidine hydrochloride (Zantac<sup>®</sup>) brought to attention the concept of role of solvent, heating, stirring and seeding techniques to control polymorphic form

and the relative stability of polymorphic forms. This litigation case also highlighted the importance of development and use of analytical methods for full characterisation of polymorphic forms (Bernstein, June 19, 2004). Regulatory authorities now ask for full characterisation of the solid-state forms during drug approval. Therefore, to avoid the risks associated with latent polymorphism, phase transformation during processing or storage and expensive patent litigations, comprehensive understanding and control of the solid-state behaviour of APIs has become an integral part of the drug development for innovator pharmaceutical companies.

## **1.6 Crystallisation**

Crystallisation is the formation of solid crystals consisting of 3-D periodic arrangement of atoms and molecules, which are held together by a wide variety of intermolecular interactions. Some interactions such as the hydrogen bonds (H-bonds) act over a short range and are directional in nature. Whilst other interactions act over longer ranges and are non-directional. All these intermolecular interactions are responsible for molecular recognition, self-assembly, cohesion of molecules in a periodic arrangement and the solid-state properties. Although crystals can be obtained from melt or gas states, crystallisation from solution is the most common and is often the method of choice for purification for industrial processes (Mullin, 1971; Ejgenberg and Mastai, 2012).

Crystallisation from solution consists of three stages: supersaturation, nucleation, and growth of the nuclei into crystals. Supersaturation is described as when a solution contains higher concentration of solute than an equilibrium concentration of the solute (Laudise, 1970). Supersaturation can be achieved by various methods including different rate of solvent evaporation, antisolvent addition, different rate of cooling, pH change, varying concentration or temperature change (Rodríguez-hornedo and Murphy, 1999). All these methods are practised with the purpose of manoeuvring solute concentration.

The next stage after achieving supersaturation is nucleation which can be defined as the assembly of ordered molecules into viable nuclei. Critical number of ordered molecules per assembly is required to prevent dissolution and produce a feasible nucleus, which acts as a template for crystal growth. With time the nucleus grows and more solute molecules join the assembly thereby increasing the adhesive solute-solute (internal) interactions which outweigh the disruptive solute-solvent (surface) interactions, leading to cessation of re-dissolution and promotion of crystal growth.

Nucleation events decide the number of particles, particle size and physical form resulting from crystallisation (Davey et al., 2002). Despite the great efforts in the field of fundamental mechanisms that drive the crystal form diversity, the nucleation phenomena are not well understood (Davey, 2004). This in part is due to the fact that there are no reliable methods that can detect nucleation experimentally at present (Hursthouse et al., 2009). New approaches for

detection and analysis of molecular clusters at low concentrations in solution are desirable to have a better understanding of nucleation (Davey et al., 2013).

In polymorphic systems, the resulting outcome of crystallisation is the consequence of the kinetic relationship between nucleation and crystal growth. The polymorph which nucleates first during crystallisation is thought to come from the assembly with faster nucleation rate as a result of its lowest free energy barrier to nucleation. However, the final crystallisation outcome, i.e. polymorph, would be determined by the combination of relative nucleation and crystal growth rates of the polymorphs (Khoshkhoo and Anwar, 1993; Bernstein et al., 1999).

The nuclei of different polymorphs can form and coexist in solution and a mixture of polymorphs can be obtained when kinetic factors prevent accomplishment of equilibrium. Therefore, as a general rule, under thermodynamic conditions (e.g. slow cooling), stable polymorphic form will be obtained and under kinetic conditions (e.g. fast cooling) a metastable form is anticipated. Therefore, it is necessary to investigate the interplay between thermodynamic and kinetics factors to discover all the relevant physical forms of a compound (Aaltonen et al., 2009). Schematic illustration of plausible scenarios of crystallisation outcomes in case of a dimorphic system is shown in Figure 1.5.

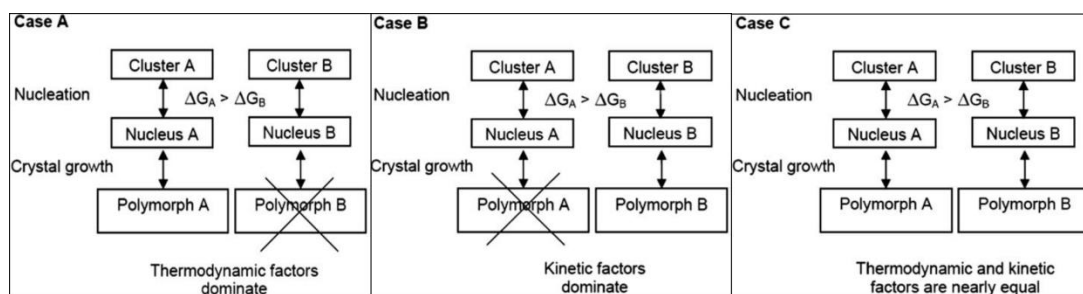


Figure 1.5. Plausible scenarios of crystallisation outcome in a dimorphic system under different circumstances. Polymorph A has higher free energy barrier ( $\Delta G$ ) to nucleation than polymorph B and is the stable form. Adapted from reference (Aaltonen et al., 2009).

## 1.7 Crystallisation Techniques

The statement made by the late Walter McCrone in 1965 that, “The number of forms of a given molecule is proportional to the time, money and experiments spent on that compound” (McCrone, 1965) has gained credence in recent years, as illustrated by the significant increase in reported physical form diversity of pharmaceutical solids (Morissette et al., 2004).

The aim of comprehensive physical form screening is to achieve the maximum coverage of experimental crystallisation space possible within the constraints of available time, material and resources. To maximise the probability of encountering all relevant forms it is highly desirable to utilise multiple crystallisation approaches and therefore sample the nucleation/crystallisation outcomes from a broad range of conditions (Morissette et al., 2004; Aaltonen et al., 2009).

In general, crystallisation techniques based on standard solution crystallisation (cooling, evaporation, anti-solvent and slurry) and solid phase methods



(sublimation, mechanical grinding, desolvation and recrystallisation from melt) are the methods of choice for physical form screening (Guillory, 1999; Morissette et al., 2004). Multiple crystallisation techniques used in this work for solid form screening are described in Section 3.2.1. Timescale of various crystallisation experiments is shown in Figure 1.6. This figure also highlights the qualitative relationship between the time required to generate a crystal form and the stability of that form, i.e. tendency to obtain metastable crystal forms is higher from faster crystallisation processes than those that are slow (Llinàs and Goodman, 2008; Bernstein, 2011).

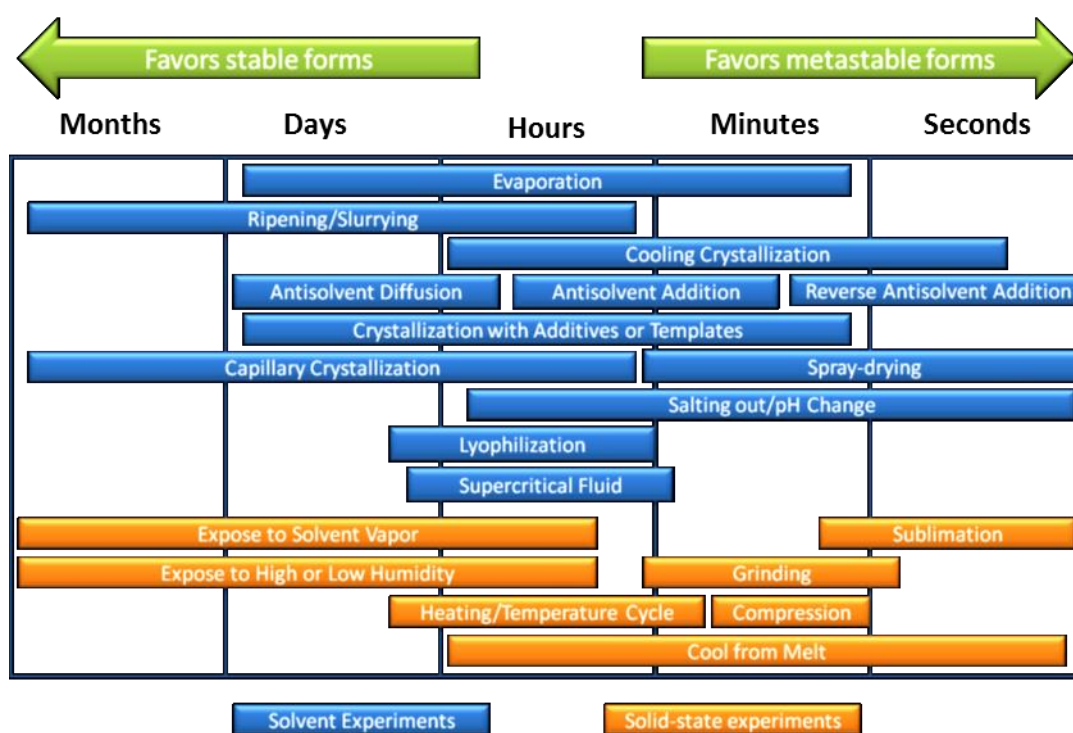


Figure 1.6. Timescale of various crystallisation techniques, which is also highlighting the techniques which can be used to obtain stable and metastable crystal forms. Adapted from reference (Anderton, 2007; Llinàs and Goodman, 2008).

Crystallisation from solution is the most widely adopted method for physical form screening for various reasons. Firstly, various polymorphs and solvates can be obtained just by changing the solvent system. Secondly, the information about propensity of solvate/hydrate formation of an API is required to develop an effective manufacturing process as API is exposed to various solvents during manufacturing and secondary processing (Aaltonen et al., 2009). Finally, in case of solvates/hydrates obtained from solution crystallisation, desolvation provides an alternative route to discover polymorphs (Nicolai et al., 2007) and was used to obtain polymorphic forms of one of the pharmaceutical molecules studied here (see Section 6.4.4).

Solution crystallisation also remains the most widely employed method for high-throughput crystallisation (HTC). HTC is a more efficient approach than the traditional manual approach and provides clues on the extent of solid-state diversity of the compound under study in a shorter period of time by utilising minimal resources. Using the HTC technique for physical form screening of a compound MK-996, over a 120 fold increase in efficiency in terms of number of experiments and time required to carry out those experiments over manual crystallisation approaches method has been achieved (Jahansouz et al., 1999; Almarsson et al., 2003). HTC is discussed in more detail in Chapter 4.

## **1.8 Control of Solid Form**

For the pharmaceutical industry, control of solid form in crystallisation process is of paramount importance due to the probabilities of striking differences in

physicochemical properties of different polymorphs and to develop robust manufacturing process (Reutzel-Edens, 2006). Control of the solid form is also critical because of the detailed scrutiny of the functional properties of the solid forms by the regulatory agencies. The appearance and disappearance of polymorphs are still largely empirical as there are no reliable methods to assess the potential extent of polymorphism in a given molecule, particularly for the typical pharmaceutical compounds with continuously increasing structural complexity (Chen et al., 2005; Lu et al., 2007). Some authors says that “perhaps there is a little of the black arts involved as findings like latent polymorphism, universal seeding and disappearing polymorphs cannot be fully explained by science of polymorphism” (Findlay and Bugay, 1998).

The interest and understanding in the area of polymorph control is growing rapidly. Over the last two decades, there has been a significant progress in the development of various methods for controlling polymorphism and a plethora of new polymorphs have been characterised which is partly due to the advancements of the methods used for characterisation of solid forms (Llinàs and Goodman, 2008). The desired crystal form can be obtained by an understanding of the structural relationship between polymorphs, the interplay between thermodynamic and kinetic factors, and by tuning the fundamental crystallisation block (Blagden and Davey, 2003; Rodríguez-Spong et al., 2004).

The knowledge of the key controlling factors is essential for a selective solid form outcome in the crystallisation process. Factors controlling crystallisation outcome can be broadly divided into two categories; primary and secondary

factors (Figure 1.7) (Kitamura, 2003). Supersaturation and temperature are the two most vital primary parameters affecting solid form outcome in crystallisation process. Three polymorphs of carbamazepine were successfully obtained from one solvent by targeting various degrees of supersaturation and multiple crystallization temperatures (Getsoian et al., 2008).

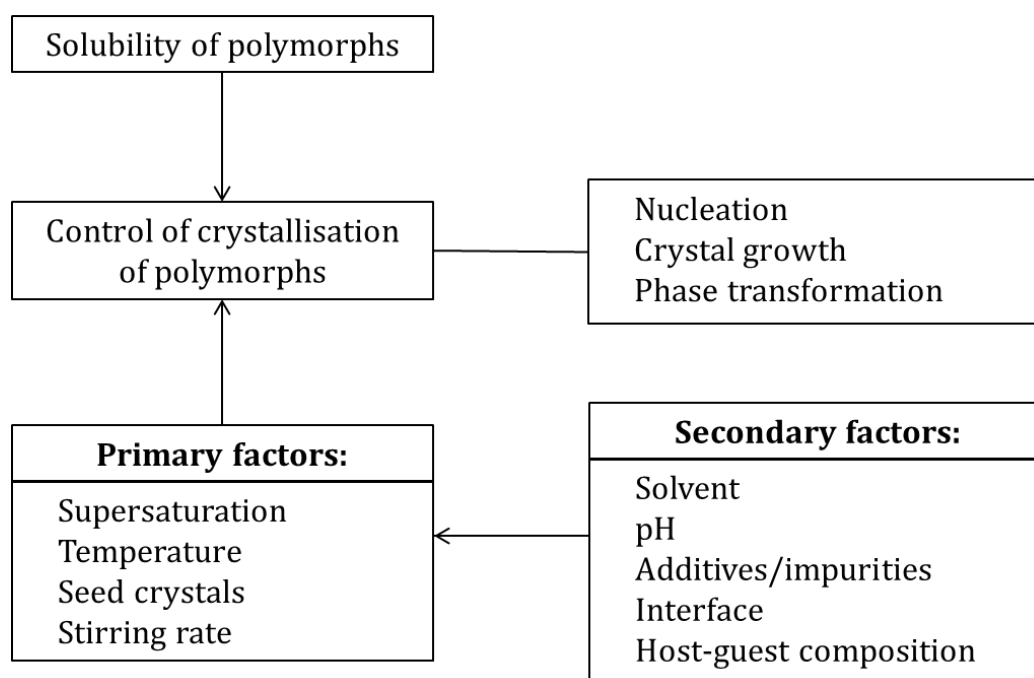


Figure 1.7. Schematic illustration of factors controlling the polymorphic outcome in crystallisation processes. Reproduced from reference (Kitamura, 2003).

Supersaturation is the driving force for crystallisation and influences the kinetics of nucleation and crystal growth. The crystallisation outcome depends on the degree of supersaturation. According to the Ostwald's rule, at higher supersaturation, the metastable form may preferentially precipitate and at low supersaturation, the stable form may tend to precipitate. Various innovative methods have been developed to control polymorphic form by changing the

degree of supersaturation. Some of them are (a) crystallisation using microporous membrane (Di Profio et al., 2007), (b) contact line crystallisation (Capes and Cameron, 2006), (c) capillary crystallisation (Capes and Cameron, 2006), (d) crystal nucleation in nanoscopic confinement (Ha et al., 2004; Hamilton et al., 2011), and (e) pH control using potentiometric cycling (Llinas et al., 2007). Majority of these methods are more suitable on a small scale, although these can generate seeds of the desirable form which can be used for further scale-up.

Temperature is the second most important factor in controlling the crystallisation outcome. Practically, crystallisation at different temperatures may produce different polymorphs (Kitamura and Nakamura, 2002). For example, during crystallisation of L-glutamic acid at 25°C the metastable  $\alpha$ -polymorph nucleated and rapidly caused the growth of the  $\alpha$ -polymorph. When the crystallisation was carried out at 45°C, though the  $\alpha$ -polymorph still nucleated, it exhibited slow transformation and resulted in the growth of the more stable  $\beta$ -polymorph (Kitamura, 1989). In the case of solvates, different stoichiometry of solvates can be obtained by changing the harvesting temperature and in particular low temperatures favour the formation of solvates with higher stoichiometry (Byrn et al., 1999). Therefore, during physical form screening, crystallisation at various temperatures is usually recommended to explore all solid forms and relationship amongst them.

Secondary factors such as pH, interface and host-guest composition affect the crystallisation outcome primarily through their effect on the degree of

supersaturation. Seeding, where solid particles of desired form are added into the crystallisation medium, is the most common and well established method used by pharmaceutical industry to control solid form and particle size distribution (Beckmann, 2000). Seeding strategy relies on the heterogeneous nucleation (promoted by crystal surfaces) and crystal growth in a controlled manner (Beckmann, 2000), while avoiding heterogeneous nucleation mediated by unknown impurities (Mullin, 2001). In general, seeding is very effective; however, in cases where the one solid form nucleates another solid form, then seeding becomes ineffective and will not produce the aimed outcome. For example, seeding of a system with metastable  $\alpha$  form of L-glutamic acid induces the nucleation of more stable  $\beta$  form as surface of the  $\alpha$  form acts as a template for its nucleation (Ferrari and Davey, 2004).

Solid form control can also be achieved by heterogeneous nucleation mediated by employing additives (Davey et al., 1997), templates (Stoica et al., 2005), self-assembled monolayers (Hiremath et al., 2004; Yang et al., 2012) and polymers (Price et al., 2005). These approaches can sometimes lead to discovery of unknown more stable polymorph(s) as seen with form IV of CBZ obtained by polymer heteronucleation approach (Price et al., 2005) and can reveal the extent of polymorphism as in 8 polymorphs of flufenamic acid (López-Mejías et al., 2012).

Crystallisation processes under thermodynamic control are least affected by the nature of the solvent (Threlfall, 2000); however it can have great effect on the polymorph appearance (Reutzel-Edens, 2006). In some cases, solvent or solvent

mixture can be used to promote or inhibit certain polymorphs. The mechanism of getting desired polymorphic forms by using different solvent is similar to the additives except that the solvent is in excess. Solvents can play a key role in directing nucleation at molecular level and both nucleation and crystal growth can be manoeuvred at the molecular level by exploring the solvent solute interactions for controlling competing crystallisation pathways (Davey et al., 2002; Blagden and Davey, 2003). These studies suggest a link between the early stages of molecular recognition and aggregation and final crystal structure.

The polymorphic outcome can also be influenced by the experimental control of crystallisation conditions e.g. by laser induced nucleation (Sun et al., 2008), ultra-sound assisted crystallisation (Gracin et al., 2005), micronisation of antisolvent (Murnane et al., 2008) or crystallisation under high pressure (Fabbiani et al., 2005).

In spite of the progress made in the methods for controlling polymorphism, the fundamental mechanisms involved in crystallisation and the molecular properties that lead to physical form diversity are poorly understood. As a result, the appearance and disappearance of polymorphs is considered only partially explicable and prediction of the extent of potential polymorphism for a given compound still remains a formidable challenge.

## 1.9 Computational Methods

### 1.9.1 Crystal Structure Prediction

Crystal structure prediction (CSP) refers to the prediction of thermodynamically feasible crystal structures of a molecule using molecular structure alone by applying *ab-initio* methods. Computational solid-state chemistry has progressed significantly in the last decade where the development was partly inspired by the promise of useful applications in the field of industrial development of pharmaceuticals (Price, 2004; Verwer and Leusen, 2007; Braga et al., 2010; Day, 2011). However, the discovery and analysis of solid forms is still dependant on systematic experimental screening to explore and understand the complete picture of potential solid form diversity of today's structurally complex pharmaceutical molecules.

Due to potential applications of CSP, a range of free or commercial software packages have been developed and are now available to explore the different ways in which organic molecules may pack in the solid-state. Work is on-going in further improvements of these programs. Most commonly used software packages include, GRACE and VASP (Neumann and Perrin, 2005; Neumann et al., 2008; Neumann, 2011), Crystal Predictor, DMACRYS, Crystal Optimiser (Karamertzanis and Pantelides, 2007; Price et al., 2010; Kazantsev et al., 2011b), Polymorph Predictor (Materials Studio; Accelrys Inc.) and CRYSTALG (Pillardry et al., 2001).



CSP is based on the assumption that molecule will crystallise in the most thermodynamic stable structure, although other crystal structures that are close in energy may also be obtained as polymorphs (Price, 2009). It also assumes that the crystallisation is controlled by thermodynamic factors and all the temperature effects can be ignored. Thus, the method looks for static 0 K crystal structure (global minimum in lattice energy) with the most energetically favourable packing (Price, 2004). CSP results are a set of the most stable computer generated structures/predicted lattice energy minima. These results are represented in a scatterplot of lattice energy vs. density or lattice energy vs. packing coefficient. These scatterplots are called crystal energy landscapes. The crystal energy landscapes can be broadly divided into four types. Schematic diagrams of the crystal energy landscapes are shown in Figure 1.8 and a brief summary is described in the following sections.

#### ***1.9.1.1 Monomorphic Crystal Energy Landscape***

Monomorphic crystal energy landscape is the one where one structure is thermodynamically most stable and hence the most preferred structure in the lattice energy (Figure 1.8a). The energy difference between this and any other structure on the landscape is higher than the plausible energy difference between polymorphs. As suggested by the current literature, monomorphic crystal energy landscapes are rare for organic molecules. Pigment yellow 74 is a notable example where experimental crystal structure is 12 kJ mol<sup>-1</sup> more stable than the next lowest energy structure on the crystal energy landscape and experimental screening also has not found any other structure.

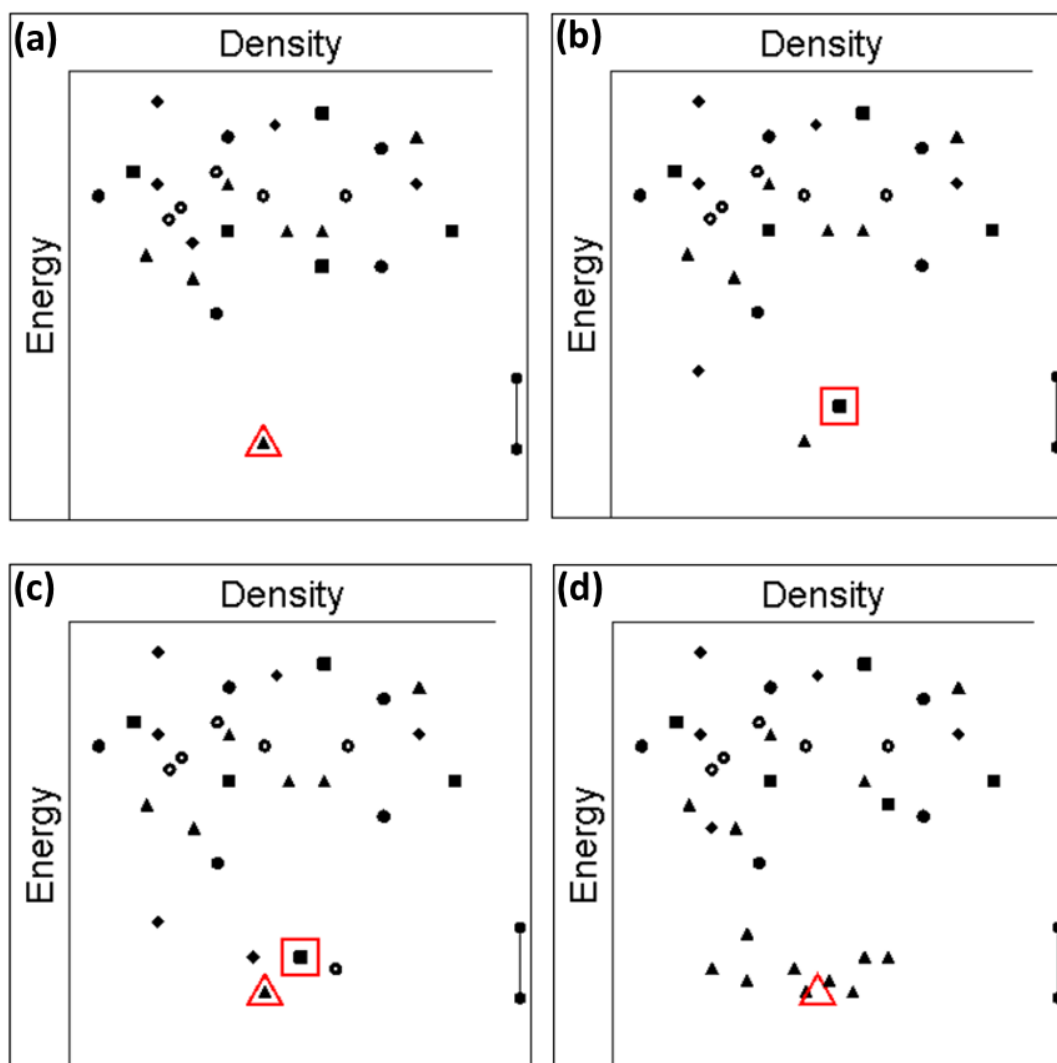


Figure 1.8. Schematic examples of the crystal energy landscapes; a, b, c and d denote monomorphic, predictive, complex and interchangeable crystal energy landscapes respectively. Each point denotes a crystal structure that is a local energy minimum, with the symbols representing significantly different type of packing such as different hydrogen bonding motifs. The experimentally known structures are denoted by empty red symbols. Reproduced from reference (Price, 2009).

### 1.9.1.2 Predictive Crystal Energy Landscape

A predictive crystal energy landscape is the one where the known most stable polymorph is not at the global minimum and an unknown structure is predicted to be more stable than this known polymorph (Figure 1.8b). This kind of result

provides strong motivation for design of crystallisation methods to find the predicted structure or rule out the possibility of any route of its nucleation. Crystallisation experiments were successfully designed to obtain predicted catemer based polymorph of CBZ which was predicted to be energetically competitive with the known  $R_2^2(8)$  dimer motifs containing polymorphs (Arlin et al., 2011).

### ***1.9.1.3 Complex Crystal Energy Landscape***

A complex crystal energy landscape (Figure 1.8c) is obtained when there are various distinct crystal structures within the low energy range and there are various competitive ways of packing of the molecule. In these cases, the structures that can be actually obtained experimentally would depend upon the kinetic factors that influence the nucleation and growth of structure provided that these will not transform into another more stable structure.

This type of landscape indicates that there is a probability of multiple solid forms and there is a packing problem with the molecule, which may result in polymorphism or may be solved by formation of multi-component system e.g. hydrate, solvate or co-crystal. Hydrochlorthiazide (HCT) exhibits complex crystal energy landscape, where a wide range of bimolecular hydrogen bonding motifs is present in the low energy region. The problem of packing these bimolecular hydrogen bonding motifs results in various distinct motifs and many of them were found in two polymorphs and seven solvates (Johnston et al., 2007).

#### ***1.9.1.4 Interchangeable Crystal Energy Landscape and Disorder***

This is a particular type of complex crystal energy landscape (Figure 1.8d), where the low energy crystal structures are associated in an interchangeable manner. This type of landscape arises when there are multiple ways of stacking the same motif (e.g. sheets/ribbons) within effectively the same energy range. This implies that multiple structures based on various arrangements of the motif, would be very close in energy. Therefore, depending on the energy barrier between different arrangements and to correct growth mistakes, this type of energy landscape may lead to disorder, polytypism, multiple stacking faults or incommensurate structures. The example of this type of crystal energy landscape is aspirin, in which the two alternate predicted equi-energetic modes of stacking of the same sheet justify the later finding of form II (Vishweshwar et al., 2005; Price, 2009) and the presence of the intergrowth of polymorphic domains within the same single crystal (Bond et al., 2007).

Contrasting the crystal energy landscape of a molecule with the experimental screening results on a thermodynamic side can be the first step towards better understanding of the factors which control crystallisation and can result in polymorphism (Davey et al., 2002; Price, 2004; Verwer and Leusen, 2007; Price, 2008b; Braga et al., 2010; Day, 2011). The calculation of relative energies of polymorphs provides a major challenge to computational chemistry because energy differences between polymorphs are quite small compared to covalent bond energies. Therefore, calculating the crystal energy landscape with a worthwhile relative accuracy in thermodynamic stability of the crystal

structures is restricted to the small pharmaceutical molecules with limited flexibility. However even after accurate calculations of the relative energies of the various polymorphs, the crystal energy landscape does not reflect the kinetic factors such as nucleation dynamics, solvent effects, temperature, pressure etc. (Price, 2008a) and inclusion of kinetic factors in the CSP remains elusive.

Despite its limitations, many studies show that CSP is a valuable complement to the physical form screening and helps in discovering new structures as well as rationalising the solid forms that are found in experimental searches (Johnston et al., 2007; Arlin et al., 2011; Price, 2013). Nowadays, one sees efforts towards simultaneous experimental and computational search for all possible polymorphs for a given organic compound (Florence et al., 2006; Braun et al., 2012).

The encouraging CSP results of 5<sup>th</sup> blind test on the flexible large molecule with 8 torsion angles and on multi-component systems i.e. salt and hydrate suggest that CSP can migrate to more complex systems from small molecules with limited flexibility, while highlighting the shortcomings where key improvements are still required (Bardwell et al., 2011). Still, it is desirable to push the limits of theoretical predictions for better understanding of the solid-state behaviour of the structurally complex pharmaceutical compounds.

### **1.9.2 Energetics of Non-bonded Interactions: The PIXEL/Semi-classical Density Sums Methods**

An equilibrium crystal structure is formed when the attractive forces and the repulsive forces are balanced within a specific packing arrangement (Callear, 2008). The total intermolecular interaction energy of a crystal and relative contribution of different forces towards the total intermolecular interaction energy can be calculated using semi-empirical PIXEL or semi-classical density sums (SCDS) method, which enables comparison of the governing forces in related solid-state structures.

PIXEL calculations also provide subdivision of lattice energy into molecule-molecule energies and information about the interaction(s)/structural determinants which dictate packing arrangement in the crystal lattice. All this information can help in a better understanding of the relationship between molecular constituents and intermolecular interactions, which is the key to the prediction and control of solid forms (Gavezzotti, 2003b).

The PIXEL method provides a more realistic and comprehensive way of examining intermolecular interactions. It takes account of the entire electron density of a molecule rather than just a few nuclei or point charges and thus provides more rational and comprehensive way of describing interactions. Some significant contributions can be missed due to simplicity associated with the classical atom-atom method. For example, description of an OH...OH interaction using a simple attractive columbic term i.e.  $H(\delta^+) \dots O(\delta^-)$  would miss

on repulsive  $H(\delta^+)...H(\delta^+)$  and  $O(\delta^-)...O(\delta^-)$  terms, which also contribute to the interaction (Dunitz and Gavezzotti, 2005a).

Four different types of forces calculated by PIXEL methods are summarised in Table 1.1 and the sum of these four terms provides the total intermolecular interaction energy. The stabilising forces are comprised of Coulombic, dispersion (London-type) and linear polarisation terms. Repulsion energy is estimated from the overlap of undeformed electron densities. As mentioned earlier, PIXEL calculations provide a list of molecule-molecule interactions (molecular pairs) in the crystal lattice, which are pairs of a reference molecules related by a symmetry operator, distance between their centre of mass and an intermolecular energy. This allows identification of the most stabilising molecule-molecule interactions.

Table 1.1. Different types of Intermolecular interactions involved in PIXEL calculations, their origin and their dependence on the intermolecular distance,  $r$ .

<b>Force</b>	<b>Origin</b>	<b>Varies as</b>
Coulombic	Charge-charge interaction	$1/r$
Polarisation	Permanent dipole-dipole interaction	$1/r^3$
Dispersion or London	Inductive dipole-dipole/ fluctuating multipoles	$1/r^6$
Repulsion (or exchange)	Repulsion between spins	$1/r^{12}$ Overlap of electron cloud

First step during calculations of intermolecular energies using PIXEL method is to calculate the electron density of the individual molecule(s) using standard

quantum mechanical programs, e.g. Gaussian<sup>®</sup>. Typical grid step size of 0.08-0.10 Å is used to calculate the electron density within a grid. The resultant grid containing electron density produced by Gaussian usually contains about  $10^6$  points (the original pixels), which is too large for routine calculations. Therefore, the grid containing pixels is then condensed into  $n \times n \times n$  super-pixels using a typical condensation level  $n$ , of 3, 4 or 5. Super-pixels with negligible electron density (less than  $10^{-6}$  electrons) are then removed and the overall molecular electron density is usually described by 10,000-20,000 pixels. Space-group symmetry operations are then applied to obtain the electron density of the molecular cluster (Dunitz and Gavezzotti, 2005a).

PIXEL mimics the results of computational expensive high-level quantum mechanical calculations within a very small fraction of the computational cost (Gavezzotti, 2003b). On a standard desktop computer, PIXEL calculations take few minutes to one hour for crystals while it takes just a few seconds for calculations on a dimer extracted from a crystal structure. Other advantages of PIXEL include decomposition of total intermolecular energy into four individual components, few parameters, delocalisation of the electrical description of the molecule and many body description of polarisation.

The main disadvantage of PIXEL is the rigid treatment of electron densities, which neglects any rearrangement of electrons that could happen upon molecular contacts. It is especially an issue for strongly H-bonded systems where covalence or charge transfer can happen. However, careful parameterisation has allowed the theory to be effective in the treatment of weak



and medium strength H-bonds (Gavezzotti, 2005). PIXEL calculations are sensitive to intermolecular bond geometries especially in the positions of hydrogen atoms and minor differences in geometries of structures can lead to differences in energies. Care should be taken while comparing the stability using PIXEL energy differences calculated on the experimental structures. An alternative would be to optimise all geometries of all the structures using standard procedure before PIXEL calculations. However, this would neglect all the conformational changes that can occur during crystal packing. Despite the caveats, PIXEL calculations have been proved to be as successful as high-level quantum chemical calculations for gas-phase dimers (Gavezzotti, 2003b). It has also proved to perform comparably or sometimes better than atom-atom force field methods in ranking crystal energies for polymorphs of organic molecules (Gavezzotti, 2003c).

PIXEL calculations have been successfully implemented for many diverse applications such as examining the effect of pressure on intermolecular interactions in crystal lattice (Wood et al., 2008), studying the effect of fluorine substitution on intermolecular energies of hydrocarbons and fluorocarbons (Dunitz et al., 2003), rationalisation of packing motifs observed in the crystal lattice of cyclohexane (Ibberson et al., 2007), identifying the stabilising factors in the crystal lattice of energetic materials (Eckhardt and Gavezzotti, 2007), and investigating the polymorphic preference under certain conditions in structurally related analogues (Cruz Cabeza et al., 2006).

## **1.10 Experimental and Computational Studies on Structurally Related Compounds**

The number of new approved pharmaceuticals is decreasing whilst costs are increasing creating additional pressures on physical form screening to deliver comprehensive information more quickly and at a lower cost whilst minimising the risk that any practically relevant forms have been missed. Also molecules that do reach the later phases of drug development are becoming more structurally complex which creates further challenges for physical form screening, e.g. large numbers of solvates or disordered structures may take longer to characterise.

There is no standard approach for physical form screening given that every compound possesses unique properties and can display very different crystallisation behaviour. Small differences in substituents can affect the molecular conformation, shape, H-bonding potential and can have significant effects on packing forces and hence the adopted crystal structures and resultant physical properties. Both structure and properties, particularly of large complex molecules, remains challenging to predict. Some examples of studies which illustrate the challenges in predicting structure of chemically related analogues are described below.

The structurally related analogues hydrochlorothiazide (HCT) and chlorothiazide (CT) exhibit varied solid form diversity and demonstrate the effect of subtle changes on hydrogen bonding potential, molecular shape and overall solid form diversity (Johnston et al., 2007; Johnston et al., 2011). For

example HCT has two polymorphs whilst CT is monomorphic under ambient conditions with both compounds forming a range of solvates. The differences in molecular structures showed an impact on the opportunities for favourable H-bonding interactions in the crystal structures of these molecules. Whilst the molecules have similar size and shape and display some similarity in local intermolecular packing, the experimental and predicted crystal structures showed no overlap. Similar shape and dominant dispersion interactions within the lattices suggest that both molecules could adopt similar structures. But the isostructures of these molecules on their respective crystal energy landscapes were found to be unfavourable i.e. very high up in energy from global minimum and have also not been observed experimentally.

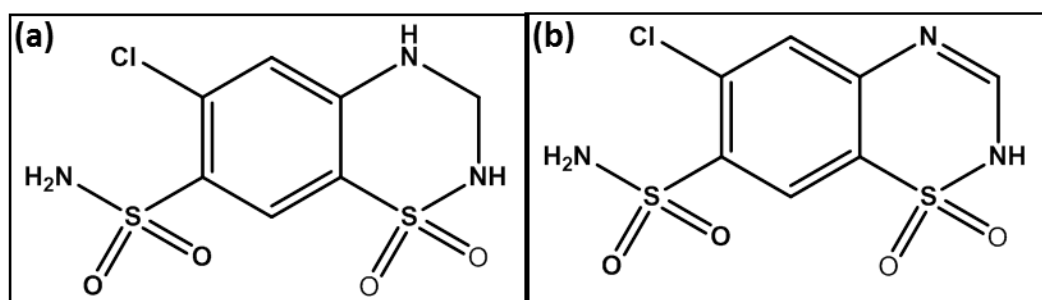


Figure 1.9. Chemical Structures of (a) hydrochlorothiazide and (b) chlorothiazide.

Similarly, no crystal packing similarity was observed in polymorphs of the antipsychotic drug aripiprazole and its structurally related metabolite dehydro-aripiprazole (Zeidan et al., 2013). Changing the anion type in a pharmaceutical salt system can also dramatically affect the observed solid form diversity and crystal packing. The hydrochloride salt of sertraline is highly polymorphic with 8 known polymorphs (Casimir and Andrew, 1997; Almarsson et al., 2003), while

its hydrobromide salt has only been shown to have one form (Remenar et al., 2003). This is more predictable perhaps due to the significant change in size of the anion i.e. chlorine vs. bromine. Similarly, changing the cation type e.g. from sodium to potassium in the salts of same active pharmaceutical ingredient has been shown to be responsible for large changes in crystal packing. This might also be due to the higher tendency of sodium salts to form hydrates (see Section 1.3) (Wood et al., 2012).

Thus from these and other examples it is clear that whilst it is often possible to propose reasons retrospectively as to why differences in molecular structure impact on differences in crystal structure, we cannot simulate *ab-initio* what these specific effects will be. Thus crystal structure prediction provides a powerful tool to assist in developing a comprehensive knowledge of solid-state behaviour of organic molecules. The next section outlines some specific examples where predictive tools and structure modelling can inform the experimental investigation of solid-state diversity.

The family of CBZ analogues i.e. CBZ, dihydrocarbamazepine (DHC), oxycarbamazepine (OCBZ) and epoxycarbamazepine (ECBZ) serves as a good example to demonstrate the effect of small changes in molecular structure on solid-state diversity and packing motifs preferences in the crystal lattice (Figure 1.10). CBZ, DHC and OCBZ have 5, 4 and 3 reported polymorphs, respectively in addition to forming solvates. However, ECBZ is monomorphic suggesting that the epoxy substituent is responsible for favouring certain structures or destabilising alternative ones and reducing the accessible forms. However, it

must also be borne in mind that the current view of the extent of diversity in this molecule could also be due to insufficient or incomplete screening for the compound i.e. not all feasible forms have actually been discovered yet.

Lutker et al. concluded that the dibenzazepine backbone and urea moiety of CBZ constitute a “polymorphophore” (Lutker and Matzger, 2010). A polymorphophore can be defined as “a structural element that, when incorporated into a molecule, favours the formation of polymorphic crystal forms” (Lutker et al., 2007).

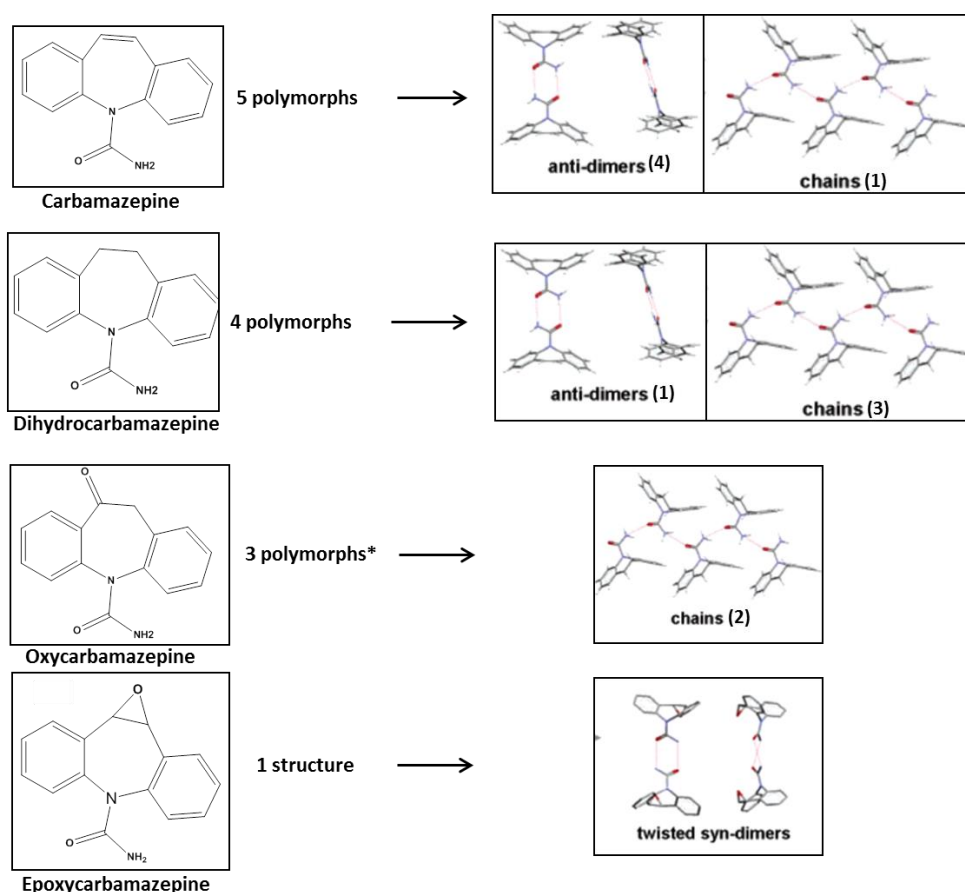


Figure 1.10. Polymorphs and hydrogen bonding motifs observed in carbamazepine and its analogues. The number of polymorphs having particular motif are shown in parentheses. \*For oxycarbamazepine, only crystal structures of two polymorphs are known.

Computational studies have shown the preference of chain motifs in the case of DHC and OCBZ due to the azepine ring distortion due to the different bonding compared to CBZ which prefers to adopt dimer motifs in the solid-state (Cruz Cabeza et al., 2006). This was also suggested by experimental studies in which comprehensive screening was carried out. However, more recently, the predicted and thermodynamically feasible dimer based polymorph of DHC was obtained from vapour phase crystallisation (Arlin et al., 2010). Similarly, the predicted chain motif based polymorph of CBZ was obtained by using DHC form II as a template (Arlin et al., 2011) exploiting the 3-D similarity between the predicted chain based form of CBZ and the experimental DHC form II. This emphasises the need for care when interpreting the outputs of computational studies. Also the relative difficulty in obtaining form V CBZ despite the predictions suggesting that they are energetically competitive with dimer based structures, highlights that current methods cannot provide kinetic favourability of individual structure formation.

Recently, CSP studies have been used to investigate the concept of polymorphophore using monomorphous fenamic acid and its analogue; tolfenamic acid for which 5 polymorphs have been reported (Figure 1.11) (Uzoh et al., 2012).

CSP studies on fenamic acid and tolfenamic acid revealed that all the crystal structures present on the crystal energy landscape consist of  $R_2^2(8)$  hydrogen bonded dimer (Etter, 1990) and differ in the packing of conformationally flexible phenyl rings linked by secondary amine. In the crystal energy landscape

of fenamic acid, only the known experimental crystal structure and a closely related crystal structure are thermodynamically favoured over all the other calculated crystal structures, thus limiting the extent of potential polymorphism.

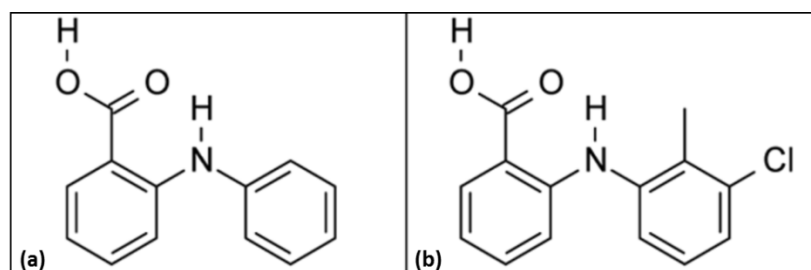


Figure 1.11. Chemical structures of (a) fenamic acid (monomorphic) and (b) tolfenamic acid (5 polymorphs).

In case of tolfenamic acid, the introduction of methyl and chlorine substituents change the crystal energy landscape and produce a range of additional energetically competitive structures, including the experimentally observed polymorphs. This study suggests that the fenamate fragment constitutes a polymorphophore and can potentially form a large number of structures due to associated conformational flexibility. However, the substituents on the fenamate fragment determine whether the specific fenamate will be polymorphic or not (Uzoh et al., 2012). This suggests that a polymorphophore may promote but does not necessarily guarantee polymorphism.

Figure 1.12 shows examples of other polymorphic fenamic acid analogues with different substituents at various positions of phenyl ring or pyridine ring instead of phenyl ring (López-Mejías et al., 2012). However there is no available information about their crystal energy landscapes.

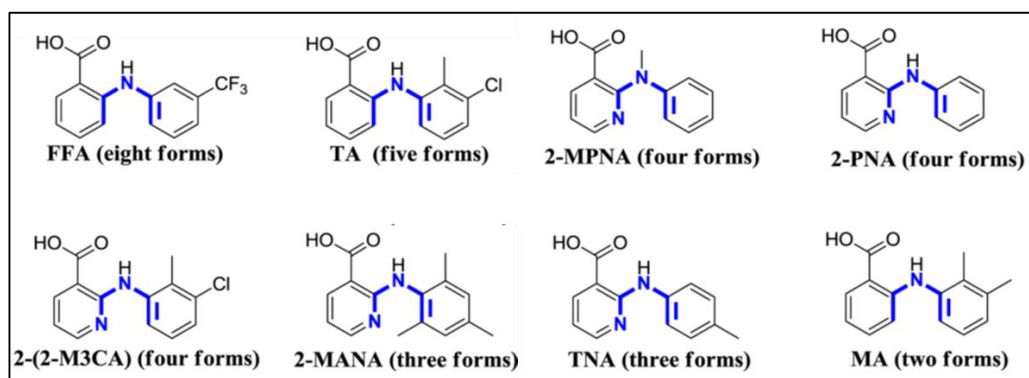


Figure 1.12. Molecular structures of fenamic acid analogues. The number of structurally characterised polymorphs for each compound is written in parentheses. Reproduced from reference (López-Mejías et al., 2012).

The above mentioned examples demonstrate that small changes in substituents can affect the molecular conformation, shape, H-bonding potential and can have significant effect on the packing forces and hence the adopted crystal structures and overall solid-state diversity. However, it is still not possible to predict the magnitude of changes and the resultant physicochemical and mechanical properties. Given the associated challenges and subtleties, it is therefore of continued interest to carry out more experimental and CSP studies to explore the basic science underpinning the structural diversity in structurally related pharmaceutical compounds.

This thesis is concerned with the methodology development for discovery of physical forms, prediction of crystallisability and demonstrates the use of combined experimental and computational studies to explore the solid-state diversity of structurally related molecules in two groups. Rigorous experimental screening utilising multiple crystallisation techniques is presented to achieve the maximum coverage of experimental crystallisation space and to maximise



the probability of encountering all relevant forms by sampling nucleation/crystallisation outcome from a broad range of conditions. CSP studies were used as a complement to experimental screening. Thus each molecule has been subjected to CSP and experimental searches to explore and understand the range of structural forms obtained and/or that may be possible.

In addition to CSP studies, PIXEL calculations were used to provide a better understanding of the stabilising factors and most important interaction(s) in the crystal lattice using a step-wise approach to obtain a holistic understanding of the structural and energetic factors involved in stabilising the different crystal structures. These calculations can help in understanding the stability of one structure over others and non-observance of hypothetical/predicted structures during experimentation. Better understanding of the role of specific interactions in stabilising the structure can, in principle, help in achieving a better control over the solid-state form of pharmaceuticals and other molecular solids.

Structural analysis using XPac and Mercury has enabled the identification of structural relationships and characterisation of the key structural motifs that underpin the formation of the experimental structures. The applications of statistical modelling techniques in diverse solvent selection and identifying the underlying relationship between crystallisation conditions and its outcome have also been demonstrated.

## **2 Aims and Objectives**

## 2.1 Aims

Control of crystal structure by crystallisation to deliver chemically and physically pure substances is of significant importance in the manufacture of safe, effective and high quality pharmaceutical products. The overall aim of this work is to investigate a range of experimental and computational approaches for the discovery and better understanding of the key factors underpinning solid-state structure and diversity of pharmaceuticals. A range of crystallisation techniques and analytical tools has been applied in combination with statistical modelling approaches, CSP, and a number of semi-empirical and quantum-mechanical approaches to study the molecular and crystal structures.

The work reported in this thesis has focussed on discovery of physical forms and exploration of the relationship between molecular structure and crystallisability. It has also concentrated on two groups each comprising of two structurally related molecules to investigate the influence of changes in the molecular structures and substituents on the conformation, intermolecular interactions, packing motifs in the crystal lattice, and overall extent of solid form diversity and to probe how predictions can aid in the interpretation of the experimental crystal energy landscape. The compounds studied are: olanzapine (OZPN) and clozapine (CZPN) in group 1, and loxapine (LXPN) and amoxapine (AXPN) in group 2 (Figure 2.1).

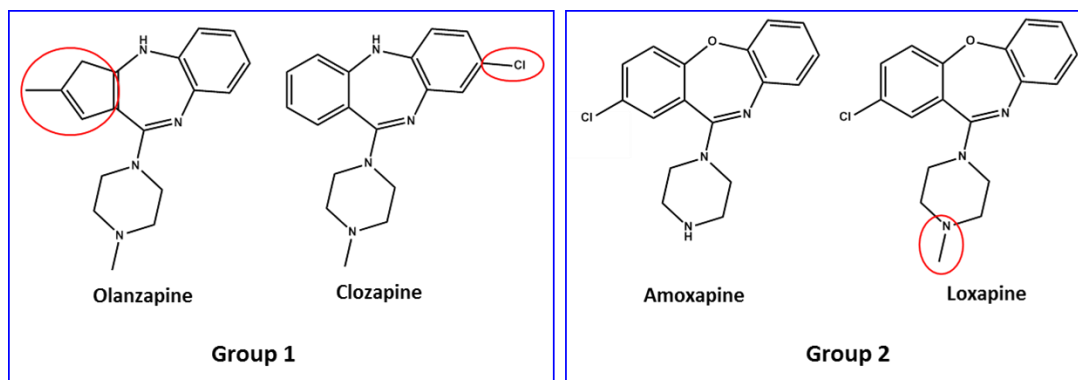


Figure 2.1. The four compounds studied in this thesis. The differences in the chemical structures within a group are highlighted in red.

## 2.2 Objectives

- Development and validation of high throughput crystallisation and analysis (HTCAA) methodology for physical form screening using quartz 96/48-well plates with an automated system for collecting high quality Raman spectra.
- Development of a statistical model for prediction of the crystallisability of a set of organic molecules based on calculated physicochemical descriptors.
- Comprehensive experimental physical form screening of molecules in group 1 (OZPN and CZPN) as well as in group 2 (AXPN and LXPN) by employing multiple crystallisation techniques to identify novel physical forms and determine their crystal structures using X-ray powder diffraction (XRPD) or single crystal X-ray diffraction (SXD).
- Exploitation of predicted crystal energy landscapes to aid the accurate interpretation of the experimental physical form screening results.
- Application of molecular packing analysis tools to the crystal structures of target molecules to identify key packing motifs and structural relationship amongst them.
- Use of computational techniques such as statistical modelling approaches and PIXEL calculations to develop a comprehensive understanding of the structural and thermodynamic factors directing crystal packing in different physical forms.
- Application of geometry optimisation calculations using CASTEP to obtain accurate H-atoms position and verification of the accuracy of the crystal structures solved by XRPD.

### **3 Material and Methods**

## 3.1 Material

Olanzapine (OZPN) was obtained from Molekula Ltd, UK and Eli Lilly, US. Clozapine (CZPN), amoxapine (AXPN) and Loxapine (LXPN) were purchased from Molekula Ltd, UK. LXPN was also procured from Eurolab Ltd, UK. All four compounds are listed in Table 3.1. X-ray powder diffraction (XRPD) was used to confirm the identity and phase purity of all the compounds. For all crystallisations, analytical/reagent grade solvents were purchased from Sigma-Aldrich or Fluka (UK).

Table 3.1. Compounds studied in this work.

Compound	Molecular Weight (g mol <sup>-1</sup> )	Melting Point (K)
Olanzapine (OZPN)	312.44	468
Clozapine (CZPN)	326.82	456
Amoxapine (AXPN)	313.78	448-449
Loxapine (LXPN)	327.81	382-383

## 3.2 Methods

### 3.2.1 Crystallisation Techniques

Various crystallisation techniques were adopted during the experimental screens of selected molecules to maximise the coverage of the experimental crystallisation space. The experimental screen was based largely on solution crystallisation using a variety of solvents selected to cover a wide range of physicochemical properties. Details of crystallisation techniques adopted in this study are mentioned in the following sections.

### ***3.2.1.1 Slow and Fast Solvent Evaporation***

Saturated solutions of the compound were prepared by introducing an excess of compound into 4-7 ml of the chosen solvent under agitation. After equilibration, samples were filtered to remove seeds and left to slowly evaporate at RT and at 277 K. For faster solvent evaporation, samples were left to evaporate on a watch glass at RT.

### ***3.2.1.2 Semiautomatic Cooling Crystallisation on React Array Platform***

Saturated solutions were prepared at a temperature,  $T_{\text{sat}}$ , equivalent to the solvent boiling point minus 10 K with a heating rate of 40 K min<sup>-1</sup>. Saturated solutions were then cooled to RT at the rate of 25 K min<sup>-1</sup>.

### ***3.2.1.3 Mechanical Grinding***

Mechanical grinding was carried out in pestle mortar or in a Retsch MM400 mixer mill or in a Retsch Cryomill. Approximately 150 mg of the compound was placed in a stainless steel grinding jar containing 2 mm stainless steel ball (with or without a drop of a chosen solvent) and ground at RT (MM400 mixer mill) or at 77 K (Cryomill) using liquid nitrogen.

### ***3.2.1.4 Crystallisation from Vapour Phase***

Approximately 30 mg of the compound was placed at the bottom of the 3/5/10 ml glass vial. A glass fibre was suspended in the selected vial on a Cu wire



placed ~1 cm above the sample. The glass vial was then sealed and kept on a hot plate at 353 K and the crystals were retrieved from the glass fibre or the walls.

#### **3.2.1.5 Melt Quenching**

The compound was melted using a hot plate at *ca.* 453 K on glass microscope slides. The slides were then either quenched in liquid nitrogen or left at RT to cool. Solidified compound was then scraped from the glass slide and analysed using XRPD and thermal techniques.

#### **3.2.1.6 Freeze Drying**

Freeze drying was carried out using bench top freeze-drier (Epsilon 2-4, Martin Christ). A saturated solution of the compound in t-butanol was prepared at RT and freeze drying was carried out using slow or fast freeze drying cycles. During slow freeze drying, the sample was transferred to the freeze drying chamber from RT, whilst during fast freeze drying, the sample was cooled to 193 K and then shifted to the freeze drying chamber. The temperature was maintained at 253 K for 48 hours. The vacuum applied during the primary and secondary drying phases was equivalent to a pressure of 1 and 0.015 mBar, respectively. The resulting samples were analysed using XRPD and thermal techniques and Raman spectroscopy.

## 3.2.2 X-ray Crystallography

### 3.2.2.1 X-ray Powder Diffraction

Polycrystalline samples were analysed by XRPD for fingerprinting, structure solution and refinement. Also XRPD was used for *in-situ* desolvation and amorphous recrystallisation studies as part of variable temperature (VT)-XRPD studies.

For sample fingerprinting, 10-50 mg of sample was placed on 28 well plate supported on a polyimide (Kapton, 7.5  $\mu\text{m}$  thickness) film. Data were collected on a Bruker AXS D8-Advance transmission diffractometer equipped with  $\theta/\theta$  geometry, primary monochromated radiation ( $\text{Cu K}\alpha_1$ ,  $\lambda = 1.54056 \text{ \AA}$ ), a Vantec 1-D position sensitive detector (PSD) and an automated multi-position  $x$ - $y$  sample stage. Data were collected in the range  $4$ - $35^\circ 2\theta$  with a  $0.015^\circ 2\theta$  step size and  $1 \text{ s step}^{-1}$  count time (Florence et al., 2003). Samples were oscillated  $\pm 0.5 \text{ mm}$  in the  $x$ - $y$  plane at a speed of  $0.3 \text{ mm sec}^{-1}$  throughout data collection to maximise particle sampling and minimise preferred orientation effects.

Capillary XRPD data were used for indexing attempts, phase identification using Pawley-type fits (Pawley, 1981) and structural analysis using Rietveld refinement (Rietveld, 1969). Samples were lightly ground in an agate pestle mortar and loaded into a  $0.7 \text{ mm}$  internal diameter borosilicate capillary. The data were collected on the Bruker AXS D8 Advance powder diffractometer equipped with transmission capillary geometry, primary monochromated radiation ( $\text{Cu K}\alpha_1$ ,  $\lambda = 1.54056 \text{ \AA}$ ) and a Bruker Lynxeye PSD. The capillary was

rotated through data collection to avoid preferred orientation effects as far as possible. Data were collected in the range of 2-40° 2 $\theta$  with a 0.015° 2 $\theta$  step size and using a 1 or 3 or 8 s step<sup>-1</sup> count time. Variable Count Time (VCT) data, typically in the range of 2-70° 2 $\theta$  with step times ranging from 2-24 s were collected for Rietveld refinement to ensure more accurate intensity estimates at high angles as required by structural analyses.

For VT experiments, sample temperature was controlled using an Oxford Cryosystem 700 Plus series device to enable *in-situ* sample transformation. All XRPD patterns were visualised using DIFFRAC<sup>plus</sup> EVA (version 2.0) program (BRUKER-AXS).

Data analysis was carried out using DASH (version 3.2/3.3) (David et al., 2006) or TOPAS Academic (version 4.1) (Coelho, 2007) software. For structure determination, the powder pattern was indexed using first 20 peaks using DICVOL04 (Boultif and Louer, 1991) or singular valued decomposition (SVD)-Index algorithm (Coelho, 2003a) in TOPAS (version 3.1). The space group was determined from volume and statistical assessment of the systematic absences as implemented in DASH structure solution package. Pawley refinement (Pawley, 1981) was used to extract the intensities and their correlations by refining lattice parameters along with background, zero-point, reflection intensities and peak shape parameters. For structure solution (David et al., 1998), an internal coordinate (z-matrix) description of the molecules was derived from the crystal structures reported in the Cambridge Structural Database (CSD) (version 5.33 and 5.34) (Frank, 2002), after a geometry check

within the MOGUL program (version 1.4 and 1.5) (Bruno et al., 2004). The O-H, N-H and C-H distances in an internal model were normalised to 0.90, 0.90 and 0.95 Å respectively. The structure was solved using 800 simulated annealing (SA) runs distributed on multi-core processor using MDASH (version 1.0) (Griffin et al., 2009). The SA solution with lowest profile  $\chi^2$  was used as a starting point for a rigid body/restrained Rietveld refinement in TOPAS Academic (version 4.1). In some cases, the SA solution was also taken for geometry optimisation in CASTEP (Cambridge Sequential Total Energy Package) (Clark et al., 2005) to verify atomic positions.

### ***3.2.2.2 Single Crystal X-ray Diffraction Analysis***

SXD data were collected on a Bruker Apex II diffractometer with charge coupled device (CCD) detector and graphite monochromated Mo  $K\alpha_1$  radiation ( $\lambda = 0.71073$  Å) with temperature control from an Oxford Cryosystems Cryostream device operating at 123 K. Diffraction data were processed (cell refinement and data reduction) using Bruker SAINT and APEX2 software (APEX2, 2005). Structure solution was carried out using direct methods in SHELXS-97 (Sheldrick, 2008) within the WinGX suite (Farrugia, 1999). The hydrogen atoms positions were placed from the difference maps or from calculated positions using a riding model. Atomic coordinates and thermal parameters were refined using full-matrix least square methods on  $|F^2|$  using all the unique data using SHELXL-97 (Sheldrick, 2008) within the WinGX suite (Farrugia, 1999; Farrugia, 2012). The refined structures were inspected using ORTEP-3 (version 2.02) (Farrugia, 2012) and Platon (version 1.15) (Spek, 2003) within the WinGX suite.

### 3.2.3 Thermal Analysis

Thermal behaviour, mass loss and melting points of all samples were monitored using simultaneous thermal analyses (STA), comprising differential scanning calorimetry (DSC) and thermal gravimetric analysis (TGA). This was carried out on a Netzsch STA 449 C thermocouple, equipped with a Netzsch CC 200 liquid nitrogen supply system and a Netzsch CC 200 C control unit. Samples were placed in 10  $\mu$ L aluminium pans (lid with a pinhole) and a heating ramp was set in the range of 273-483 K at a rate of 2-20 K min<sup>-1</sup>. A thermal heat-cool-heat/melt-recrystallisation-melt cycle for AXPN (see Section 7.3.2) comprises of heating of the sample to 457 K from RT at the rate of 10 K min<sup>-1</sup>, followed by cooling to 263 K at the rate of 40 K min<sup>-1</sup> and then again heating to 480 K at the rate of 10 K min<sup>-1</sup>. All analyses were performed using the Netzsch Proteus thermal analysis program<sup>®</sup> (version 4.3.1) (NETZSCH-Gerätebau GmbH).

### 3.2.4 Raman and Infrared Spectroscopy

Raman spectra were collected using DXR Raman instrument (Thermo Scientific) equipped with a high precision x, y, z motorised stage. A 780 nm (frequency-stabilised single mode diode laser) or 532 nm (diode-pumped, solid state) laser with maximum laser power of 10 mW was used as an excitation source, coupled with air cooled CCD for collecting the Raman spectra in the shift region of 50-3400 cm<sup>-1</sup>. Spectra were collected using a long working distance 10/20x objective with 50  $\mu$ m aperture, with 1-10 s exposure and a spectral resolution of

5 cm<sup>-1</sup>. All the spectra were visualised and analysed with OMNIC™ software (version 8.2) (Thermo Scientific).

Fourier Transform Infrared (FT-IR) spectra were collected using Smart iTR NICOLET iS10 (Thermo Scientific). The spectra were recorded in the range of 4000 to 525 cm<sup>-1</sup> at an instrument spectral resolution of 4 cm<sup>-1</sup> (64 scans per spectrum). The spectra were visualised and analysed with OMNIC™ software (version 8.1) (Thermo Scientific).

### **3.2.5 PIXEL Calculations**

PIXEL (Gavezzotti, 2002; Gavezzotti, 2003a; Gavezzotti, 2003c) calculations were carried out to estimate the total lattice energy, contribution of polarisation ( $E_p$ ), dispersion ( $E_d$ ), Coulombic ( $E_c$ ) and repulsive ( $E_r$ ) interactions towards the total lattice energy ( $E_{tot}$ ) and identify the most significant intermolecular contacts within the crystal structures. The molecular conformation was taken from each crystal structure, with hydrogen bond lengths adjusted to standard neutron values for the experimentally determined structures (1.08 Å and 1.00 Å for C-H and N-H respectively). The electron density of the structures was calculated using GAUSSIAN® program (version 6.1) (GAUSSIAN03W, 1995) at the MP2/6-31G\*\* level of theory. A cube representing the electron density was generated using medium cube settings and a step size of 0.08 Å. Pixels were then condensed into superpixels with a condensation level  $n=4$ . Lattice energy calculations were carried out on a cluster of molecules within a maximum distance of 18 Å from a central molecule and a maximum radius for the search

of 35 Å. In case of OZPN, the effect of thermal expansion on the calculated energy contributions was also investigated by carrying out the PIXEL calculations on the experimental structures (123 K and 298 K) and predicted structures (0 K).

### 3.2.6 Crystal Packing Analysis

XPac (version 2.0.1 and 2.0.2) (Gelbrich and Hursthouse, 2005) and Mercury (version 3.0 and 3.1) (Macrae et al., 2006) were used to analyse the experimental and predicted crystal structures. For the XPac analysis, routine medium cut-off parameters ( $\delta_{\text{ang}} = 10^\circ$ ,  $\delta_{\text{tor}}$  and  $\delta_{\text{dhd}} = 14^\circ$ ) and a van der Waals search radius of 1.5 Å were used. The similarity of pairs of the crystal structures and the molecular conformations in them were measured by calculating the root mean square deviation (RMSD<sub>n</sub>) of the *n* molecule overlay using the Crystal Packing Similarity tool within Mercury. Packing coefficients were calculated using academic version of Mercury available at University College London.

**4 Development and Validation of High-Throughput  
Crystallisation and Analysis (HTCAA) Methodology  
for Physical Form Screening**



## 4.1 Introduction

Physical form screening, i.e. generating and analysing various physical forms of an API has become an essential part of drug development and product lifecycle management (Gardner et al., 2004; Aaltonen et al., 2009). Physical form screening is not only of concern at the preformulation stage during drug development, but is monitored continuously during scale-up and manufacturing to detect late appearing polymorphs and possible API-exciipient interactions. The aim of comprehensive physical form screening is to achieve the maximum coverage of experimental crystallisation space or to achieve the greatest possible knowledge within the constraints of available time, material and resources (Florence, 2009b).

It is generally believed that a large number of crystallisation trials increase the likelihood of finding the most suitable physical form for drug development (Aaltonen et al., 2009). HTC studies rely on carrying out a large number of small-scale (mg of solute with  $\mu\text{l}$  of solvent) parallel crystallisations commonly using a multi-well plate (MWP) or multi-rack format. These approaches rely on extensive use of automation to carry out a large number of crystallisations using robotics and computer control to execute some or all steps involved in crystallisation (Florence, 2009b) and *in-situ* analysis using fingerprinting techniques such as Raman spectroscopy or XRPD. Data mining tools are then applied to identify the new solid-state forms and the HTC results are stored in databases for future use (Aaltonen et al., 2009).

A number of approaches with a range of capabilities and benefits have been applied for HTC in search of physical forms of the pharmaceuticals. Some examples of physical form screening using automated methods, conditions and their results are summarised in Table 4.1.

HTC offers the advantage of exploring a large property space of a molecule requiring less material and time compared with traditional manual approaches. It is worth noting that in spite of the large number of crystallisation trials, the associated hit rate (number of experiments where API crystallise) with HTC can be very low e.g. between 2.5 to 13.0% (Peterson et al., 2002; Almarsson et al., 2003; Morissette et al., 2003). The ultimate advantage of HTC lies in carrying out a large number of experiments, utilising mg of API,  $\mu$ l of solvent with little or no manual intervention (Aaltonen et al., 2009).

The need for HTC is driven by the requirement to take the most favourable physical form (polymorph, solvate or salt) to the market in the minimum amount of time. HTC can help not only in eliminating the risk associated with the development compound quickly, but also in finding ways to mitigate the risk of failures during later stage of drug development (Balbach and Korn, 2004). Comprehensive information about various physical forms from HTC also provides an opportunity for maximum coverage of intellectual property space and helps in addressing the key regulatory questions e.g. number of physical forms of an API and the relationships amongst them (Morissette et al., 2004).

Table 4.1. Examples of high-throughput crystallisation carried out for physical form screening of pharmaceuticals. Reproduced from reference (Aaltonen et al., 2009).

API	Platform/Company	No. of Crystallisations/ No. of Solvents	HT Crystallisation Method	HT Analytical Method	No. of Physical Forms Found	Reference
Carbamazepine	Chemspeed Accelerator SLT100	594 crystallisations/66 solvents	Cooling, evaporation, cooling and evaporation	XRPD	3 polymorphs (forms I, II and III), 9 solvates	(Florence et al., 2006)
Cimetidine	Symyx technologies, Inc.	288 crystallisations/20 solvents, 84 solvent mixtures	Cooling, evaporation, precipitation	Raman, XRPD, polarising light microscopy	3 polymorphs	(Desrosiers, 2004)
Carbamazepine	Solvias AG	Number not available/43 solvents and mixtures	Evaporation, suspension, desolvation	Raman	4 known polymorphs, 2 known solvates (dihydrate and acetone), 5 suggested new forms (incl. 2 solvates)	(Hilfiker et al., 2003)
Angiotensin II receptor antagonist MK-996	CrystalMax™/ TransForm Pharmaceuticals, Inc.	1440 crystallisations/ 21 solvents and mixtures	Cooling (3 nominal concentration levels), separate hydrate screen	Raman, XRPD	18 crystalline forms	(Almarsson et al., 2003)
Sertraline HCl	CrystalMax™/ TransForm Pharmaceuticals, Inc.	3072 crystallisations/ 24 solvents and mixtures	Cooling (2 nominal concentration levels)	Raman, XRPD	10 forms (4 additional forms by follow up studies)	(Almarsson et al., 2003)
Ritonavir	CrystalMax™/ TransForm Pharmaceuticals, Inc.	2000+ crystallisations/ 24 solvents (single and binary mixtures)	Cooling (4 nominal concentration levels)	Raman	3 polymorphs (I, II, IV), 1 solvate (III). Follow up studies revealed a trihydrate (Form V)	(Morissette et al., 2003)
Acetaminophen	CrystalMax™/ TransForm Pharmaceuticals, Inc.	7776 crystallisations (first iteration)/16 solvents (single and binary mixtures)	Cooling (3 nominal concentration levels)	Raman	2 polymorphs (I and II). Form III was found in subsequent studies	(Peterson et al., 2002)

This chapter reports the development and validation of a High-Throughput Crystallisation and Analysis (HTCAA) methodology using 96/48 quartz well plates combined with automated data collection using Raman spectroscopy for physical form screening. This methodology is complementary to larger scale (3-8 mL) parallel approaches and enables studies where only small quantities (1-2 mg per crystallisation) of a compound are available. A schematic illustration of the experimental workflow for the HTCAA approach is shown in Figure 4.1. Development and validation of steps 1-3 in Figure 4.1 were included in this investigation. Novel physical forms identified during cluster analysis (step 3 in Figure 4.1) were then prepared on a scale of ~ 150-200 mg. This was followed by the characterisation of the novel physical forms using thermal analysis, XRPD and/or SXD.

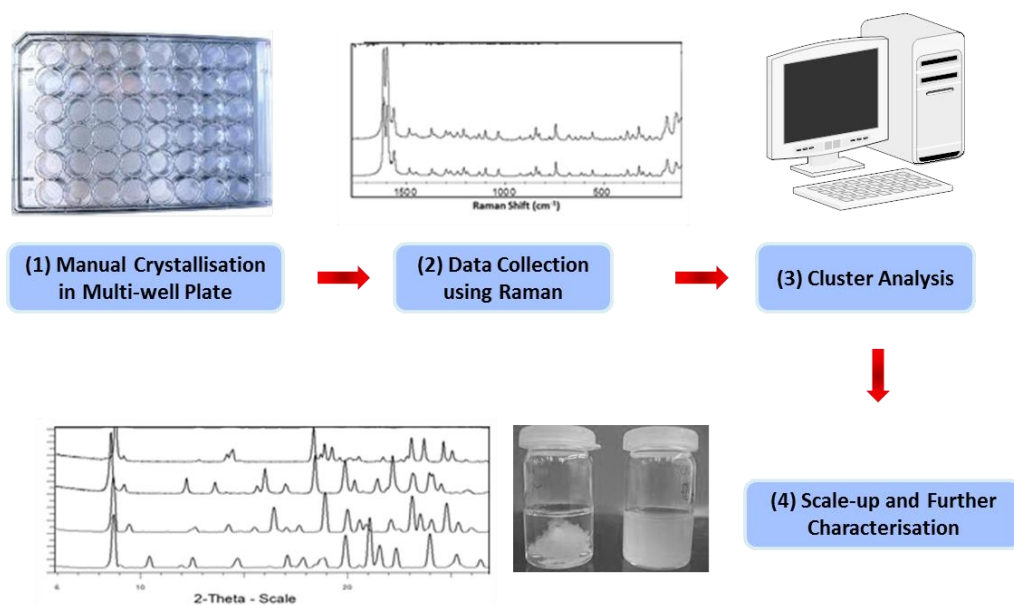


Figure 4.1. Illustration of experimental workflow for high-throughput crystallisation and analysis methodology.

The HTCAA methodology developed here is not intended for comparison with highly automated and sophisticated systems (e.g. Crystal Max™ and Symyx) summarised in Table 4.1 or with other larger scale systems e.g. fully automated parallel solution crystallisation platform which requires 150 mg of solute with 3-8 ml of solvent per crystallisation (Florence et al., 2006). The target was development of a complementary, low-cost, and efficient integrated methodology for small scale (1-2 mg) crystallisations and analysis to provide information on solid-state behaviour of a compound from a range of solvents when only limited amount of material is available. This would involve establishing and testing a robust workflow combining small scale crystallisation, Raman microscopy and chemometrics for fast and cost-effective physical form screening.

Raman spectroscopy was chosen as an analytical method as it is an efficient tool in differentiating between different physical forms like polymorphs, solvates, co-crystals and salts (Findlay and Bugay, 1998). Raman spectroscopy is based on inelastic scattering of monochromatic excitation source (usually a laser source) upon contact with a sample. The change in the frequencies of the absorbed and remitted photons of the monochromatic excitation source by the sample is measured. This shift provides information about the vibrational, rotational and other low-frequency modes in molecules. Raman spectroscopy offers various advantages (e.g. no specific sample preparation, less material requirement and faster data collection) over other analytical techniques (Bugay, 2001). It also provides physical and chemical information about a compound (Peterson et al., 2002; Kojima et al., 2006; Al-Dulaimi et al., 2010).

There are a few literature reports of semi-automated salts screening (Balbach and Korn, 2004; Kojima et al., 2006) and co-crystals screening (Kojima et al., 2010) in MWP using Raman spectroscopy, however no chemometric analysis was carried out on the Raman spectra to aid application in a HTC context. Another report using Raman and optical microscopy has demonstrated the discovery and selection of polymorphs of CBZ, acetaminophen, sulfamethoxazole, and a pharmaceutical intermediate 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile (ROY) using polymer heteronuclei (Balbach and Korn, 2004; Price et al., 2005).

The volume of the data generated from MWP approaches can be overwhelming for individual spectral analysis and hence require effective data analysis tools. In this work a non-supervised chemometric tool called group analysis was used to identify novel physical forms from the Raman spectra collected. Whilst using adequate data pre-treatment to remove baseline effects and appropriate grouping (Lowry et al., 2006), useful information was extracted from the spectral data generated from MWP, which was subsequently utilised to guide larger scale (~150-200 mg) crystallisation efforts. Specifically, the application of HTCAA methodology has been developed and assessed using salt screening of AXPN and polymorph/solvate screening of OZPN and CZPN (see Section 2.2).

Salt screening for an API is a common process in pharmaceutical industry. The increased number of poorly soluble compounds emerging from modern drug discovery approaches leads to increased utilisation of salt forms to increase aqueous solubility and hence bioavailability where the compound contains

ionisable functional groups. Salts formation may also be employed to increase chemical stability or to reduce the solubility of an API in sustained released dosage forms (Ashby and Wang, 1996). Furthermore, salts and other crystal forms (solvates, hydrates, and co-crystals) of a given molecular entity can enhance the intellectual property estate enabling extended protection of valuable market shares through patents (Clark et al., 2005). For a comprehensive understanding of physical form diversity, it is advisable to combine salt selection studies with simultaneous polymorph screening (Remenar et al., 2003).

AXPN is a basic drug with calculated pKa values of 9.0, 8.2 and 4.6 for the secondary amine, tertiary amine and imine nitrogen atoms respectively (Figure 4.2) estimated using Pipeline Pilot Professional Client interface (Accelrys, 2010). In addition to salt screening of AXPN using HTCAA, solvent incorporation in salts and polymorphic nature of salts was also assessed.

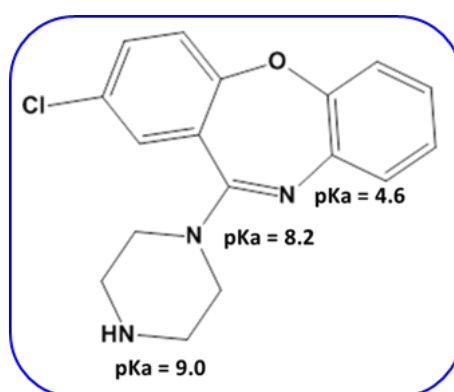


Figure 4.2. Calculated pKa of nitrogen atoms in amoxapine

The HTCAA method was also applied to physical form (polymorph/solvate) screening of OZPN and CZPN (Figure 4.3) to obtain a preliminary view of their

solid-state behaviour in a selected solvent system. Although hydrates are sometimes utilised, solvates are not generally suitable for drug development, although may be used as precursors to the final form. Hence a robust knowledge of solvate formation, particularly in process relevant solvents, and their identification is essential to understand the solid-state behaviour of the compound (Anderton, 2007).

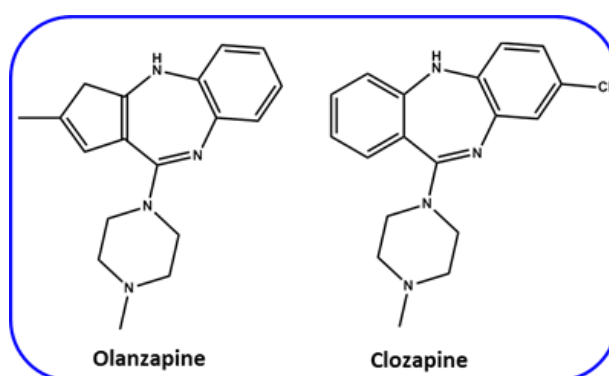


Figure 4.3. Chemical structures of olanzapine and clozapine

## 4.2 Sample Preparation and Methodology Development

### 4.2.1 96/48 Quartz Multi-well Plate

There are many MWP available in terms of size, number of wells and material of construction. MWP selection was based on the following criteria: solvent/chemical compatibility, low/no fluorescence, reusable, and high temperature tolerance ( $\sim 800^{\circ}\text{C}$ ). Quartz 96 and 48-well plates customised for use with the DXR Raman microscope (Thermoscientific, Madison, USA) were procured from Hellma, Germany. The layout of both 96 and 48-well plates



showing typical configurations for salt screening and anti-solvent screening respectively are shown in Figure 4.4.

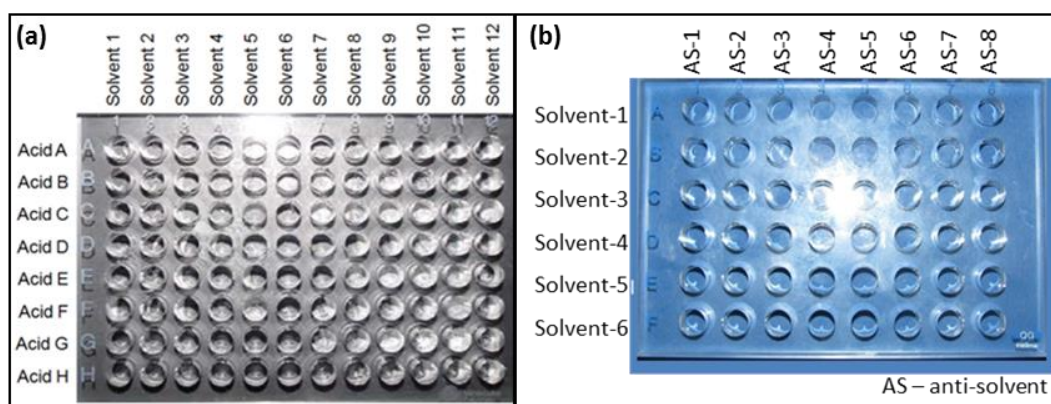
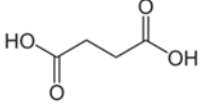
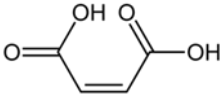
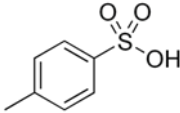



Figure 4.4. Layout of (a) salt screening in the 96-well plate, and (b) anti-solvent screening in the 48-well plate. A polymorph screen could be carried out using a different solvent in each well for example.

#### 4.2.2 Preparation of 96-well Plate for Salt Screening of Amoxapine

Salt screening of AXPN was carried out in 96-well plate using four different acids and two concentrations of AXPN in 12 solvents. A set of 4 pharmacologically safe mono and dibasic acids (Table 4.2) were selected for salt screening purpose and purchased from Sigma-Aldrich, UK. These acids were expected to form stable salts with AXPN as the pKa difference between ionisable group on AXPN and counter acid is more than 3 units (Bastin et al., 2000).

Table 4.2. Counter ions used for salt screening of amoxapine

Name	Structure	pKa	Melting Point (K)
Succinic acid		4.2, 5.6	457
Maleic acid		1.9, 6.1	408
para-toluene sulfonic acid (PTSA)		-2.8	311
Acetic acid		4.8	289

All the crystallisations were set up manually. The solutions (75 mM) of AXPB and the acids were prepared in ethyl acetate prior to use. The solution of AXPB was placed in each well of rows A, C, E, G and H (60  $\mu$ l per well) and B, D and F (120  $\mu$ l per well) of the 96-well plate (Figure 4.4). To each well in the first seven rows, 60  $\mu$ l of ethyl acetate solution of succinic acid (row A-B) or maleic acid (row C-D) or PTSA (row E-F) or acetic acid (row G) was added. This resulted in a binary mixture of AXPB and acid in the molar ratio of 1:1 in rows A, C, E, G and 2:1 in rows B, D, F. Row H contained only AXPB in ethyl acetate as a control. The plate was then left in a fume hood at RT to allow solvent evaporation. After complete evaporation of solvent, 200  $\mu$ l of crystallisation solvent [one solvent per column from methanol, ethanol, acetonitrile, acetone, ethyl acetate, isobutyl acetate, nitromethane, cyclohexane, toluene, diisopropyl ether, tetrahydrofuran (THF) and water] was added to each column of the plate. The MWP was then sealed using transparent sealing films (ThermaSeal RTS™) and shaken using an

orbital shaker at RT for 4 hours to allow dissolution of solid. Following dissolution, the seal was removed and the MWP was left in the fume hood to allow slow solvent evaporation at RT. The MWP was then checked for crystals using polarised light microscopy and analysed using the Raman microscope.

### **4.2.3 Preparation of 48-well Plate for Physical Form Screening of Clozapine**

Two physical form screening experiments (PS-1 and PS-2) were carried out using a single concentration of solute and a 96-solvents system at RT in a 48-well plate. A solution of CZPN (75 mM) was prepared in ethyl acetate prior to use and 100  $\mu$ l was placed in each well of the 48-well quartz plate. The MWP was kept in the fume hood at RT for evaporation of the ethyl acetate to leave approximately 2.4 mg of solid after drying. 250  $\mu$ l of the crystallisation solvent (one per well chosen from 24 solvents in the same sequence as listed in Table 4.3) was added into the first 24 wells (A1-C8). 125  $\mu$ l of the same crystallisation solvent (one per well, in the same sequence as in wells A1-C8) was placed in the next 24 wells (D1-F8). In these wells (D1-F8), a further 125  $\mu$ l of water was added to make a 1:1 v/v mixture of water and each solvent. The MWP was sealed using transparent sealing films (ThermaSeal RTS™) and placed on an orbital shaker for 2 hours to redissolve the CZPN. The plate was visually inspected to check dissolution of CZPN. Following dissolution of CZPN in the majority of wells, the seal was removed and the MWP was left in a fume hood at RT to allow slow evaporation of the solvents. The plate was inspected regularly

for evidence of crystallisation. Following appearance of crystals, data were collected using the Raman microscope (see Section 4.2.5). A similar procedure was repeated for PS-2 with solvents listed in Table 4.4.

Table 4.3. Solvents used for physical form screening of clozapine in experiment 1 (PS-1) using 48-well plate.

S. No	Solvent	S. No	Solvent	S. No	Solvent
1	Methanol	9	Pentyl acetate	17	Acetonitrile
2	Ethanol	10	Butyl acetate	18	1,4-dioxane
3	1-propanol	11	Acetone	19	Pyridine
4	2-propanol	12	Methyl-ethyl ketone	20	Nitro methane
5	1-butanol	13	Diethyl ether	21	Tetrahydrofuran
6	2-butanol	14	Tertiary-butyl methyl ether	22	Dichloro methane
7	Methyl acetate	15	Toluene	23	Water
8	Ethyl acetate	16	1,1-dichloroethane	24	Acetic acid

Table 4.4. Solvents used for physical form screening of clozapine in experiment 2 (PS-2) using 48-well plate.

S. No	Solvent	S. No	Solvent	S. No	Solvent
1	1-pentanol	9	Pentyl acetate	17	4-methyl-2-pentanone
2	2-pentanol	10	Trichloroethylene	18	2-ethoxy ethanol
3	1-hexanol	11	Chloroform	19	2,2,2-trifluoro ethanol
4	Cyclohexanol	12	Carbon tetrachloride	20	2-methoxy ethanol
5	Isoamyl alcohol	13	Fluoro toluene	21	1,2-dimethoxyethane
6	2-methyl-1-propanol	14	Cyclohexane	22	N,N-dimethyl formamide
7	Isopropyl acetate	15	Diethyl carbonate	23	N,N-dimethyl acetamide
8	Isobutyl acetate	16	Diisopropyl ether	24	Dimethyl sulfoxide

#### 4.2.4 Preparation of 96-well Plate for Physical Form Screening of Olanzapine

Physical form screening of OZPN was carried out using two concentrations of solute and 48 solvents at RT in a 96-well plate. A solution of OZPN (50 mM) in ethyl acetate was prepared prior to use. 70  $\mu$ L and 140  $\mu$ L of OZPN solution were placed in wells 1-48 (A1-D12) and 49-96 (E1-H12) respectively. The MWP was then left in the fume hood at RT for evaporation of the ethyl acetate to leave approximately 1.1 and 2.2 mg of solid in wells 1-48 (A1-D12) and 49-96 (E1-H12) respectively after drying. 200  $\mu$ L of each of the crystallisation solvent was added into the first 48 wells (A1 to D12) in the same sequence as listed in Table 4.5.

Table 4.5. Solvents used for physical form screening of olanzapine using 96-well plate

S. No	Solvent	S. No	Solvent	S. No	Solvent
1	Methanol	17	Tetrahydrofuran	33	2,2,2-trifluoroethanol
2	Ethanol	18	Pyridine	34	N,N,N-triethyl amine
3	1-propanol	19	Diethyl carbonate	35	Aniline
4	2-propanol	20	Diethyl ether	36	Hexane
5	1-butanol	21	Tertiary-butyl methyl ether	37	N,N-dimethyl formamide
6	2-butanol	22	Acetone	38	N,N-dimethyl acetamide
7	1-pentanol	23	Methyl-ethyl ketone	39	2-butoxy ethanol
8	2-pentanol	24	Toluene	40	2-methoxy ethanol
9	Methyl acetate	25	Fluoro toluene	41	Dimethyl sulfoxide
10	Ethyl acetate	26	Isoamyl alcohol	42	N-methyl pyrrolidone
11	Butyl acetate	27	Carbon tetrachloride	43	2-phenyl ethanol
12	Isobutyl acetate	28	Dichloromethane	44	1-methyl naphthalene
13	Pentyl acetate	29	1,2-dichloroethane	45	1,2-dimethoxyethane
14	1,4-dioxane	30	Tetrachloro ethylene	46	Formic acid
15	Acetonitrile	31	1-chlorobutane	47	Acetic acid
16	Nitro methane	32	1-bromobutane	48	Water

The solvent addition step was repeated for wells E1-H12. The MWP was then sealed using transparent sealing film (ThermaSeal RTS™) and placed on an orbital shaker for 2 hours to redissolve the OZPN. The plate was visually inspected to check for complete dissolution of OZPN. Next, the seal was removed and the MWP was left in the fume hood at RT to allow slow evaporation of the solvents. The plate was inspected regularly for evidence of crystallisation. Following appearance of the crystals, data were collected using the Raman microscope (see Section 4.2.5).

#### **4.2.5 Raman Microscopy**

Raman spectra were collected using a DXR Raman microscope (ThermoScientific, Madison, USA) equipped with a high precision x, y, z motorised stage with a holder for MWP (Figure 4.5). The instrument employs a diode-pumped laser (532 nm) or frequency-stabilised single mode diode laser (780 nm) as an excitation source and an air cooled CCD detector. The maximum power of laser (10 mW) was used for data collection.

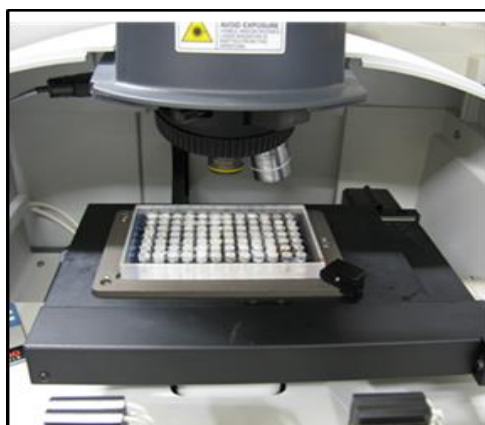


Figure 4.5. DXR Raman microscope with motorised x, y, z stage and multi-well plate reader.

There are various data collection modes available, including (a) an average of multiple spectra collected from the well, (b) a spectrum collected from the centre of the well, (c) multiple spectra collected from a grid of defined area and step size across the well, (d) manual assignment of individual data collection locations, and (e) an automated search for the strongest Raman signal within the well available for data acquisition. Optical images are also automatically acquired for each data collection location to allow confirmation that spectra were collected from representative material.

Spectra were collected using a long working distance 10x objective with 50  $\mu\text{m}$  aperture, 1s exposure and a spectral resolution of 5  $\text{cm}^{-1}$ . An autofocus option was employed for collecting spectra to optimise the signal to noise ratio from the sample and fluorescence correction was applied during data collection. The OMNIC™ Array Automation program (version 8.2.1) (Thermo Scientific) within OMNIC™ software (version 8.2) (Thermo Scientific) was used to define the data collection strategy. A minimum of 4 positions per well were assigned manually for analysis by inspecting each well for various crystal habits. A manual approach was used for selecting points for data collection due to the non-uniform distribution of crystals at the bottom of the wells. This ensured that Raman spectra from different morphologies would be analysed maximising data quality albeit at the expense of time required to select individual data collection locations. By collecting multiple spectra from each well the likelihood of detecting a mixture of forms where the minor component is only a low percentage of the bulk is increased. An example of the real time data collection

interface is shown in Figure 4.6. Once data were acquired chemometric analysis was applied using statistical techniques (see Section 4.2.6).

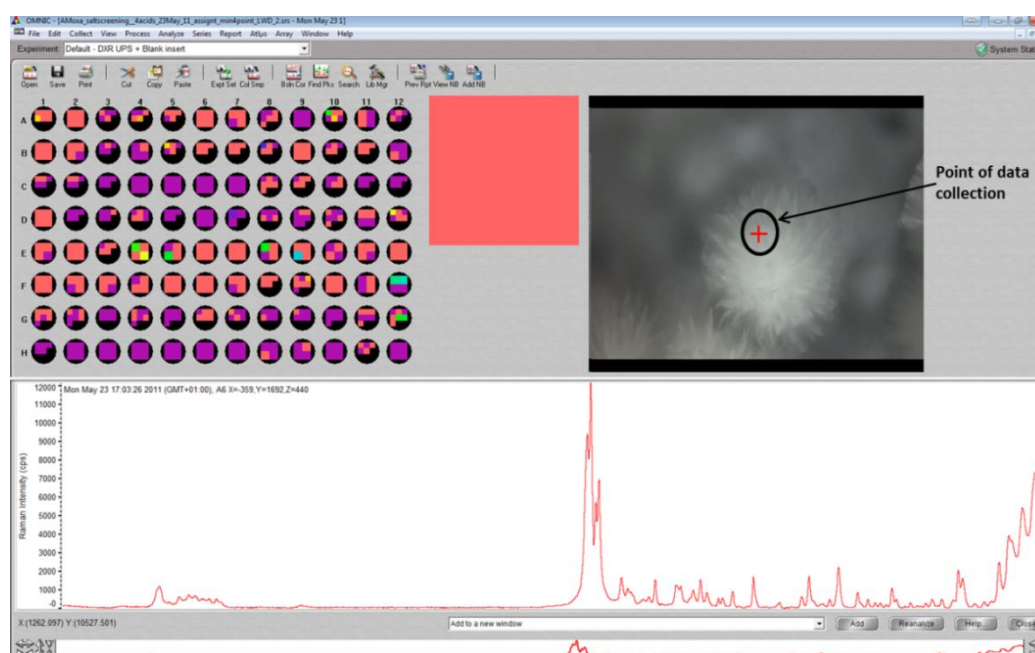


Figure 4.6. A screenshot of the data acquisition interface set-up for use with the bespoke 96-well quartz plate. Each circle in the 8 x 12 grid (top left) represents a well on the plate and each square within a well represents a point of data collection (i.e. a Raman spectrum). The point (spot size-1  $\mu\text{m}$ ) from where the spectrum is being collected is shown by a plus sign on the image (top right) with the respective Raman spectrum shown (bottom). Spectra in the grid are coloured based on the magnitude of the measured Raman intensity. Red and violet colours indicate the highest and lowest intensity respectively.

#### 4.2.6 Chemometric Analysis

Pre-processing of each spectrum was done by using a Norris derivative (Norris and Williams, 1984) for baseline correction which smooths and performs the derivatisation in a single step within the Array Automation program (version



8.2.1) of the OMNIC™ software package (version 8.2). This treatment removes artefacts, variation in the spectra due to the presentation of crystals in each well (orientation effect) and baseline variations arising due to fluorescence or residual laser scatter (O'Connell et al., 2005). After preprocessing, spectra were clustered using the unsupervised group analysis technique in the region of  $\sim 1400\text{-}1700\text{ cm}^{-1}$  of Raman shift, with a similarity cut off value of 90% within array program. This region of Raman shift was used because of the presence of bands with high signal to noise ratio for these compounds. Various characteristic Raman bands for organic compounds within the Raman shift region of  $1400\text{-}1700\text{ cm}^{-1}$  are listed in Table 4.6.

Table 4.6. Raman bands within the region of  $1400\text{-}1700\text{ cm}^{-1}$  of Raman shift

Functional Group / Vibration	Region ( $\text{cm}^{-1}$ )	Raman
$\nu(\text{C}=\text{N})$	1610-1680	Strong
$\delta(\text{CH}_2)$ $\delta(\text{CH}_3)$ asym	1410-1470	Medium
$\nu(\text{C}-\text{C})$ aromatic ring chain	1580-1600	Strong
vibration	1450-1500	Medium

Being an unsupervised technique, group analysis makes no assumption about the number of classes or class assignment before the analysis (Kastanos et al., 2012). It compares all the spectra in the data set and sorts the spectra into groups or clusters based on their "similarity". The similarity value for a group of spectra is a function of the cross-correlation coefficient between all the spectra in the set. After group analysis, the colour assigned to each well within MWP

corresponds to the groups assigned as the result of the spectral data analysis enabling easy detection of different single or multiple forms (Lowry et al., 2006).

#### **4.2.7 Scale-up and Characterisation of Novel Forms**

Novel forms identified during each HTCAA study were scaled-up to a 100-200 mg scale by evaporative solution crystallisation in a 3-5 ml vial to allow further characterisation of each new solid form.

Novel salts forms of AXPN were prepared using the same conditions (solvent and molar ratio of AXPN and counter acid) as identified from the MWP screen. For OZPN and CZPN (only PS-1), evaporative crystallisations in vials (see Section 3.2.1.1) were carried out with all the solvents used during HTCAA to compare with the results obtained at smaller scale. In some cases, liquid-assisted grinding using Retsch MM400 mixer mill (see Section 3.2.1.3) and cooling crystallisations (see Section 3.2.1.2) were also used to confirm the results of evaporative solution crystallisation or obtain the pure form. All the novel forms were characterised using Raman microscopy (see Sections 3.2.4 and 4.2.5), XRPD, SXD where suitable samples could be obtained (see Section 3.2.2) and thermal analysis using simultaneous DSC and TGA (see Section 3.2.3).

## 4.3 Results and Discussion

### 4.3.1 Salt Screening of Amoxapine

Out of the 96 recrystallisation experiments on the 96-well plate (~1.4-3.6 mg solid per well), crystals were observed in 91 wells (hit rate of 94.8%) as indicated by polarised light microscopy. In other cases, a sticky/oily material was formed which could not be analysed further.

Spectra obtained from the 96-well plate were grouped using the group analysis tool in the region of 1470-1700  $\text{cm}^{-1}$  of Raman shift with a cut off value of 90% for similarity. 466 spectra were grouped into a total of 11 groups (Figure 4.7, Table 4.7) and the remaining 171 spectra were outliers and did not fall into any of the groups. These ungrouped spectra were analysed individually (Figure 4.7).

Table 4.7. Details of 11 groups observed on the 96-well plate and their respective colour shown in Figure 4.6

Group	Counter Anion	Colour in Figure 4.6
1	Succinic acid, maleic acid and acetic acid	Blue
2	Succinic acid	Fluorescent green
3	Succinic acid	Dark orange
4	Maleic acid	Cyan
5	PTSA	Green
6	PTSA	Yellow
7	Maleic acid	Orange
8	PTSA	Dark yellow
9	Maleic acid	Light green
10	Maleic acid and acetic acid	Pink-red
11	NA-only AXP	Violet

Acid (AXPN:Acid)	Methanol	Ethanol	Acetonitrile	Acetone	Ethyl acetate	Isobutyl acetate	Nitromethane	Cyclohexane	Toluene	Diisopropylether	Tetrahydrofuran	Water	
Succinic acid (1:1)	A	1	2	3	4	5	6	7	8	9	10	11	12
Succinic acid (2:1)	B												
Maleic acid (1:1)	C												
Maleic acid (2:1)	D												
PTSA (1:1)	E												
PTSA (2:1)	F												
Acetic acid (1:1)	G												
Amoxapine	H												

Figure 4.7. Salt screening results of amoxapine after group analysis on a 96-well plate. Each square (spectrum) from the wells is colour coded according to classification group. Group 1 to 11 are shown by blue, fluorescent green, dark orange, cyan, green, yellow, orange, dark yellow, light green, pink-red and violet colour respectively. All other spectra shown by violet colour are outliers.

Out of the 11 groups, groups 1-6 represent novel salts and group 7 represents poor quality spectra with no peak information. The spectra from groups 8-10 were similar to one of the novel salts identified in groups 1-6 but were different enough (i.e. poor Raman scattering band resolution) for the group analysis to classify accurately. In addition to sample presentation effects highlighted above (see Section 4.2.6), minor changes in Raman spectra can also arise from differences in particle size, purity and residual solvent (Remenar et al., 2003).

The spectra in group 11 match that of the AXPN starting material indicating that no polymorphs or solvates are obtained with the studied solvents. Two novel

salts of AXPN with acetic acid were obtained after analysing the spectra which could not be grouped in any of the 11 groups. The remainder of the ungrouped spectra were confirmed as belonging to one of the above groups (1-11) but were misclassified due to the variations in peak resolution or poor signal to noise ratio. The Raman spectra in the region of 1520-1640  $\text{cm}^{-1}$  of Raman shift for all 10 novel salts are shown in Figure 4.8.

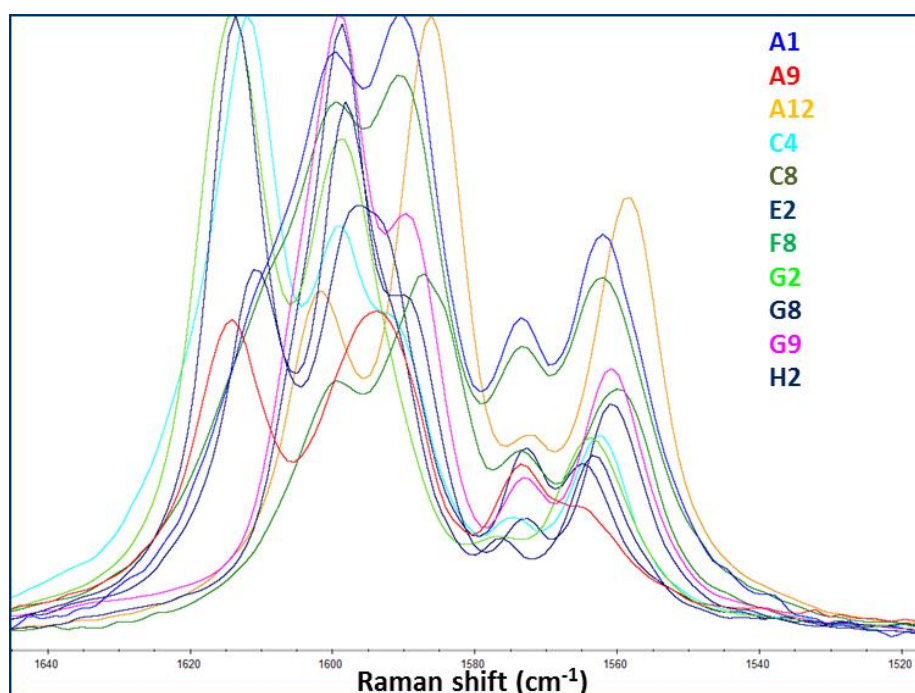


Figure 4.8. Characteristic Raman spectra for 10 novel salts observed during salt screening of amoxapine. Each spectrum is labelled according to the well number from which they have been obtained on the 96-well plate.

Some maleic, succinic and acetic acid counter ion spectra were grouped together (group 1, shown by blue colour in Figure 4.7) as these structures display similar peak positions in the 1400-1700  $\text{cm}^{-1}$  region. On comparing the full spectra of these different salts, the differences were clearly evident from other regions of Raman spectra. Thus, a narrow wavelength region is sufficient

to allow rapid and effective cluster analysis. A full range spectrum contains more spectral information but can lead to less effective clustering due to the increased information content.

The total number of AXPN salts identified with each counter acid is reported in Table 4.8. Salts in the binary mixtures of AXPN and counter acid exhibiting different spectra are labelled according to the initials of the counter acid (Table 4.8 and Table 4.9).

Table 4.8. Total number amoxapine salts identified from each counter acid during HTCAA.

<b>Group</b>	<b>Counter Acid</b>	<b>Number of Crystalline Forms</b>	<b>Label</b>
1	Succinic acid	3	S1, S2 and S3
2	Maleic acid	2	M1 and M2
3	PTSA	2	E1 and E2
4	Acetic acid	3	AA1, AA2 and AA3

Details of the crystallisation conditions (counter acid, stoichiometry and solvent) under which the individual salt was obtained are given in Table 4.9. A mixture of AXPN with its succinate (S1), maleate (M1 or M2) and acetate (AA1, AA2 and AA3) salts was detected in 1, 4 and 4 wells respectively. A mixture of maleate (M1 and M2) salts was observed in 4 wells. Similarly, a mixture of tosylate (E1 and E2) salts was detected in 6 wells (Table 4.9).

Table 4.9. Amoxapine salts classified by Raman spectra using group analysis on the results obtained in the screening experiment on the 96 well plate.

		Solvents																
		1	2	3	4	5	6	7	8	9	10	11	12					
CA	AXPN:CA	MeOH	EtOH	MeCN	Acetone	EtOAc	IBA	NM	cHex	Toluene	DIPE	THF	Water					
Succinic acid	1:1	S1	S1	S1	S1	S1	S1	S1	S1	S1	S1	---	S3					
Succinic acid	2:1	S1	S1	---	S1	S1	S1	S1	S1	S1	S1	S1	S3					
Maleic acid	1:1	M1	M1	M2	M1	M1	M1	M1	M1	M2	M2	M1	M2	M1	M1	M1		
Maleic acid	2:1	M2	M2	M2	AXPN	M2	M1	M1	M1	M2	AXPN	AXPN	M1	M2	AXPN	M2	M2	M1
PTSA	1:1	E1	E1	E1	E1	E1	E1	E2	E1	E1	E1	E1	E1	E1	E1	E1	E1	
PTSA	2:1	E1	E1	E2	E1	E2	E1	E1	E2	E1	E2	E1	E2	E1	E2	---	E1	
Acetic acid	1:1	AA1	AA1	AXPN	AA1	AXPN	AXPN	AA1	AXPN	AA1	AA2	AXPN	AA3	AXPN	AXPN	---	AA1	
AXPN		AXPN	AXPN	AXPN	AXPN	AXPN	AXPN	AXPN	AXPN	AXPN	AXPN	AXPN	AXPN	AXPN	AXPN	---	AXPN	

Abbreviations used in table: CA-counter acid, AXPN-amoxapine, MeOH-methanol, EtOH-ethanol, MeCN-acetonitrile, EtOAc-ethyl acetate, IBA-Isobutyl acetate, NM-nitromethane, cHex-cyclohexane, DIPE-diisopropyl ether, THF-tetrahydrofuran.

Raman spectra of salts from AXPN and succinic acid (1:1, 2:1) are classified as S1 or S2 or S3,

Raman spectra of salts from AXPN and Maleic acid (1:1, 2:1) are denoted as M1 or M2,

Raman spectra of salts from AXPN and PTSA (1:1, 2:1) are represented as E1 or E2,

Raman spectra of salts from AXPN and acetic acid (1:1) are classified as AA1 or AA2 or AA3.

In total, 10 novel salts were observed using the 4 acids in molar ratios (AXPN:counter acid) of 1:1 and 2:1 utilising a minimum amount of material i.e. ~187 mg of AXP, ~50 mg of the 4 counter acids and ~20 ml of the 12 solvents. With appropriate data treatment and grouping techniques, a variety of chemical information was obtained. For example, two potential solid-state forms were observed for maleate as well as tosylate salts. Similarly three potential solid-state forms were detected for both succinate and acetate salts. However, it is not clear if the salts were hydrates, solvates or a combination. In addition, the same crystalline forms were observed in the wells with 1:1 and 2:1 molar ratio of AXP and counter acid.

The presence of both M1 and M2 in wells with 1:1 and 2:1 molar ratio of AXP and maleic acid and AXP free base in some wells with 2:1 molar ratio of AXP and maleic acid suggest that the AXP tends to form salts with maleic acid in a molar ratio of 1:1. The identity of the novel salts was confirmed using SXD/XRPD and STA analyses during scale-up experiments (see Section 4.3.1.1).

#### ***4.3.1.1 Scale-up and Characterisation of Novel Salts of Amoxapine***

##### ***4.3.1.1.1 Succinate and Acetate Salt Forms of Amoxapine***

All the succinate (S1, S2 and S3) and acetate (AA1, AA2 and AA3) salt forms could be obtained by solvent evaporation from the respective solvents as identified during the screen. An overlay of Raman spectra of the succinate and acetate salt forms observed from the 96 well plate and after scale-up is shown in Figure 4.9.



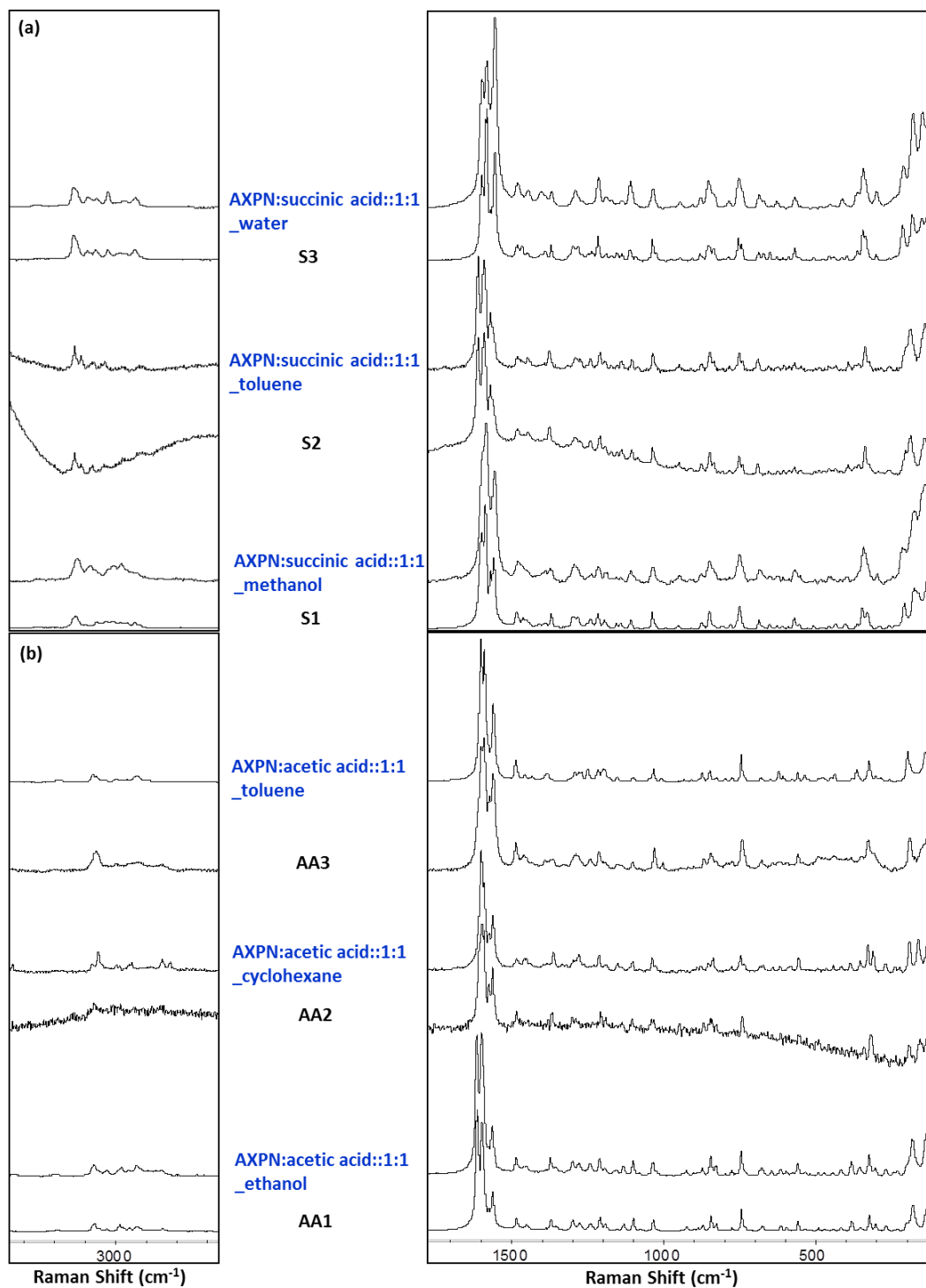


Figure 4.9. Stack plots of Raman spectra collected from samples of salts of (a) amoxapine and succinic acid, (b) amoxapine and acetic acid, at 96-well plate and on 150 mg scales. The Raman spectra of salts obtained from 96-well plate and on 150 mg scales are labelled in black and blue colour respectively.

An overlay of XRPD patterns of the novel salts including the three different succinate and acetate salt forms obtained after scale-up is shown in Figure 4.10.

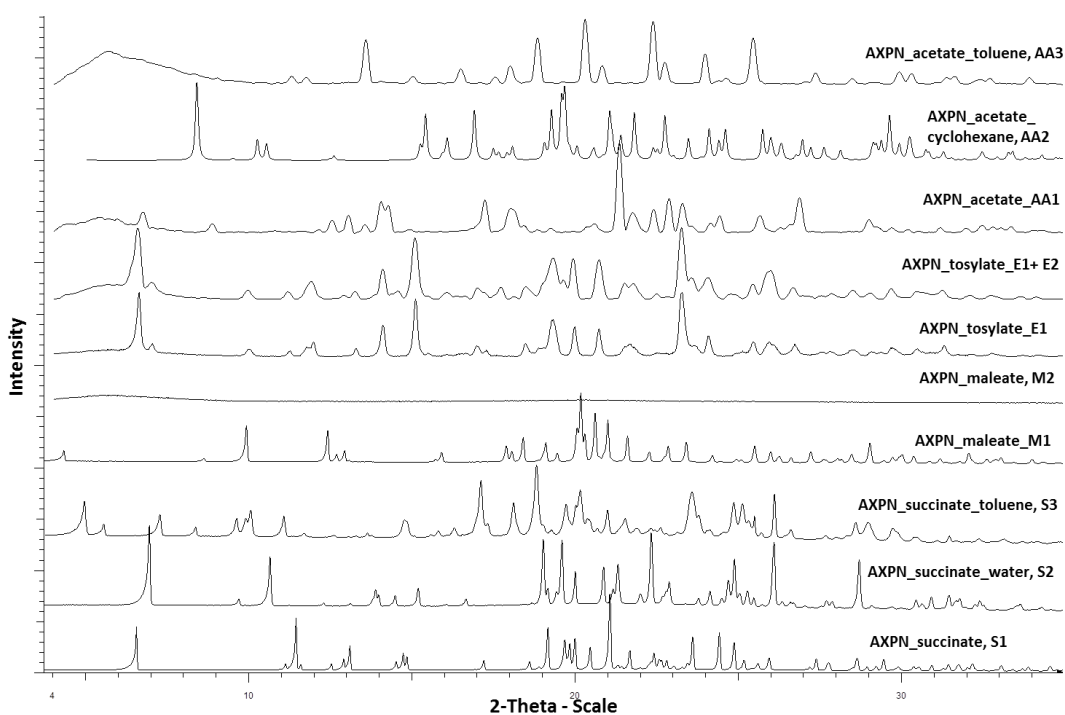


Figure 4.10. Stack plot of XRPD patterns of the 10 novel amoxapine salts obtained on 150 mg scales in the range of 4-35° 2 $\theta$ .

S1, AA1, AA2 and AA3 form a part of a group of six salts, for which single crystal data were also collected and lattice parameters are reported in Table 4.10. Structural analysis was not the key goal of this study; however a brief overview of key structural features of these 6 salts is presented in Table 4.11.

Table 4.10. Lattice parameters of the amoxapine salts identified from the salt screen. All parameters were determined from single-crystal diffraction data collected at 123(2) K (see Section 3.2.2).

Structure ID	Identifier	Space Group	$a(\text{\AA})$	$b(\text{\AA})$	$c(\text{\AA})$	$\beta$ (°)	Volume ( $\text{\AA}^3$ )
1	AXPN_succinate, S1	<i>Pbcn</i>	24.2941(11)	21.0763(11)	8.5064(4)	90.00	4355.5(4)
2	AXPN_maleate, M1	<i>P2<sub>1</sub>/c</i>	20.6004(9)	9.8341(4)	9.7939(4)	96.779(1)	1970.2(1)
3	AXPN_PTSA, E1	<i>C2/c</i>	31.5462(13)	5.8832(3)	27.6448(12)	119.299(2)	4474.3(4)
4	AXPN_acetate, AA1	<i>P2<sub>1</sub>/n</i>	14.5022(13)	7.8015(7)	19.0516(17)	105.906(4)	2072.9(3)
5	AXPN_acetate_cyclohexane, AA2	<i>P2<sub>1</sub>/c</i>	21.0726(12)	6.0393(3)	18.6087(10)	92.096(2)	2366.6(2)
6	AXPN_acetate_toluene, AA3	<i>P2<sub>1</sub>/c</i>	13.5397(5)	7.8076(2)	19.5838(7)	97.089(2)	2054.4(1)

Table 4.11. Content of asymmetric unit of six novel salts of amoxapine reported in Table 4.10.

Identifier		Asymmetric unit
S1	AXPN_succinate	$[\text{H-AXPN}]_2^+ [\text{succinate}]_2^- [\text{succinic acid}]_2 [\text{H}_2\text{O}]_2$
AA1	AXPN_acetate	$[\text{H-AXPN}]^+ [\text{acetate}]^- [\text{acetic acid}]$
AA2	AXPN_acetate_cyclohexane	$[\text{H-AXPN}]_2^+ [\text{acetate}]_2^- [\text{acetic acid}]_2 [\text{cyclohexane}]$
AA3	AXPN_acetate_toluene	$[\text{H-AXPN}]^+ [\text{acetate}]^- [\text{toluene}]$
M1	AXPN_maleate	$[\text{H-AXPN}]^+ [\text{H-maleate}]^-$
E1	AXPN_PTSA	$[\text{H-AXPN}]^+ [\text{tosylate}]^-$

#### 4.3.1.1.2 Maleate Salts of Amoxapine

M1 is a non-solvated 1:1 maleate salt (as indicated by SXD data and Thermal analysis) of AXPN and was obtained from solvent evaporation from methanol at RT. The stoichiometry information obtained here corroborates the inference of salt screening on the 96-well plate. M2 was observed concurrently with M1 from solvent evaporation of AXPN and maleic acid in a molar ratio of 2:1 from methanol or ethanol. The XRPD pattern of M2 however revealed that it is non-crystalline or diffraction amorphous (Figure 4.10). To confirm this, cooling crystallisations and liquid-assisted grinding using ethanol and methanol were carried out. XRPD data from these samples were consistent with samples comprising M1 and AXPN. No evidence of diffraction amorphous samples was found from these methods. The appearance of M2 could be due to rapid onset of supersaturation, resulting in oiling out or rapid precipitation of an amorphous

solid of AXPN maleate salt (Morissette et al., 2004). This is supported by the morphology analysis of M2 during HTCAA as visual inspection of the morphology indicates that M2 consists of poor quality crystals (Figure 4.11).

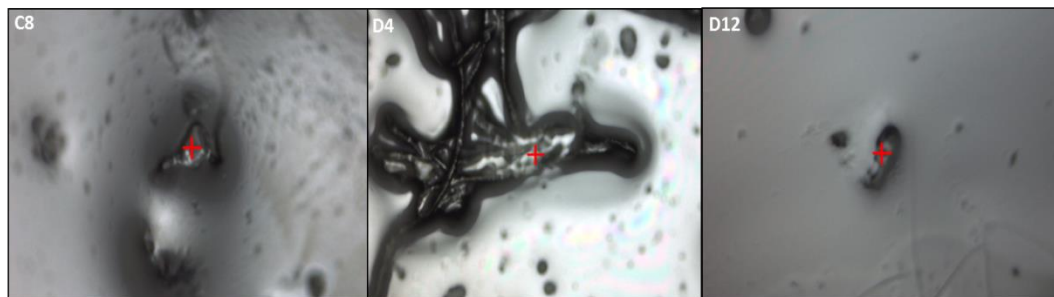


Figure 4.11. Microscope images (20x) of the M2 crystals observed in wells C8, D4 and D12 of 96-well plate.

#### ***4.3.1.1.3 Tosylate salts of Amoxapine***

E1 and E2 were obtained in the initial screen as a mixture from solvent evaporation of a 1:1 molar ratio of AXPN and PTSA in cyclohexane at RT. E1, E2 and AXPN were obtained from solvent evaporation from a 2:1 molar ratio of AXPN and PTSA in cyclohexane. SXD analysis revealed that E1 is a 1:1 non-solvated tosylate salt of AXPN. Only pure E2 was obtained on 96-well plate from nitromethane and cyclohexane with 1:1 and 2:1 molar ratios of AXPN and PTSA. But in vials, E2 was always found in the presence of E1 or AXPN. This suggests that E1 or AXPN might have been missed by the Raman analysis in these wells during data collection. Alternatively, this change (i.e. obtaining a mixed phase) is due to the effect of the change in solvent volume (Aaltonen et al., 2009). Another reason might be the lack of control over the nucleation and growth process in the small scale evaporation trials resulting in varying rates of supersaturation,

hence increasing the possibility of formation of a mixed phase sample. E1 and E2 were obtained as a mixture from 1:1 molar ratios of AXP and PTSA, however, when a 2:1 molar ratio of AXP and PTSA was used, excess AXP was observed suggesting these two forms are polymorphic rather than different stoichiometric salts. Overlays of the Raman spectra of the maleate and tosylate salts are shown in Figure 4.12.

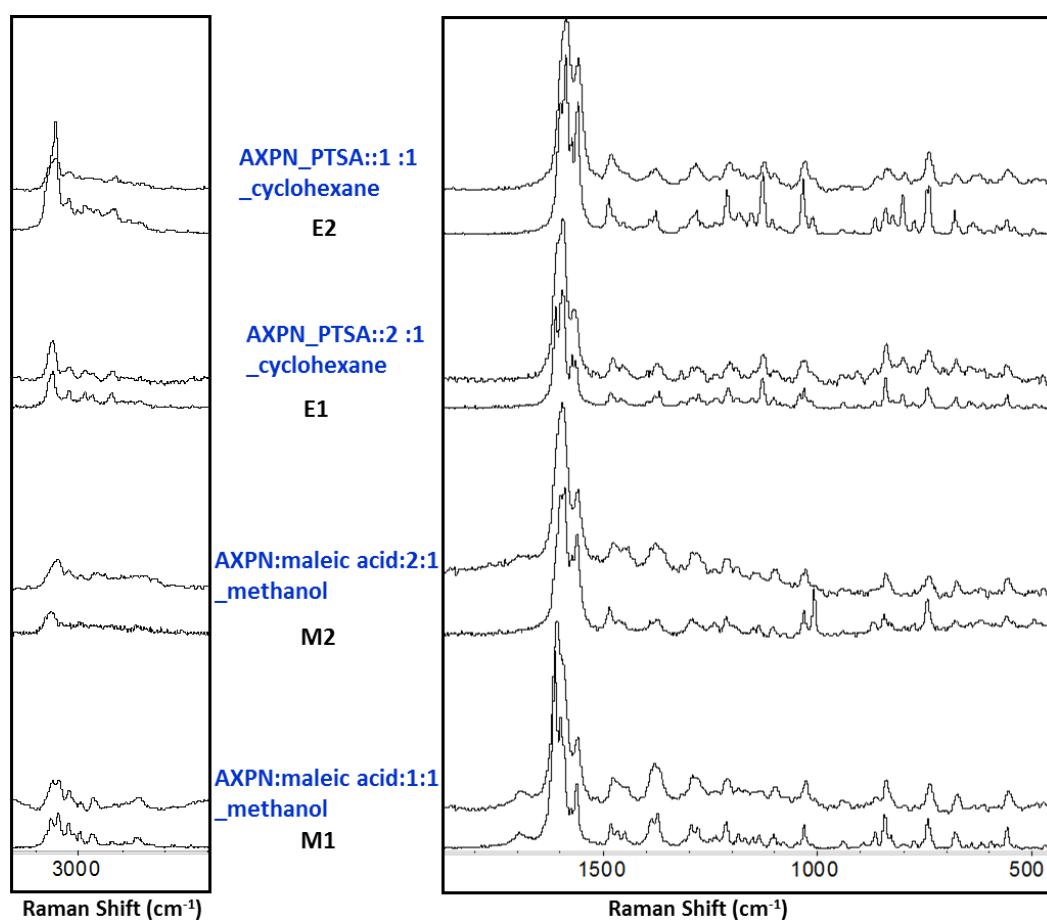


Figure 4.12. Stack plots of Raman spectra collected from samples of salts of (a) amoxapine and maleic acid and (b) amoxapine and para-toluene sulfonic acid at 96-well plate and on 150 mg scales. The Raman spectra of salts obtained from 96-well plate and on 150 mg scales are labelled in black and blue respectively.

To summarise, 10 novel AXPN salts were observed from the novel HTCAA method developed and were successfully scaled-up to mg quantities. The crystal structures of 6 salts were determined using SXD (Table 4.10 and Table 4.11) and the remaining 4 salts were characterised using XRPD, Raman and thermal analysis. Further work would be required to obtain suitable samples to enable structure determination of these four salts. SDPD was not attempted for these 4 salts though has been used in Chapter 7 in this work.

## 4.3.2 Physical Form Screening of Clozapine

### 4.3.2.1 Experiment 1, PS-1

Following the method outlined in section 4.2.3, crystalline solid was obtained in 38 out of 48 wells (hit rate=79.2%) as indicated by polarised light microscopy (Table 4.12). Raman spectra obtained from all the samples on the plate were grouped by employing group analysis tool (see Section 4.2.6) using data in the region of 1495-1678  $\text{cm}^{-1}$ . 229 spectra were grouped into 5 groups (Figure 4.13) and the remaining 66 spectra were outliers and did not fall into any of the groups (Figure 4.13). These ungrouped spectra were analysed individually. Spectra from group 1 (Figure 4.13) were present in 34 wells and represent the starting form of CZPN (Petcher and Weber, 1976), designated as form A. Spectra in groups 2 and 3 were present in 11 and 18 wells respectively. Spectra from group 4 were only found in single well. The physical forms representing group 2-4 have been designated as form B, C and D respectively. Visual inspection of the recrystallised material in all the wells in group 3 suggested that it was oily or poorly crystalline. Spectra from group 5 (well C3 and E8) did not produce any spectral information. A mixture of physical forms was observed in 14 wells. Ungrouped spectra either do not have any spectral information or fall into one of the groups (1-5) but were misclassified due to variations in peak resolution or poor signal to noise ratio in the experimental spectra.



Table 4.12. Solvent systems used in physical form screening, PS-1 of clozapine using a 48-well plate. Wells where no crystals were observed are coloured green. Group analysis of all the Raman spectra is shown in Figure 4.13.

	1	2	3	4	5	6	7	8
A	MT	ET	PP	PP-2	BT	2-butanol	MA	EA
B	PA	BA	Acetone	MEK	DEE	TBMe	Toluene	DCE
C	CH <sub>3</sub> CN	DXN	Pyridine	NM	THF	DCM	Water	AA
D	MT + water	ET + water	PP + water	PP-2 + water	BT + water	2-butanol + water	MA + water	EA + water
E	PA + water	BA + water	Acetone + water	MEK + water	DEE + water	TBMe + water	Toluene + water	DCE + water
F	CH <sub>3</sub> CN + water	DXN + water	Pyridine + water	NM + water	THF + water	DCM + water	Water + water	AA + water

Abbreviations: MT-methanol, ET-ethanol, PP-1-propanol, PP2-2-propanol, BT-1-butanol, MA-methyl acetate, EA-ethyl acetate, PA-propyl acetate, BA-butyl acetate, MEK-methyl ethyl ketone, DEE-diethyl ether, TBMe-t-butyl methyl ether, DCE-1,1-dichloroethane, CH<sub>3</sub>CN-acetonitrile, DXN-1,4-dioxane, NM-nitromethane, THF-tetrahydrofuran, DCM-1,2-dichloroethane, AA-acetic acid

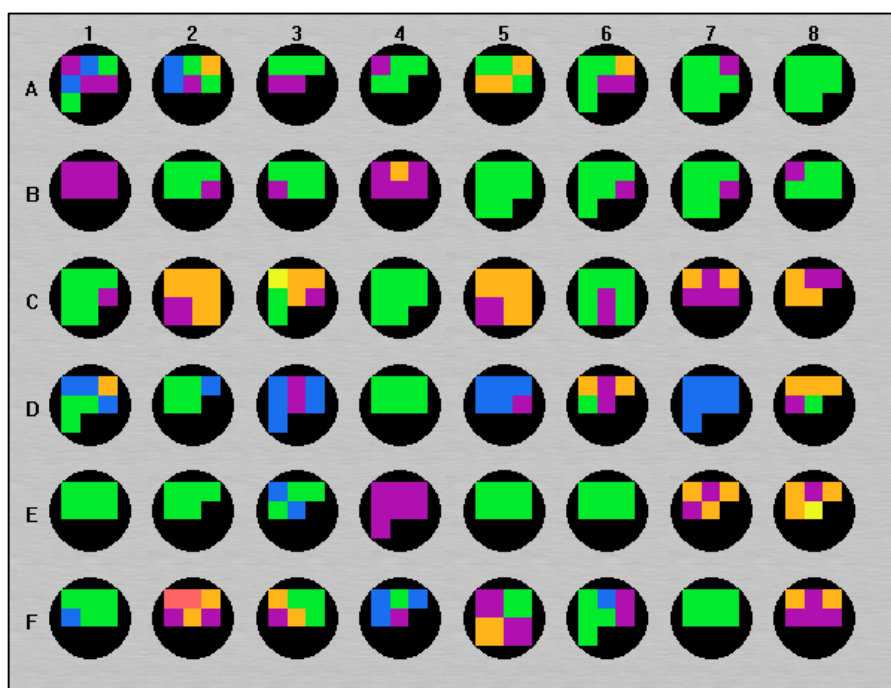


Figure 4.13. Physical form screening (PS-1) results of clozapine after group analysis on the 48-well plate using the conditions shown in Table 4.12. The wells are colour coded according to classification group. Group 1, 2, 3, 4 and 5 are shown by green, blue, orange, red and yellow colour, respectively. All other spectra shown by magenta colour are outliers.

An overlay of the Raman spectra representative of physical forms A, B, C and D in the region of 1480-1680  $\text{cm}^{-1}$  is shown in Figure 4.14.

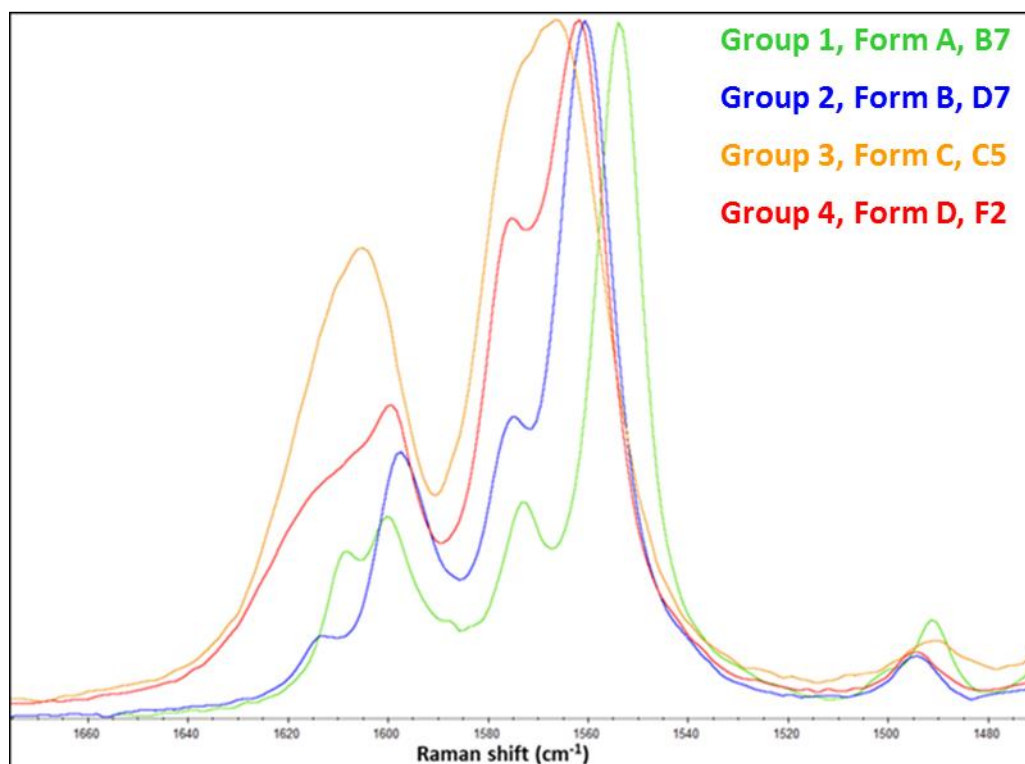


Figure 4.14. Overlay of Raman spectra in the region of 1450-1700  $\text{cm}^{-1}$  of Raman shift for four physical forms (A, B, C and D) observed during physical form screening of clozapine.

#### 4.3.2.2 Experiment 2, PS-2

A second plate for physical form screening (PS-2) was prepared following the method outlined in section 4.2.3, crystalline solid was obtained in 42 wells on the 48-well plate (hit rate = 87.5%) as indicated by polarised light microscopy (PS-2; Table 4.13). Raman spectra obtained from all the samples on the plate were grouped using the group analysis tool in the region of 1495-1678  $\text{cm}^{-1}$ .

Table 4.13. Solvent systems used in physical form screening, PS-2 of clozapine using a 48-well plate. Wells where no crystals were observed (based on visual inspection) are coloured pink. Group analysis of all the Raman spectra is shown in Figure 4.15.

	1	2	3	4	5	6	7	8
A	PNT	PNT-2	HN	CH	IAA	MPN	IPA	IBA
B	PA	TCE	CHCl <sub>3</sub>	CCL <sub>4</sub>	FT	CH	DEC	DIPE
C	MP	EXE	TFE	MXE	DME	DMF	DMA	DMSO
D	PNT + water	PNT-2 + water	HN+ water	CH + water	IAA + water	MPN + water	IPA + water	IBA + water
E	PA + water	TCE + water	CHCl <sub>3</sub> + water	CCL <sub>4</sub> + water	FT + water	CH + water	DEC + water	DIPE + water
F	MP + water	EXE + water	TFE + water	MXE + water	DME + water	DMF + water	DMA+ water	DMSO + water

Abbreviations: PNT-1-pentanol, PNT-2, 2-pentanol, HN-1-hexanol, CH- cyclohexanol, IAA-isoamyl alcohol, MPN-2-methyl-1-propanol, IPA-isopropyl acetate, IBA-isobutyl acetate, PA-pentyl acetate, TCE-trichloroethylene, CHCl<sub>3</sub>-chloroform, CCL<sub>4</sub>-carbon tetrachloride, FT-4-fluorotoluene, CH- cyclohexane, DEC-diethyl carbonate, DIPE-diisopropyl ether, MP- 4-methyl-2-pentanone, EXE-2-ethoxyethanol, TFE-2,2,2-trifluoroethanol, MXE-2-methoxyethanol, DME-1,2-dimethoxyethane, DMF-N,N-dimethylformamide, DMA-N,N -dimethylacetamide, DMSO-dimethylsulfoxide

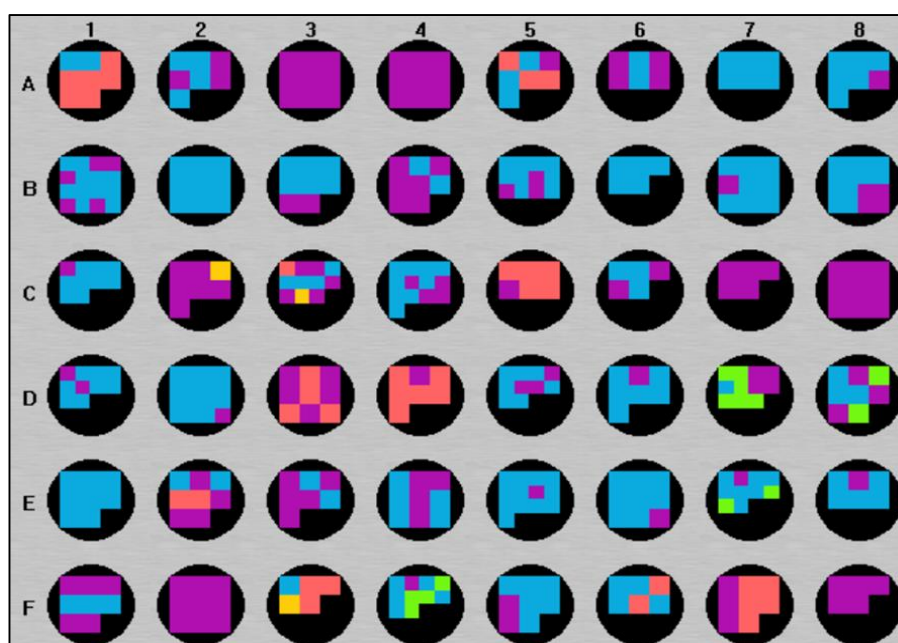


Figure 4.15. Physical form screening (PS-2) results of clozapine after group analysis on the 48-well plate using the conditions shown in Table 4.13. The wells are colour coded according to classification group. Group 1, 2, 3 and 4 are shown by blue, green, orange and yellow colour, respectively. All other spectra shown by magenta colour are outliers.

270 spectra were grouped into 4 groups and the remaining 120 spectra were outliers and did not fall into any of the four groups (Figure 4.15). These ungrouped spectra were analysed individually to confirm their identity. Samples exhibiting spectra that belong to group 1 were present in 37 wells (Figure 4.15) and relate to the CZPN starting material. Spectra from group 2 and 3 were obtained from samples in 4 and 10 wells respectively (Figure 4.15) and represent form B and C of CZPN, as observed from PS-1. Group 4 represents samples that returned no spectral information. PS-2 also yielded the same physical forms (form A, B, and C) as were observed in PS-1.

To summarise, two HTCAA experiments encompassing 96 crystallisations, utilising a minimum amount of material i.e. ~235 mg of solute and ~24 ml of 96 solvent systems yielded 3 physical forms of CZPN. On the basis of these results, CZPN appears to have a more limited range of solid-state diversity as compared to OZPN (studied in Section 4.3.3).

#### ***4.3.2.3 Characterisation of Novel Forms of Clozapine after Scale-up, PS-1***

Each of the three forms were successfully scaled up and fully characterised. Form A and B were obtained from the solvent systems as identified during the HTCAA screen. Crystal structure determination of form B revealed that it is a monohydrate of CZPN (Siva Lakshmi Devi et al., 2011).

Form C that was present in 18 wells was an oily or poorly crystalline material. On scale-up, it was obtained from acetic acid, 1,4-dioxane, methyl ethyl ketone and 1,4-dioxane:water (1:1) and is characterised as an amorphous phase based

on the lack of diffraction observed from XRPD. Other solvent systems which yielded form C on a 48-well plate provided either form A or B or mixture of both on scale-up.

A stack plot of the characteristic Raman spectra of forms B, C, and D obtained from the screen and scale-up experiments are shown in Figure 4.16. XRPD patterns of the anhydrous, monohydrate and non-crystalline phases of CZPN are shown in Figure 4.17.

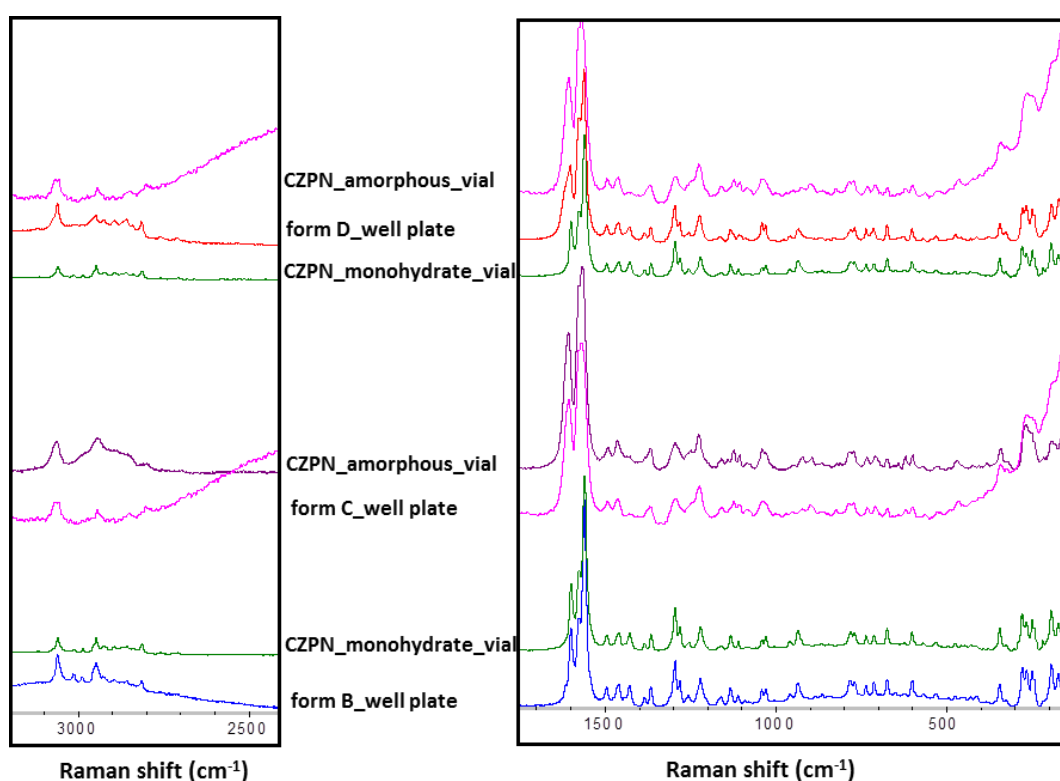


Figure 4.16. Stack plot of the Raman spectra collected from samples of physical forms of clozapine observed on the 48-well plate and on 150 mg scales. The Raman spectra are labelled according to the source of their origin.

The Raman spectrum of form D (obtained from well F2 during PS-1 from 1:1 mixture of 1,4-dioxane and water) appeared very similar to that of form B. This was confirmed by merger of groups 2 and 4 into one group by decreasing the

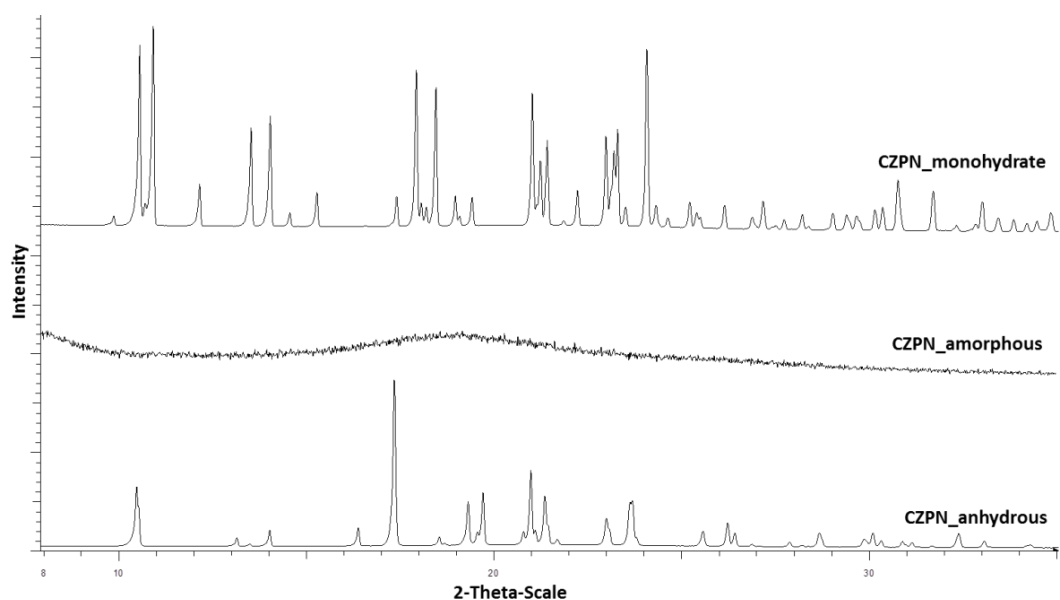


Figure 4.17. Stack plot of the characteristic XRPD patterns of the 3 physical forms of clozapine obtained on 150 mg scales in the range of 4-35° 2 $\theta$ . CZPN\_anhydrous, CZPN\_amorphous and CZPN\_monohydrate represent form A, C, and B observed on the 48-well plate.

cut-off of similarity to 85% in subsequent group analysis of all spectra obtained from PS-1 and manual inspection. Evaporative crystallisations in 1:1 mixture of 1,4-dioxane and water yielded either a monohydrate or a non-crystalline phase. Closer inspection of the full Raman spectrum of form D revealed that it has Raman bands of both form B (CZPN monohydrate) and form C (non-crystalline phase) and is therefore a mixture of both. During group analysis of all the spectra obtained from MWP, form D spectra were classified as a separate group due to presence of peaks from both form B and C in the region of 1448-1680  $\text{cm}^{-1}$  which was used in the analysis.

To summarise, 3 physical forms were identified from the HTCAA screen and successfully obtained after scale-up. Detailed analysis of all the physical forms has been described in Chapter 7.

### 4.3.3 Physical Form Screening of Olanzapine

Following the method outlined in section 4.2.4, crystalline solid was obtained in 68 out of 96 wells (hit rate=70.8%) as indicated by polarised light microscopy. Group analysis (see Section 4.2.6) of the Raman spectra obtained from all the wells in the region of 1400-1650  $\text{cm}^{-1}$  of Raman shift yielded 8 groups (Table 4.14). Spectra belonging to group 1 represent the starting material of OZPN and were observed in 19 and 18 wells from rows A-D and E-H respectively (Table 4.14). Spectra that belong to the remaining seven groups denote the novel physical forms of OZPN (Table 4.14). A mixture of form I and a novel form was detected in 5 wells.

Table 4.14. Physical form screening results of olanzapine using 96-well plate after group analysis based on Raman spectra. The wells are colour coded according to the classification group. Wells with no fill represent cases where no crystals were observed.

	conc	1	2	3	4	5	6	7	8	9	10	11	12				
<b>A</b>	1mg	Methanol new	Ethanol new	1-propanol new	2-propanol	1-butanol form I	2-butanol	1-pentanol form I	2-pentanol form I	MA new	EA form I	BA form I	IBA form I				
<b>B</b>	1mg	PA form I	DX new	CH <sub>3</sub> CN new	NM new	THF new	Pyridine	DEC form I	DEE form I	TBMe	Aceton e	MEK	Toluene form I				
<b>C</b>	1mg	FT form I	IAA New	CCl <sub>4</sub> form I	DCM form I	DCE form I	TCE form I	CB form I	BB	TFE new	TEA	Aniline	Hexane form I				
<b>D</b>	1mg	DMF form I	DMA form I	BXE	MXE		DMSO		NMP	PE	MNP	DME	FA	AA new	Water		
					new	form I	new	DMSO									
<b>E</b>	2mg	Methanol		Ethanol new	1-propanol new	2-propanol	1-butanol new	2-butanol form I	1-pentanol form I	2-pentanol form I	MA		EA new	BA		IBA	
		new	form I								new	form I		new	form I	new	form I
<b>F</b>	2mg	PA form I	DX new	CH <sub>3</sub> CN new	NM		THF new	Pyridine	DEC form I	DEE form I	TBMe form I	Aceton e new	MEK	Toluene form I			
					new	form I											
<b>G</b>	2mg	FT form I	IAA New	CCl <sub>4</sub> form I	DCM form I	DCE form I	TCE form I	CB form I	BB	TFE new	TEA form I	Aniline	Hexane form I				
<b>H</b>	2mg	DMF form I	DMA	BXE new	MXE		DMSO		NMP	PE	MNP	DME	FA	AA new	Water		
					new	form I	new	DMSO									

Abbreviations: conc-concentration, MA-methyl acetate, EA-ethyl acetate, BA-butyl acetate, IBA-isobutyl acetate-, PA-pentyl acetate, DX-1,4-dioxane, CH<sub>3</sub>CN-acetonitrile, NM- nitromethane, THF-tetrahydrofuran, DEC-diethyl carbonate, DEE-diethyl ether, FT-4-fluorotoluene, TBMe-tertiary butyl methyl ether, MEK-methyl ethyl ketone, IAA-isoamyl alcohol, CCl<sub>4</sub>- carbon tetrachloride, DCM-dichloromethane, DCE-1,2-dichloro ethane, TCE-tetrachloro ethylene, CB- 1-chloro butane, BB-1-bromo butane, TFE-2,2,2-trifluoroethanol, TEA-N,N,N-triethylamine, DMF-N,N-dimethylformamide, DMA-N,N-dimethylacetamide, BXE-2-butoxyethanol, MXE-2-methoxyethanol, DMSO-dimethylsulfoxide, NMP-N-methylpyrrolidone, PE-2-phenylethanol, MNP-1-methylnaphthalene, DME-1,2-dimethoxyethane, FA-formic acid, AA-acetic acid



Significant differences were found in the region of 1650-1400  $\text{cm}^{-1}$  of the Raman spectra for all the novel physical forms of OZPN (Figure 4.18). The same physical forms were typically observed irrespective of the solute concentration suggesting the robustness of the crystallisation of each form in the presence of the selected solvent. Exception includes wells A10 and E10 with ethyl acetate. In the case of F10 and H3 with acetone and 2-butoxyethanol respectively, novel physical forms were observed with concentration of 2.2 mg/well, while no crystals were observed in respective wells with 1.1 mg/well.

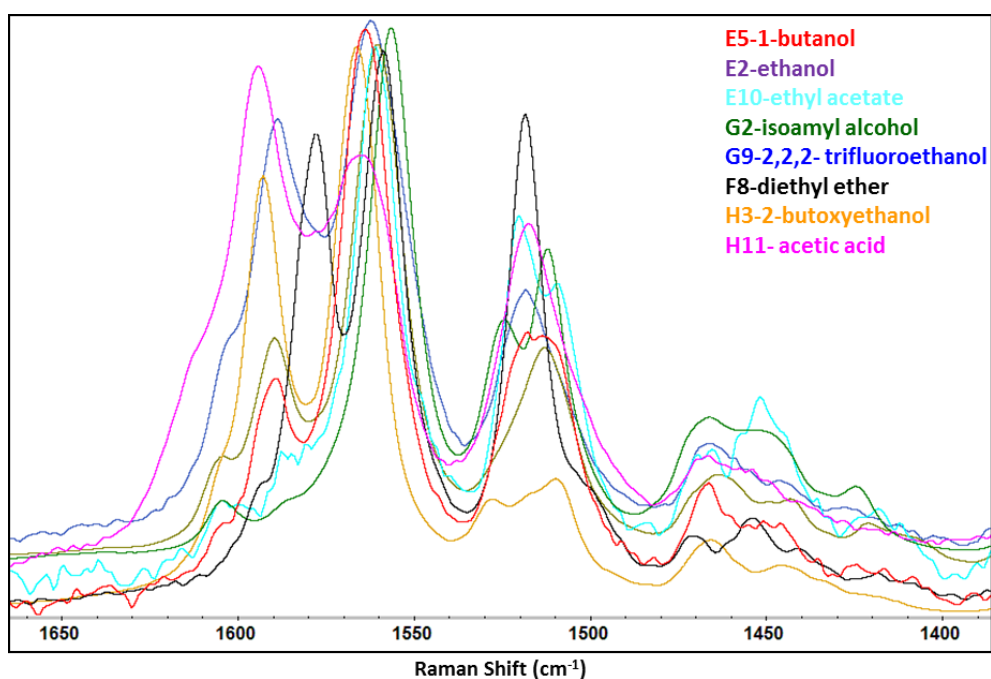


Figure 4.18. Characteristic Raman spectra in the 1480-1650  $\text{cm}^{-1}$  region for 7 novel physical forms and form I of olanzapine identified during HTCAA. Spectra are coloured and labelled according to the well from which they have been detected.

To summarise, 96 crystallisations utilising a total of  $\sim 158$  mg of solute with  $\sim 20$  ml of 48 solvents yielded 7 novel forms of OZPN.

#### ***4.3.3.1 Scale -up of Novel Physical Forms of Olanzapine***

All 7 novel physical forms were obtained from the respective solvents as per plate samples. A stack plot of the characteristic Raman spectra of all 7 physical forms obtained from the screen and on scale-up experiments is shown in Figure 4.19. Six physical forms were characterised as OZPN solvates i.e. acetic acid, ethanol, 2,2,2-trifluoroethanol, 2-butoxyethanol, isoamylalcohol hydrate and 1-butanol hydrate. Crystal structure determination of the 7<sup>th</sup> physical form (detected on wells E10, E11 and E12) revealed that it is a dihydrate B of OZPN (further details on various dihydrates of OZPN are provided in Chapter 6).

Whilst group analysis of Raman spectra in selected range of the Raman shift yielded only 8 distinct groups, SXD/XRPD and thermal analysis revealed that the methanol, ethanol, 1-propanol, 1,4-dioxane, acetonitrile, nitromethane, 2-methoxyethanol, acetone, DMSO, methyl acetate and 1-butanol form different solvates with OZPN. Whilst during group analysis of Raman spectra, novel physical forms from methanol and ethanol were grouped together, the other techniques revealed differences in their characteristic data. Novel physical forms obtained from 1-propanol, 1-butanol, methyl acetate, 1,4-dioxane, acetonitrile, nitromethane, THF, acetone, 2-methoxyethanol and DMSO were also combined in one group from the Raman analysis. SXD analysis revealed that these groups consist of isostructural solvates which have the same long range packing but differ in the solvent and solvent mediated interactions in crystal lattice (discussed in Section 6.4.2 of Chapter 6). Due to the presence of similar

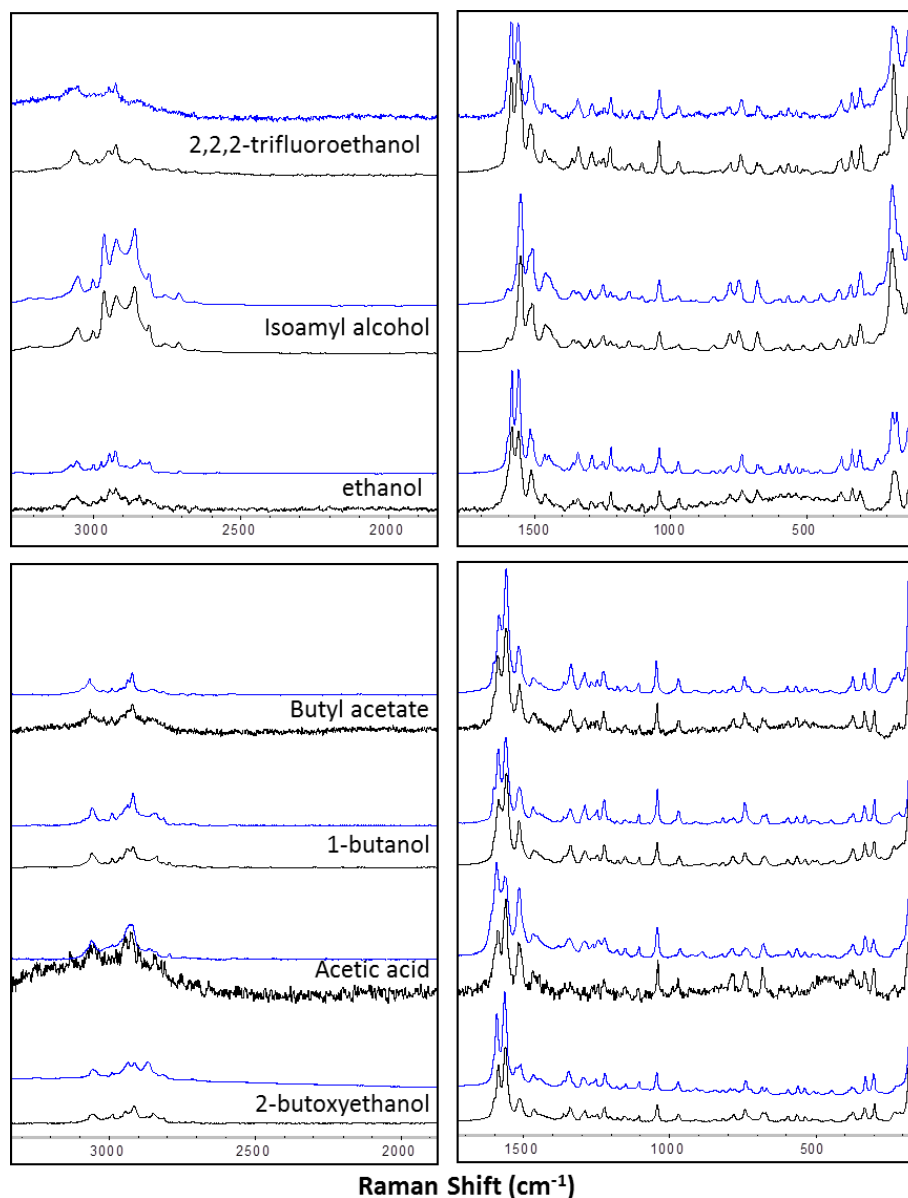


Figure 4.19. Stack plot of Raman spectra collected from samples of physical forms of olanzapine at 96-well plate (black colour) and 150 mg scales (blue colour).

Raman peaks in the region of 1400-1650  $\text{cm}^{-1}$ , these isostructural solvates have been grouped together using a 90% cut off similarity. This emphasises the need for caution in relying in automated data analysis. The use of multiple techniques is always recommended to maximise the certainty of correctly identifying discrete forms.

By closer inspection of the Raman spectra of methanol and ethanol solvates, small differences were observed in the higher frequency and phonon regions which could be due to the different alkyl groups in the solvent moiety and changes in the lattice vibrations respectively (Figure 4.20).

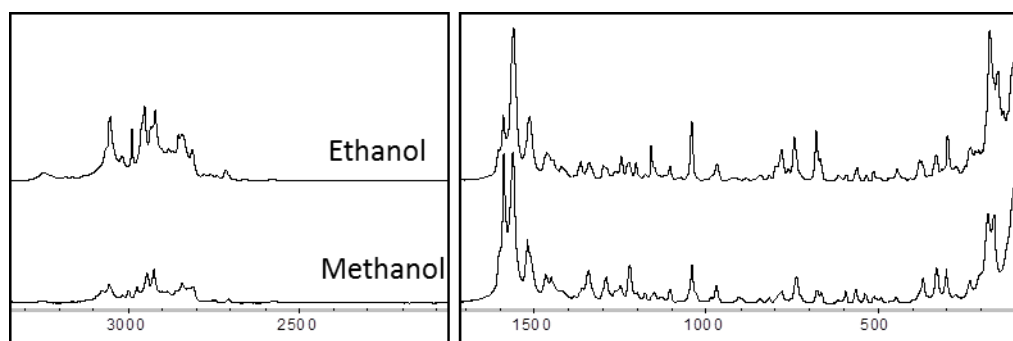


Figure 4.20. Raman spectra of isostructural methanol and ethanol solvates of olanzapine.

Subtle differences were observed in another group of isostructural solvates (hydrated OZPN solvates of 1-propanol, 1-butanol, methyl acetate, 1,4-dioxane, acetonitrile, nitromethane, THF, acetone, 2-methoxyethanol and DMSO) in the regions of Raman shift other than the one used for group analysis ( $1400\text{--}1650\text{ cm}^{-1}$ , Figure 4.21). These differences are due to the different functional groups present in the solvent molecules, i.e. vibration due to cyanide group at  $2255\text{ cm}^{-1}$  in acetonitrile hydrate; C-O-C band at  $835\text{ cm}^{-1}$  in 1,4-dioxane hydrate; variations in higher Raman shift region ( $3100\text{--}3650\text{ cm}^{-1}$ ) due to hydroxyl group in hydrate solvates of 1-propanol, 1-butanol and due to methyl groups in hydrate solvates of acetonitrile, nitromethane, acetone and methyl acetate (Smith and Dent, 2005a).

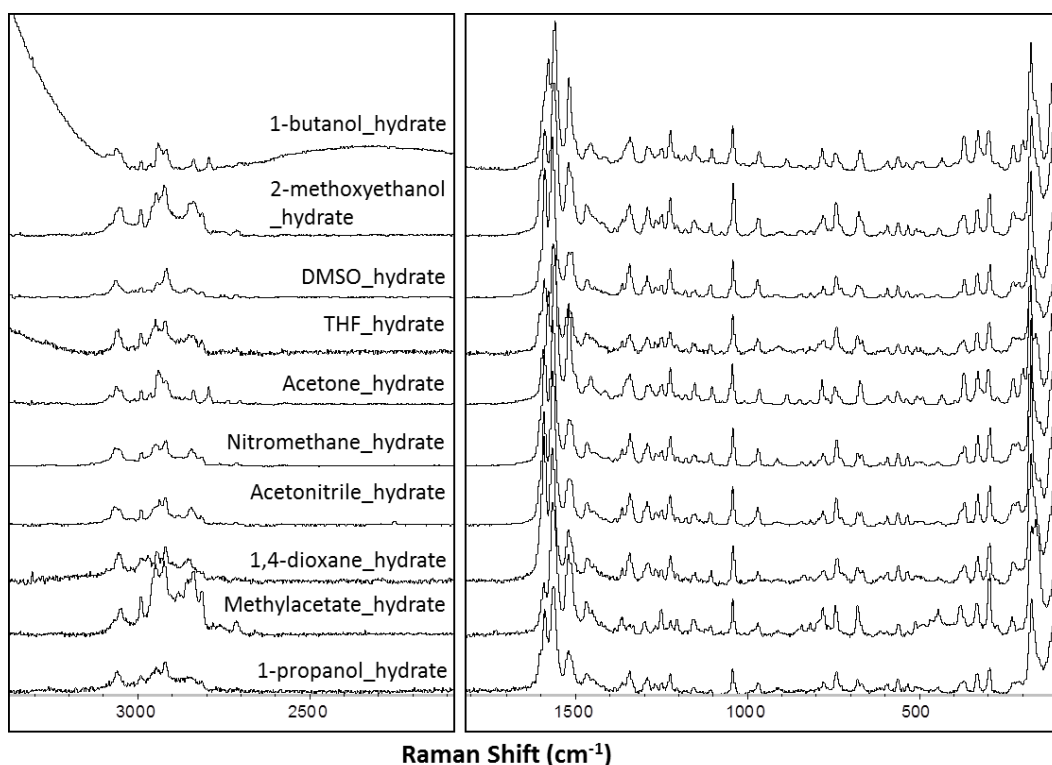


Figure 4.21. Raman spectra of isostructural hydrated solvates of olanzapine.

16 novel OZPN solvates were observed from the HTCAA method and successfully scaled up to mg quantities. A stack plot of XRPD patterns of form I and 16 novel solvates of OZPN is shown in Figure 4.22.

In addition to this, another 7 novel OZPN solvates from DMF, 2-butanol, 2-pentanol, pyridine, TBMe, 1,2-dimethoxyethane and formic acid were also obtained on 150 mg scales. Either form I (4 cases, might be due to desolvation of solvates) or no crystals (3 cases) were observed from these solvents on the 96-well plate. No polymorph of OZPN was observed on 96-well plate or 150 mg scales. Further details of all novel OZPN solvates are provided in Chapter 6.

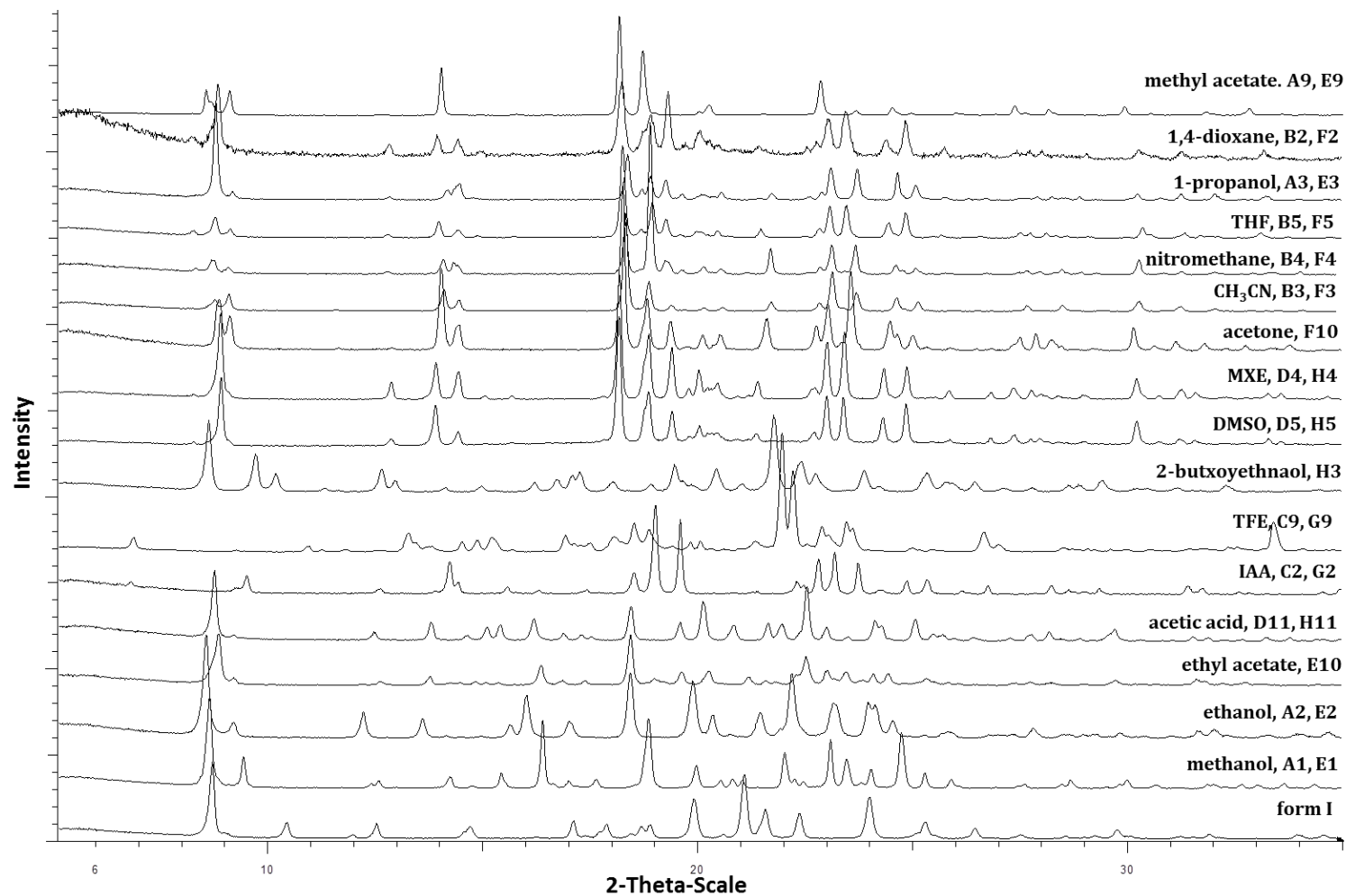


Figure 4.22. Stack plot of XRPD patterns of the 16 novel solvates and form I of olanzapine obtained on 150 mg scales. Abbreviations used in the figure are: THF-tetrahydrofuran, CH<sub>3</sub>CN-acetonitrile, MXE-2-methoxyethanol, DMSO-dimethylsulfoxide, TFE-2,2,2-trifluoroethanol and IAA-isoamyl alcohol.

#### **4.4 Key Findings of the Developed HTCAA Methodology**

The methodology developed in this work has been shown to be useful in finding the initial scope of physical form diversity for a compound using minimal material. This includes salts, polymorphs, solvates and non-crystalline solid forms. Clearly given the ability of Raman to distinguish different lattices and chemical compositions the approach could also have potential applications for co-crystal screening studies.

Typical studies took 2-3 days from set-up to data analysis highlighting the relative speed of the HTCAA approach. By comparison the vial recrystallisations took 6-7 days. The results from the multi-well plate approaches are complementary to larger scale automated methods and can be used to guide large-scale crystallisation efforts (up to 500 mg).

Many HTC approaches have been described in the literature using sophisticated, complex technologies (summarised in Section 4.1). In contrast, this approach represents a simple, fast and cost-effective method, well-suited for initial evaluation of solid-state behaviour of compounds available in mg quantities in the laboratory. Potential advantages and disadvantages for this approach are summarised in Table 4.15.

This methodology is capable of exploring various crystallisation techniques specifically here, cooling crystallisations and solvent evaporation at various temperatures. Also, this approach and equipment would be well suited

Table 4.15. Potential advantages and disadvantages of HTCAA methodology.

<b>Advantages</b>	<b>Disadvantages</b>
Low initial setup cost and low maintenance cost	Prone to human error during manual dispensing
Low material requirement and cost-effective	Only one temperature zone for whole plate
Robust and flexible	Poor control during solvent evaporation
Can be implemented to variety of other diverse crystallisation conditions	No stirring and filtration at RT or under temperature control, thus poor control on compound dissolution
Fast data acquisition and analysis	Limitations of chemometric technique e.g. grouping of isostructural crystal forms
Sensitive analytical technique	Difficult to handle compound exhibiting fluorescence
Quick initial overview of a compound's solid-state behaviour	Potential of cross contamination
	False positives/false negatives

for other crystallisation techniques e.g. slurry methods, crystallisation from melt, co-crystal screening and crystallisation involving heterogeneous nucleation. It can inform candidate form selection by providing additional information (e.g. polymorphic and stoichiometric information of the salt, solvent incorporation in lattice, crystal habit) and to gain a comprehensive understanding of the propensity and extent of solid form diversity. Incorporation of solvents in the crystal lattice is crucial information especially during process development. These solvents can be avoided during manufacturing or during optimisation of crystallisation processes (Morissette et al., 2004).

The use of multi-well plates in this manual approach also has limitations in the range of conditions that can be controlled. It is not always possible to obtain all



the physical forms on a multi-well plate. As has been stated elsewhere (Morissette et al., 2004), an effective and rigorous solid-form screen benefits from multiple crystallisation techniques. This approach can support several of them but is complementary to a few techniques e.g. reverse sublimation, grinding and spray drying. For example, in a study for OZPN, HTCAA found 19 solvates but a further 7 were discovered during further scale-up experiments (using 100-200 mg of OZPN with 3-5 ml of solvent in vials). Other novel solvates of OZPN were also observed using different crystallisation techniques as a part of larger crystallisation screen (see Chapter 6). No polymorphs of OZPN were observed from this methodology, whilst 2 additional polymorphs were observed using other crystallisation methods (see Chapter 6). This highlights the fact that a single method cannot guarantee all physical forms of an API. It is advisable to complement the HTC methodology with various in-depth experiments using multiple crystallisation techniques e.g. grinding, desolvation of solvates, recrystallisation from melt and reverse sublimation, in order to explore the complete experimental crystal energy landscape of an API.

Some of the limitations encountered during development and validation of this methodology relate to the sole reliance on Raman spectroscopy. Raman spectra acquired from isostructural solvates were very similar and sometimes difficult to classify as separate spectra types. New physical forms that are structurally closely related could therefore be missed if no further investigation is done using other analytical techniques including XRPD, SXD and thermal analysis. For example, 9 solvates of OZPN which were grouped in the same cluster during HTCAA were successfully identified by further investigation with other

analytical techniques. Raman spectra with small changes due to particle size, orientation and residual solvent can be mistaken as discrete physical forms and characterisation using other analytical techniques is advisable to avoid false positives. Whilst Raman allows the presence of a new form to be established it was not always possible to determine whether this new form was a polymorph, hydrate or solvate.

Other frequently encountered challenges associated with Raman microscopy of molecular solids are fluorescence and sample charring. Fluorescence is the “process of emission of electromagnetic radiation in the form of light as a result of absorption of light from another source” (Smith and Dent, 2005b). This emission appears as a strong background in the Raman spectrum when the sample is illuminated by the excitation laser. Fluorescence usually appears as a very broad curve extending across hundreds of wavenumbers. An example of a Raman spectrum exhibiting fluorescence is shown in Figure 4.23.

Fluorescence may be an inherent property of the compound under study though it can also occur due to the presence of an impurity or residual solvent or from the well plate material itself (hence quartz well plates were used in this work which exhibit minimal fluorescence). As a result, a relatively large fluorescence background may be encountered that can dominate the emission spectrum resulting in degradation or complete loss of the Raman signal (Matousek et al., 2002; Pommier, 2005; Lowry et al., 2006; Romero-Torres et al., 2009).

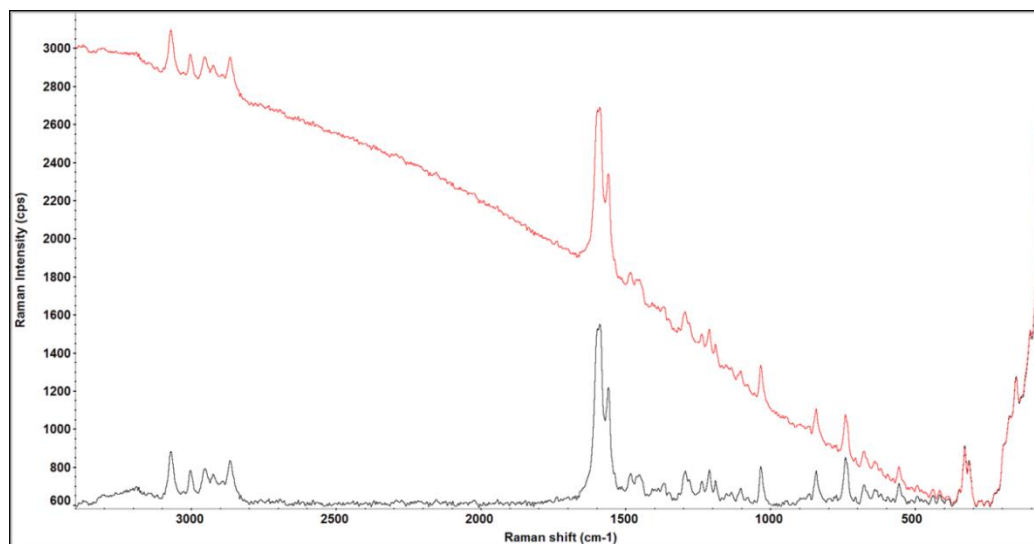


Figure 4.23. Raman spectra of the sample exhibiting fluorescence (red) and after applying fluorescence correction (black).

Fluorescence can be minimised using a number of approaches. Lasers with longer wavelengths e.g. in near Infra-red region (NIR; 800-2500nm) can be used to minimise the fluorescence as the fluorescence intensity usually increases at shorter wavelength (Matousek et al., 2002). In the case of pharmaceutical compounds studied here, the intensity of Raman bands was low with 780 nm laser. In comparison, the intensity from the 532 nm laser was suitable for data analysis and hence was used in majority of this work. Where fluorescence is an inherent property of the compound under investigation and sharp Raman peaks are visible on the back of broad background, baseline correction can be applied to subtract it from the background. In some cases, derivation of the spectrum is required to remove majority of the broadband background. However, by doing this, the noise of the spectrum can also get amplified. Therefore, smoothing is often applied in conjunction with derivatives to remove high frequency noise from the spectrum. Care must be taken when selecting the smoothing function

as it can affect the spectral resolution and ability to differentiate between physical forms with subtle changes in spectra. Ideally a smoothing function is preferred which would provide the highest signal to noise ratio with the least amount of smoothing (Pommier, 2005; Lowry et al., 2006).

For impure samples, recrystallisation of the material to remove the impurities responsible for the fluorescence is another approach to remove/minimise it (Pommier, 2005), although impractical in the current context as it may have impact on polymorphism. In one of the test compounds used in this study, raw material of CZPN showed high fluorescence. However, no fluorescence was observed in the sample obtained from vapour phase crystallisation. This suggests that fluorescence in the raw material might be due to the presence of a fluorescent impurity.

Sometimes the laser can cause thermal decomposition due to sample heating by high excitation energies of the laser. This was observed with OZPN and few samples of AXP. This was avoided by using lower laser power or selecting a 780 nm laser, however this also impacts signal to noise ratio or results in longer data collection times.

## 4.5 Summary

In this chapter, the development and validation of HTCAA methodology for physical form screening using Raman microscopy and a quartz multi-well plate has been demonstrated. This methodology is complementary to the existing large scale parallel approaches and offers various advantages including reduced sample requirement (~1 mg per crystallisation), no sample preparation prior to data collection, rapid data acquisition using a non-destructive analytical technique and rapid analysis of the spectral data using chemometric techniques. It has been shown to be an effective tool for preliminary investigation of polymorphism, solvate formation and salt formation providing data to guide further efforts in the most productive direction. HTCAA provides knowledge of suitable crystallisation conditions that give specific forms and provides a basis for subsequent scale-up.

A variety of physical as well as chemical information was obtained using Raman microscopy during salt screening of AXPN. Chemical information suggests salt formation, stoichiometry with polyprotic counter acid and (in some cases) solvent incorporation into the crystal lattice. Physical information includes possible polymorphism and crystallinity of the salts.

As a relatively low-cost screening tool, HTCAA provided a useful overview and an initial scope of the solid-state diversity of OZPN and CZPN with the given set of solvents in a shorter period of time. One HTCAA experiment involving 96 crystallisations on a 96-well plate indicated a complex crystal structure landscape for OZPN. Approximately 158 mg of OZPN with ~20 ml of 48 solvents

yielded 16 physical forms. In contrast, only 3 physical forms were detected for CZPN with the same number of crystallisations carried out in two experiments on a 48 well plate by utilising ~235 mg of CZPN and ~24 ml of 96 solvent systems. Both compounds have similar solubility profile in selected solvent systems so the experiments are comparable. It took 2-3 days for experiments on OZPN and CZPN whereas screening carried out using traditional methods for the same compounds using same solvent systems took 1-2 weeks.

Following the initial discovery, the majority of the forms were successfully obtained on 150 mgs scale in 3-8 ml vials. However, 7 new solvates of OZPN were obtained during scale-up that had not been identified by the initial MWP experiments. This demonstrates that solvent volume and the inexact control of conditions during small scale crystallisation can play a role in deciding crystallisation outcome.

## **5 Predicting Crystallisability of Organic Molecules using Statistical Modelling Techniques**

## 5.1 Statistical Modelling Techniques

### 5.1.1 Introduction

Statistical modelling and analysis used for data mining, machine learning and prediction involves careful selection and appropriate application of statistical analysis techniques. This requires certain underlying assumptions to be fulfilled, data analysis, correlation, hypothesis and testing of hypothesis to validate conclusions. During analysis of data, there are generally two goals; prediction and inference. Prediction involves analysis of the existing data and drawing of conclusions which can be used to predict the hypothetical or future data. Inference involves investigation of the nature of correlation between various components of complicated processes (Mytkolli and Calitoiu, 2009). Various stages involved in a typical modelling process are shown in Figure 5.1.

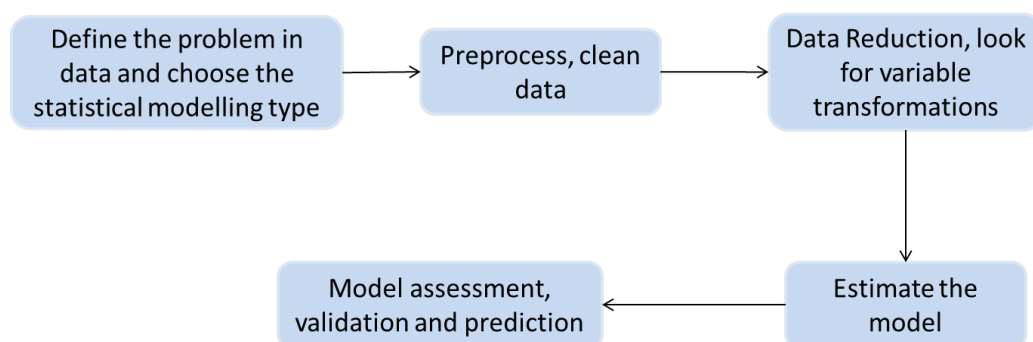


Figure 5.1. Schematic workflow of various stages of a typical statistical modelling process.

Various statistical modelling techniques are used in the pharmaceutical industry at different stages of drug discovery and development. These include principal



component analysis (PCA) (Wold et al., 1987), artificial neural networks (ANN) (Kauffman and Jurs, 2001; Doniger et al., 2004), multiple linear regression (MLR) (Kauffman and Jurs, 2001), partial least squares (PLS) (Sheridan et al., 1994), hierarchical clustering (HC)(Makretsov et al., 2004), Random Forest (RF) (Breiman, 2001) and support vector machine (SVM) (Doniger et al., 2004). All these approaches have their associated advantages and disadvantages due to their varying complexities. Two methods; PCA and RF programs, which have been used for this work, are described in the sections below.

### 5.1.2 Principal Component Analysis

PCA is a powerful tool for reducing the number of observed variables to a smaller number of artificial variables known as principal components (PC) (Figure 5.2). The first PC accounts for maximum possible variance in the dataset and each subsequent PC accounts for the remaining variance to the fullest extent (Wiklund, 2008; Suh et al., 2009). Typically, the first two PCs will describe 60-70% of the overall variance in the dataset.

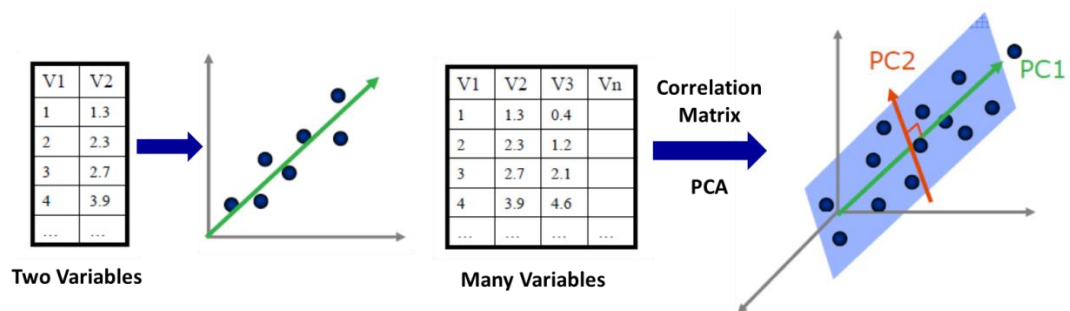


Figure 5.2. Principal Component Analysis. Reproduced from reference (Wiklund, 2008).

PCA effectively removes redundancy (i.e. correlated variables which measure the same response) which is important as highly correlated variables can exert a disproportionate influence and bias results (Eder et al., 1994). PCA is helpful in identifying patterns in the data and highlighting the similarities and differences within the data. It is also a useful tool for classification and of particular use when a data reduction procedure is required which makes no assumptions of an underlying causal structure that is responsible for co-variation in the data.

Two interpretable plots are obtained from PCA namely; score plots and loading plots (Figure 5.3). The score plot provides information about the relationship amongst objects whereas the loading plot depicts the relationship amongst variables. The following guidelines can be used to interpret the score and loading plots.

- Similar objects are close to each other on score plot.
- Variables with high correlation are close to each other on loading plot.
- Variables on opposite side of origin are negatively correlated.
- Objects on the right of score plot are dominated by variables on the right of loading plot and same applies for the objects on left of the score plot (Wiklund, 2008; Suh et al., 2009).

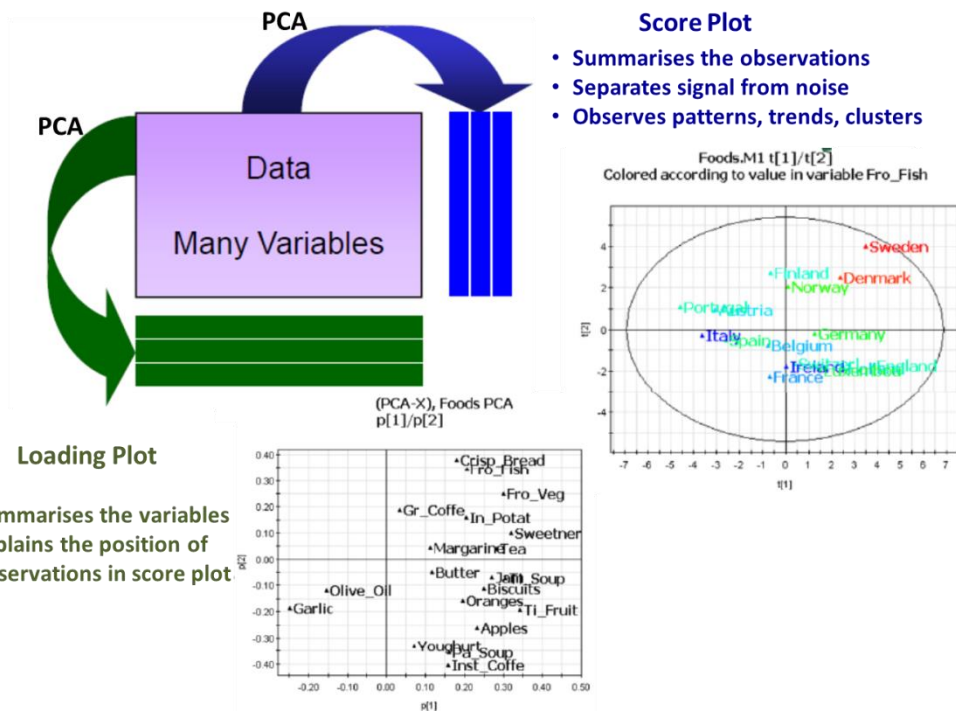


Figure 5.3. Example of Score and loading plots obtained from Principal Component Analysis while investigating the patterns of food consumption of 20 food products in 16 countries. Score and loading plots relate to observations (countries in this case) and variables (food products in this case) respectively. Reproduced from reference (Wiklund, 2008).

### 5.1.3 Criteria for Deciding the Number of Principal Components

In general, only the first few PCs account for meaningful variance in the data, and the higher PCs will tend to account for only trivial variance. Four criteria are used while deciding the number of components during PCA analysis. The 1<sup>st</sup> criterion is known as “Eigenvalue-one criterion/Kaiser criterion” which means that any PC with an eigenvalue greater than 1.00 is retained (Kaiser, 1960). Eigenvalues provide information on the content of individual PCs. According to the 2<sup>nd</sup> criterion, “The Cattell Scree test”, the eigenvalues associated with each

component are plotted and the plot is then assessed to identify a break between the components with relatively large eigenvalues and those with small eigenvalues. The components that appear before the break are assumed to be meaningful and are retained; those appearing after the break are assumed to be unimportant and are discarded (Cattell, 1966). The other way is to find a place in the plot, where the smooth decrease of eigenvalues appears to level off to the right of the plot. There is no set criterion for deciding this level. For example, a scree plot for a dataset (Figure 5.4) may suggest as many as 8 PCs but in this case 5 PCs were used based on combined results of both Eigenvalue-one and Cattell Scree test criteria (Xu and Redman-Furey, 2007).

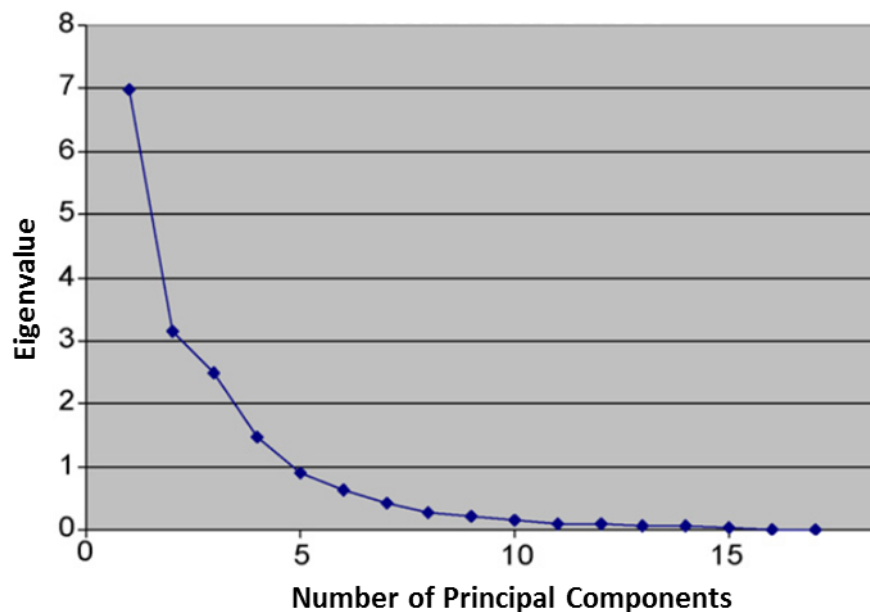


Figure 5.4. Scree plot of the number of principal components vs. eigenvalues. Adapted from reference (Xu and Redman-Furey, 2007).

The 3<sup>rd</sup> criterion is based on the “percentage of the variance explained by the component in the data set”. This can be calculated by dividing the eigenvalue for

the component of interest by total eigenvalues of the correlation matrix. Components which explain 5-10 % of variance can be retained. A cumulative percentage of variance can also be used as a criterion for selecting components. For example, how many components are required to explain at least 70% of variance of data set. The 4<sup>th</sup> criterion is known as “Interpretability criteria” which includes the interpretation or substantive meaning of the retained components using any of the above mentioned three criteria and verifying that this interpretation makes sense in terms of what is known about the system under investigation (Wold et al., 1987; Ferré, 1995).

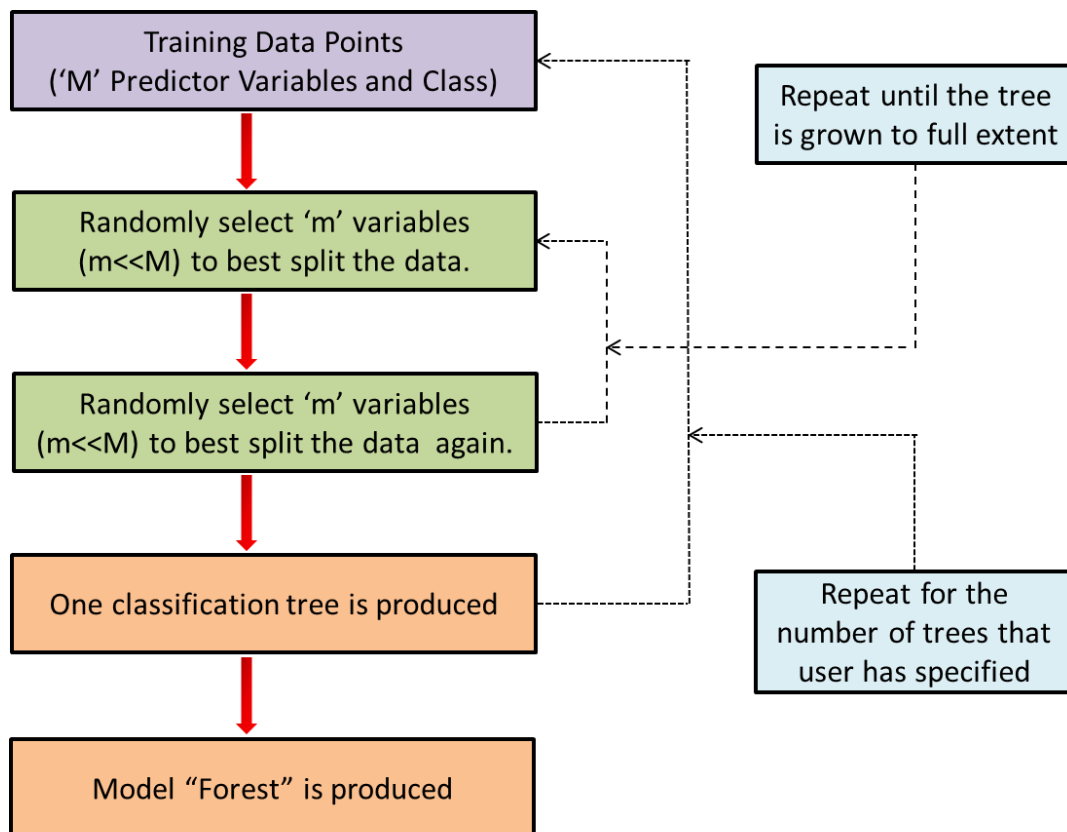
#### **5.1.4 Random Forests Classification Method**

Random forests (RF) is a method for classification and regression (Breiman, 2001; Liaw and Wiener, 2002; Svetnik et al., 2003). RF is an ensemble classifier that consists of many decision trees. Prediction of a continuous variable is provided as an average of the predictions of all trees. RF offers various advantages over other methods which are stated below;

- No over-fitting of the data, as the generalisation error almost converges with increase in number of trees.
- Provides an estimate of internal error, strength and correlation.
- Can handle large number of descriptors and employs a good method to estimate the missing value.
- Provides a measure of descriptor’s importance.
- Can handle multiple outcomes.

- High predictive accuracy and is applicable even in high-dimensional problems with highly correlated variables.
- Relatively robust to outliers and noise.

A schematic diagram of an RF workflow is shown in Figure 5.5. Bootstrap sampling and random selection of input descriptors are used to induce randomness in the input data used to develop the RF model. This ensures that the classification trees grown in the forest are dissimilar and not correlated to each other.



### Random Forest ---- Building the Model Forest

Figure 5.5. Schematic workflow of building of Random Forests model.

Using bootstrap sampling, classification trees are grown using 2/3<sup>rd</sup> of the dataset and remaining 1/3<sup>rd</sup> of the dataset [Out Of Bag (OOB) data] is employed to obtain unbiased estimates of correct classification rates (internal estimates of error). Compared to a single classification tree, this algorithm yields better prediction rates and is more robust in dealing with noise in the data set because the forest of trees are grown to the full extent. The generalisation error of a forest of trees classifiers depends on the strength of the individual tree in the forest and the correlation between them. As part of a forest construction, RF computes proximities (overall similarities and dissimilarities) between the objects and provides a proximity matrix which can be used for clustering and detecting outliers (Breiman, 2001; Liaw and Wiener, 2002; Svetnik et al., 2003). The proximity matrix from the RF is generally analysed using multi-dimensional scaling (MDS) in 2-D/3-D plot. The closer the two observations in the MDS plot, the larger the proximity between them (Steyvers, 2006).

The RF algorithm also assesses the importance of descriptors used in building of the classification model. It is assessed by replacing each descriptor in turn by random noise and the resulting deterioration in the model quality is a measure of descriptor importance. The deterioration in the RF model quality is assessed by mean decrease in accuracy (based on OOB data) and/or mean decrease in Gini (Gini index used in splitting) (Li et al., 2005; Palmer et al., 2006).

RF is widely used in many scientific areas, e.g. for measuring aqueous solubility (Palmer et al., 2006), predicting mutagenicity, (Zhang and Aires-de-Sousa, 2006), QSAR studies (Debeljak et al., 2007), for building of drug likeness

classification models (Good and Hermsmeier, 2006), and selecting relevant predictor variables in the analysis of microarray data, in DNA sequencing and other applications in life sciences (Lunetta et al., 2004; Bureau et al., 2005; Huang et al., 2005; Li et al., 2005; Qi et al., 2006). There is only one report of RF application in the area of crystallisation of small molecules (Johnston et al., 2008).

### **5.1.5 Applications of Statistical Modelling Techniques in Pharmaceutical Industry**

Statistical modelling techniques are widely used at various stages of drug discovery and development in the pharmaceutical industry (Tye, 2004; Alsenz and Kansy, 2007). Due to advances in high throughput instrumentation and robotics being applied in drug discovery and development activities, large volumes of data are routinely generated. Statistical modelling techniques can be used in experimental design and optimisation (Rose, 2002). They can also be used for analysis of complex multivariate datasets in an efficient and robust way which then allows the user to sift through the information and reveal trends between factors and responses within the data. Statistical models can be applied to yield quantitative structure property relationships (QSPR) for predicting physicochemical properties (water solubility, lipophilicity etc.) during hit identification and lead optimisation stages of drug discovery (Taskinen and Yliruusi, 2003; Alsenz and Kansy, 2007). This area has benefited from advances in computing power enabling rapid calculation of 2-D and 3-D molecular



descriptors and generation of fast outputs from the range of commercial and freely available analysis programs. This capability leads to accelerated validation of QSPR models (Hughes et al., 2008).

Statistical techniques have also been used in various aspects of solid form selection and during the development of robust crystallisation processes. Statistical modelling techniques are also of relevance in the context of Quality by Design (QbD) approaches in pharmaceutical manufacturing (Kourti, 2009). Some of the examples of applications of statistical modelling techniques in crystallisation area are briefly described.

#### ***5.1.5.1 Solvent Classification/Selection***

The aim of physical form screening is to cover the maximum area of crystallisation space efficiently. For solution crystallisations this can be aided by judicious choice of solvents with diverse physicochemical properties. Solvent variation is a straightforward means of designing diversity in the experimental crystallisation screen and manipulating the crystallisation outcome. Using statistical approach, an attempt is made for grouping of solvents by quantifying the similarities and dissimilarities of the solvents in a given library based on various experimental and calculated physicochemical descriptors of solvent molecules, thus providing a basis for rational solvent selection. Ideally these physicochemical descriptors are directly relevant to the solvent's potential influence over the nucleation and growth of the solute. The selection of suitable solvent physicochemical descriptors for grouping is a persistent challenge

because of the poorly defined relationship between solvent, nucleation and crystal growth. Furthermore, crystallisation is influenced by many experimental parameters other than solvent identity (Florence, 2009b).

Gu. et al. clustered 96 solvents into 15 solvent groups using a HC method (Gu et al., 2004). A total of 8 physicochemical descriptors for each solvent were employed for this clustering. In another case, 57 solvents were clustered into 20 groups based on 17 physicochemical descriptors using cluster analysis (Xu and Redman-Furey, 2007). These approaches classify different solvents in specific groups allowing the selection of diverse conditions using one solvent for each cluster initially. Similarly, in an in-house study, 24 solvent clusters were obtained using ClusterSim approach on 94 solvents. Using this ClusterSim approach, both polymorphs of HCT were obtained by using 24 representative solvents from the entire library, one selected from each cluster as well as identifying likely solvate formers. A similar approach was employed for polymorph screening of OZPN and no polymorph was observed from any of the 24 solvents used for screening. This is believed to be due to OZPN's prolific tendency of solvate formation. To date only the thermodynamically stable form has been obtained directly from solution crystallisation and all other polymorphs can be obtained from alternative crystallisation methods (see Chapter 6).

Although the grouping and related approaches for assessing solvent diversity cannot guarantee that the multiple solid forms of the solute will be obtained from solution, they still add considerable value while selecting solvent libraries.

Analysis of the property space of the solvent library can help in identifying the under-represented regions that would benefit from the addition of new solvents or identifying solvents with largely similar properties which may be considered for exclusion (Florence, 2009b).

#### ***5.1.5.2 Design of Experiment***

It is well known that the chances of obtaining all physical forms increase with the number of properly designed experiments. Therefore, Design of Experiment (DOE) approaches can be employed for selecting the optimum number of experiments required to maximise the coverage of experimental crystallisation space within the constraints of available material, equipment and time. It can also be implemented in optimising the outcome of the crystallisation experiment e.g. maximising the chances of getting the desired form (McCabe, 2010). Application of DOE for polymorph screening of CBZ has showed that not only new forms can be found, but also analysis of input variables against outcome response allows prediction of the crystallisation outcome (i.e. polymorphic form) of crystallisation experiments (McCabe, 2010).

#### ***5.1.5.3 Assessment of Completeness of Experimental Screens***

Statistical modelling techniques can be used to assess the completeness of an experimental screen. An attempt can also be made to identify the correlation between the conditions tested and the outcomes. Retrospective inspection of the crystallisation outcomes of a comprehensive physical form screen for CBZ

against the crystallisation conditions using RF led to three novel crystalline solvates of CBZ obtained from solvents predicted to lead to solvates (Johnston et al., 2008).

#### ***5.1.5.4 Revealing Trends in Larger Datasets***

Statistical analyses are quite useful in revealing trends in large and complex datasets, e.g. statistical analysis of a database of 34770 organic crystal structures from the CSD has revealed that the hydrate formation is not affected by the donor/acceptor ratio or the molecular weight of a compound, but that the probability of hydrate formation increases with an increase in the polar surface area (Infantes et al., 2007). A statistical analysis of 974 co-crystal structures from the CSD showed that shape and polarity of constituent molecules of the co-crystals tend to be similar. Interestingly the complementarity of the number of hydrogen bond donor/acceptors in the constituent molecules of the co-crystals was not found to be a dominant factor (Fabian, 2009).

#### ***5.1.5.5 Predicting Crystallisation of Small Molecules from the Melt***

Taylor and co-workers have reported a classification model using PCA to evaluate the crystallisation tendency of organic molecules from undercooled melts. A relationship between molecular and physical properties of organic molecules with glass forming ability (GFA) and glass stability (GS) tendency was observed. This model has applications in assessing the glass forming ability and

the stability of new chemical entities, allowing the evaluation of the viability of amorphous formulation during early stages of drug development. (Baird et al., 2010). The same group also reported that viscosity is an important parameter in determining the crystallisation tendency of organic compounds from undercooled melts (Baird et al., 2012).

#### ***5.1.5.6 Predicting Crystallisability of Proteins***

Structural characterisation of proteins has been hindered because of the inherent difficulty in obtaining good quality crystals suitable for diffraction due to their highly dynamic conformations. To overcome this, various crystallisability predictor models have been developed using the data available in Structural Genomic consortia (Nuria Sanchez-Puig et al., 2012). These predictive models require only the protein sequence as input to predict crystallisability. They are based on various input parameters including energy and hydrophobicity indices, composition of certain amino acid types, secondary structure, solvent accessibility, predicted disorder and content of certain buried and exposed residues.

Various crystallisation propensity prediction models include PPCpred (Mizianty and Kurgan, 2011), OBScore (Overton and Barton, 2006), XtalPred (Slabinski et al., 2007), ParCrys (Overton et al., 2008), CRYSTALP2 (Kurgan et al., 2009), MetaPPCP (Mizianty and Kurgan, 2009) and SVMCrys (Kandaswamy et al., 2010). Depending on the program, the output can be simple categorical descriptors e.g. “crystallisable” or “non-crystallisable” or a category 1-5, where

the highest value corresponds to proteins that are “very difficult” to crystallise and the lowest value corresponds to proteins that are “optimal” to crystallise. Only PPCpred can be used to predict the propensity of each of four steps in a full chain which includes i) production of diffraction quality crystals, (ii) production of crystals, (iii) purification and (iv) production of the protein material and inform about the step responsible for crystallisation failure.

Comparative analysis has suggested that these predictive models have a prediction accuracy of ~70-80%. These predictive models take into account only the intra-molecular factors encoded within protein chains. Therefore, these crystallisation propensity predictor models may fail to provide dependable predictions when the inter-molecular factors such as the protein–protein and/or protein–precipitant interactions, characteristics of the expression systems, buffer composition, precipitant diffusion method, etc., must be considered (Mizianty and Kurgan, 2011).

There are no reports of crystallisability prediction of small molecules from solution; therefore, it was of interest to use the applications of statistical modelling techniques in prediction of crystallisation propensities for small molecules. Organic compounds exhibit different crystallisation propensities. Some organic molecules crystallise well or quickly, while others do so badly or slowly. Poor crystallisation behaviour can include a collection of outcomes including nano/micro crystals formation, oiling out, no significant improvement in purity, and/or agglomeration. Despite great efforts towards better theoretical understanding of crystal nucleation and growth, it is still not possible to

understand how to improve difficult and troublesome crystallisations. Hence in most practical situations, experience, trial-and-error and empirical knowledge are largely relied upon to achieve a desirable outcome (Hursthouse et al., 2009).

In this chapter, PCA and RF were applied to predict the crystallisation propensities (“crystallisability”) of a published dataset of small organic molecules (Hursthouse et al., 2009). Here, crystallisability means that the molecule is crystallising to give single crystals suitable for SXD. An RF model was built using calculated 2-D and 3-D molecular descriptors of organic molecules and was used to model the property space responsible for the different observed crystallisation outcomes. The developed model was also used to find important calculated molecular descriptors governing the varied crystallisation outcomes. A comparison of these descriptors with experimentally known factors responsible for reduced crystallisation tendency was also performed.

## **5.2 Descriptor Calculations, Model Building and Validation**

### **5.2.1 Training Dataset and 2- and 3-Dimensional Descriptors Calculations**

There are very few examples of schematic crystallisation studies on a series of organic molecules, however, Hursthouse, *et al.* published a dataset (Hursthouse et al., 2009) which was suitable for predicting the crystallisation propensities. In the reported study, crystallisation was attempted from ethanol in the first

instance. Other solvents used for crystallisation were water, toluene, benzene, aqueous ethanol, cyclohexane-benzene mixtures, DMSO, DMF and pyridine. However, no detailed information is provided in the report and it is also not known whether every solvent was used for crystallisation for each molecule. While selecting this dataset, it was assumed that that crystallisation was well controlled.

This dataset reported, consists of 382 molecules with a common acylanilide skeleton (Figure 5.6). The skeleton was modified to create 382 analogues by alteration of substituents of the R and X groups. This created a diverse library of molecules that share a common molecular nucleus.

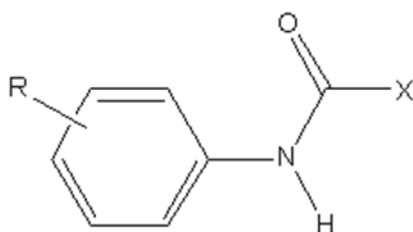


Figure 5.6. Basic skeleton of acylanilide molecules, R and X represent the two points of modifications. X includes H, CH<sub>3</sub>, C<sub>2</sub>H<sub>5</sub>, C<sub>3</sub>H<sub>7</sub>, C(CH<sub>3</sub>)<sub>3</sub>, CF<sub>3</sub>, OCH<sub>3</sub>, OC<sub>2</sub>H<sub>5</sub>NH<sub>2</sub> and Cl. R includes H, CH<sub>3</sub>, C<sub>2</sub>H<sub>5</sub>, C(CH<sub>3</sub>)<sub>3</sub>, OCH<sub>3</sub>, OC<sub>2</sub>H<sub>5</sub>, OCF<sub>3</sub>, F, Cl, Br, I, CF<sub>3</sub>, OH, NH<sub>2</sub> and COOH.

All the molecules were drawn using Chemdraw Ultra (version 11.0) and reliable 3-D conformations generated in Discovery Studio using the Pipeline Pilot interface (Accelrys, 2010). 2-D (185) and internal 3-D (i3-D) (123) molecular descriptors were calculated using MOE (MOE, 2002). 2-D molecular descriptors are defined to be numerical properties and calculated from the atoms and



connection table of the molecule. 2-D molecular descriptors included physical properties descriptors (e.g. molecular weight, sum of the atomic polarisabilities), atom and bond count descriptors (number of atoms, number of rotatable bonds), connectivity and shape indices descriptors (atomic connectivity index, first kappa shape index), adjacency and distance matrix descriptors (largest value in the distance matrix, balaban's connectivity topological index), pharmacophore features descriptors (e.g. number of hydrogen bond donors and acceptors) and partial charge descriptors (atomic partial charges calculated using Partial Equalisation of Orbital Electronegativities (PEOE) method). i3-D molecular descriptors are calculated using 3-D coordinate information of the molecules. The i3-D molecular descriptors included potential energy descriptors (e.g. total potential energy and torsion potential energy), volume, surface area and shape descriptors (e.g. van der Waals volume and water accessible surface area) and conformation dependant charge descriptors (e.g. dipole moment calculated from the partial charges of the molecule).

A correlation matrix was prepared using a Pearson correlation coefficient by using a Pipeline Pilot interface. Molecular descriptors which showed zero variance and covariance (threshold of Pearson correlation coefficient >90%) were removed from the dataset. The resultant training dataset comprised of 151 calculated molecular descriptors. The two known crystallisation outcomes from the original report (Hursthouse et al., 2009) were put into the training set and described as: class 1, where single crystal suitable for SXD was observed and class 2, where no single crystal suitable for SXD was observed.

## 5.2.2 Training the Statistical Model

A PCA of crystallisability was carried out using the SIMCA program (version 12.0 and 13.0) (Umetrics) (SIMCA, 2012) by inputting 2-D and 3-D calculated physicochemical descriptors for molecules. PCA was carried out with varying number of PCs to explore the effectiveness of procedure in describing the data variability.

RF analysis was carried out using the 'RandomForests' package available in the statistical computing and graphics environment 'R' (version 2.10.1) (R Development Core Team, 2006) denoted as RF/R and a commercially available RandomForests® package (Salford Systems) represented as RF/Salford. The optimal values of parameters; seed, mtry, ntree, nodesize and jclasswt were selected until no further improvement was seen in the classification model.

Here, 'ntree' refers to the number of trees grown during model building and was increased incrementally until no further improvement was observed in the model. 'mtry' is the number of different molecular descriptors tried at each split and the default value is the square root of the total number of input descriptors. 'jclasswt' allows weightings to adjust error rates between classes that have very different number of observations. 'nodesize' refers to the minimum nodesize below which leaves are not further subdivided and the default value is 1. 'Seed' refers to any non-zero integer number which controls the random number generator. It was arbitrarily set to 45 to provide reproducibility in the random numbers required by the RF. OOB error of estimate was used as a guide during model training process.

The RF model was trained using all 151 calculated descriptors and 2 crystallisation outcomes using the following parameters: ntree = 20000, mtry = 12, jclasswt = 1 (for class 1) and 1950000 (for class 2), nodesize = 1, seed = 45.

The RF model reports the crystallisation prediction as probabilities, which correspond to the percentage votes across all trees for a molecule as each crystallisation outcome (class 1 vs. class 2). For each molecule, RF prediction provides a distribution of percentage votes for each defined outcome, totalling 100%. The mean decrease in accuracy method (see Section 5.1.4) was used to assess the relative importance of molecular descriptors responsible for the final classification model.

## 5.3 Results and Discussion

### 5.3.1 Principal Component Analysis

PCA revealed that only 51% variance was explained using two PCs. Thirteen PCs were required to explain only 81% variance suggesting that PCA is not able to define a simplistic trend in a complex dataset. No significant clustering of two classes (crystals vs. no crystals) was observed from the PCA (Figure 5.7). Due to inability of the PCA to provide a useful model of the relationship in the dataset, no further investigation was done and a decision was taken to use RF technique.

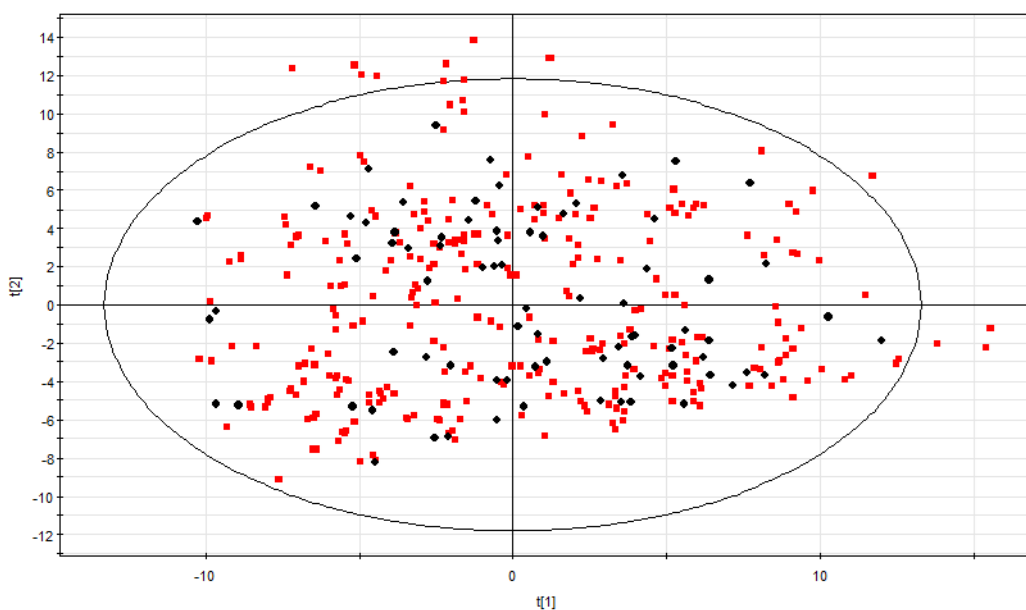


Figure 5.7. Scatter plot of properties of molecules, shown with an ellipse that represents the Hotelling T2 with 95% confidence. Each of the 382 points on the plot represents a molecule from the dataset. Molecules are coloured according to the crystallisation outcome: class 1 (crystals, red) and class 2 (no crystals, black).

### 5.3.2 Random Forests Classification Model

During RF classification model building, the overall error rate was converged with an increase in the number of trees. The final RF model has classified the molecules in two classes with an overall OOB error of prediction of 32.6% and prediction accuracy of 67.4% (Figure 5.8). In procedure, this translates to a 67% chance of accurately predicting the current outcome of a molecule based on the RF model. This model is better than the random accuracy of 50% expected in a system with binary outcome. Similar results were obtained from both the programs; RF/R and RF/Salford.

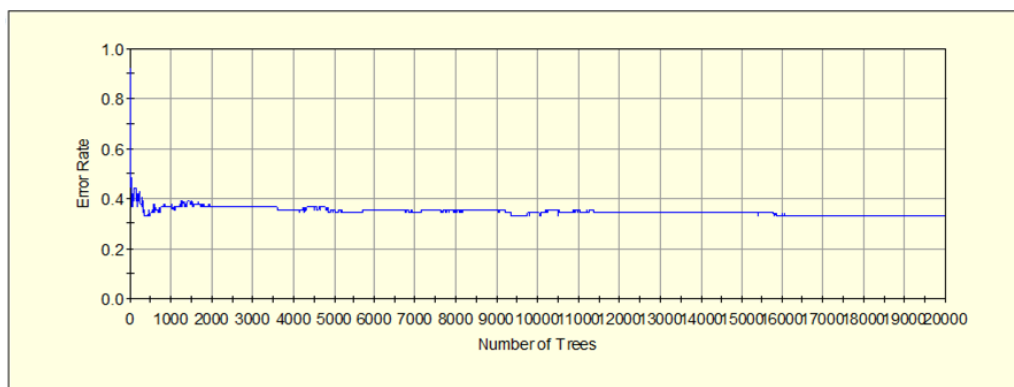


Figure 5.8. Error plot for the Random Forests classification model trained using 2-D and 3-D molecular descriptors of the molecules present in the dataset. Blue line shows the evolvement of overall OOB error of prediction with the addition of number of trees.

The error plot in Figure 5.8 provides an overall OOB error of prediction and prediction accuracy for the classification model and a confusion matrix in Table 5.1 provides information on the prediction accuracy and OOB error rate associated with each class. Data in Table 5.1 confirm that the RF model has

predicted the number of molecules in each class with similar percentage accuracy.

Table 5.1. Confusion matrix generated by Random Forests for classification of dataset of 382 molecules. Class 1 represents the molecules which crystallised and class 2 represents the molecules which did not crystallise.

<b>Actual Class</b>	<b>Total Cases</b>	<b>% Correct</b>	<b>1, N=231</b>	<b>2, N=151</b>	<b>Class Error</b>
1	303	67.7	205	98	32.34
2	79	67.1	26	53	32.91

The RF program computed the proximities between pairs of molecules which were then scaled down into two dimensions using MDS. In contrast to PCA, the MDS shows some distinct clustering. The MDS plot of scaling coordinates 1 vs. 2 (Figure 5.9), obtained from the proximity matrix generated by RF/Salford showed two distinct zones belonging to two classes and an overlapped zone which comprises molecules from both the classes (Figure 5.9). This model can be further used to predict the crystallisability of molecules from the same series for which no crystallisation attempts have been made. The chances of accurate crystallisability prediction for these molecules would be ~70%.

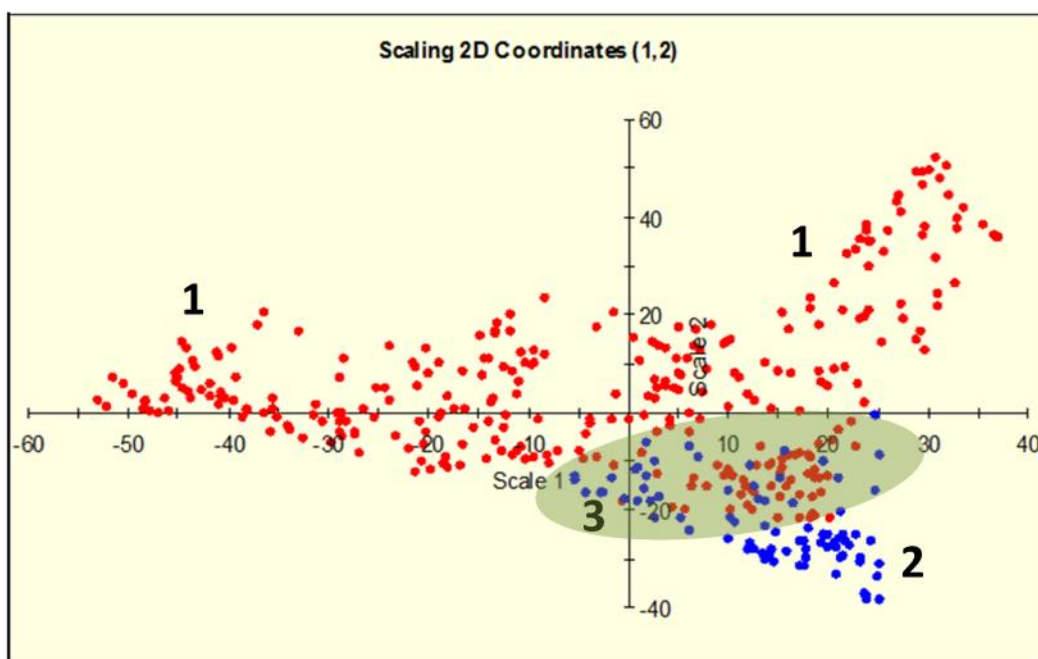


Figure 5.9. The multidimensional scaling plot of the scaling coordinates (1,2) obtained from the Random Forests classification proximity matrix. Each of the 382 points on the multidimensional scaling plot represents a molecule from the dataset and is coloured according to crystallisation outcome: class 1 (red) and class 2 (blue). 1 and 2 denote the two separate zones correspond to molecules from class1 and class 2. 3 (encircled in green) denotes the overlapped zone with molecules from both classes.

The MDS plot (Figure 5.9) highlights the overlapped region, where molecules from both the classes coexist. Therefore, a convex hull plot which is an alternative to the MDS plot and offers a useful representation of large datasets with a considerable overlap of points between them (Vidmar and Pohar, 2005) was generated. The convex hull plot shows that the molecules in class 2 were confined in a limited space of the plot indicating a common set of features that is consistent with poor crystallisability while molecules in class 1 were present across a larger area (Figure 5.10).

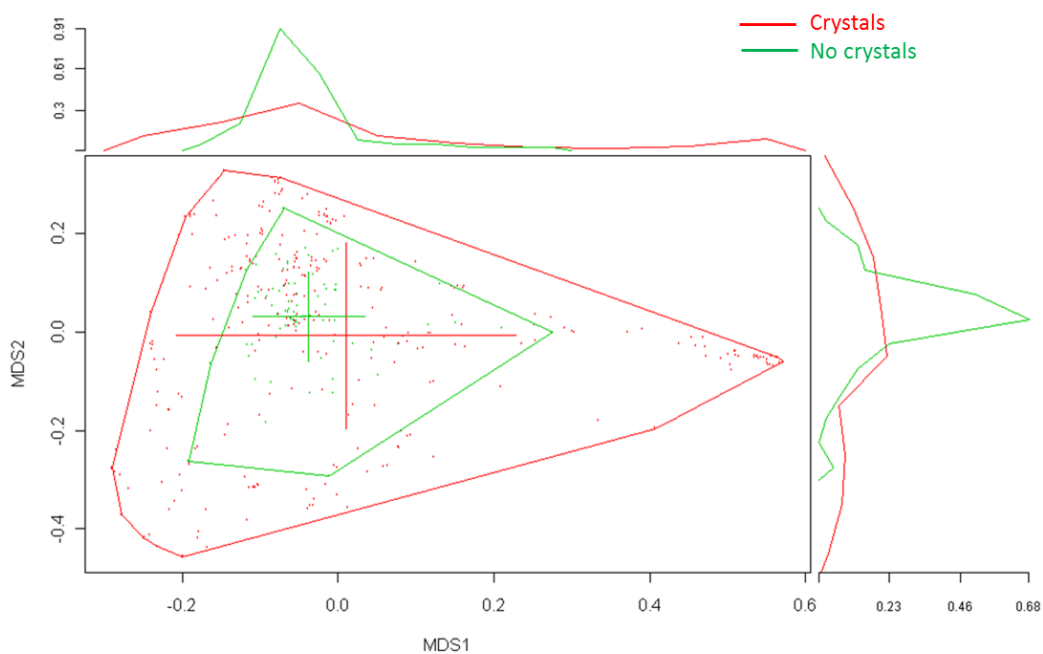


Figure 5.10. Convex hull plot of scaling coordinates obtained from Random Forests proximity matrix. Molecules in class 1 and 2 are represented by red and green points respectively. The cross sign is the mean of MDS1 and MDS2 for each group.

The top 10 most important molecular descriptors responsible for this classification model include the descriptors that describe the relative energies of the molecules, their atoms connectivity, molecular conformations and number of rotatable bonds in the molecules (Table 5.2).



Table 5.2. Ten most important descriptors selected by Random Forests for classification model. Reproduced from web reference

<http://www.juit.ac.in/attachments/podophyllotoxin/introduction.htm>

<b>Descriptor</b>	<b>Rank</b>	<b>Descriptor Definition</b>	<b>Further Details of Descriptors</b>
E_TOR	1	Torsion (proper and improper) potential energy	The energy descriptors use the MOE potential energy model to calculate energetic quantities from stored 3D conformations
E_VDW	2	van der Waals component of the potential energy	
GCUT_SMR_0	3	Atomic contribution to molar refractivity	The GCUT descriptors using atomic contribution to molar refractivity instead of partial charge
AM1_EELE	4	Electronic energy calculated using the AM1 Hamiltonian	Can be calculated from the connection table (with no dependence on conformation) of a molecule
BCUT_PEOE_3	5	Calculated from Adjacency Matrix of a chemical structure	It is calculated from the eigenvalues of a modified adjacency matrix.
B_1ROTR	6	Fraction of rotatable single bonds: b_1rotN divided by b_count.	The atom count and bond count descriptors are functions of the counts of atoms and bonds
VSURF_CW2	7	Capacity factor	Depend on the structure connectivity and conformation
E	8	Value of the potential energy	Same description as for descriptors ranked from 1-2
B_ROTTR	9	Fraction of rotatable bonds: b_rotN divided by b_count.	Same description as for descriptors ranked at 6
GCUT_SLOGP_1	10	Atomic contribution to logP	The GCUT descriptors using atomic contribution to logP instead of partial charge.

### **5.3.3 Model Optimisation Attempts**

The RF model was retrained after removing the least important descriptors and was evaluated for prediction accuracy using OOB error of estimate. This process was iteratively repeated and from every training model the least important descriptors were removed. However, no significant improvement in the overall/class-wise prediction accuracy was observed. RF model trained using only 33 descriptors (1/5<sup>th</sup> of the total number of descriptors) gave an overall OOB error rate of estimate of 32.3% (for class 1-31.7% and for class 2-32.9%). The predictive accuracy of this model is very close to the starting model developed using 151 descriptors. This suggests that only 33 descriptors are required to develop the classification model.

To further improve the predictive accuracy of this model, it was subjected to a backward elimination algorithm. During this, only one descriptor was removed at a time and the model was trained using the remaining descriptors. However, no further improvement was observed in the classification model.

### **5.3.4 Important Descriptors Assessment**

Achieving a ~70% accuracy in the predictive model provides a useful indication based on its ability to identify problematic systems with reasonable confidence thereby flagging problematic compounds and identifying early targets for more detailed efforts to improve. The most important descriptors listed by RF showed that the torsion energy, van der Waals/steric energy, structure connectivity,

conformation and numbers of rotatable bonds in the molecules are contributing towards different crystallisation behaviour.

Some of these descriptors represent the conformational flexibility in the molecule and it is known that conformational flexibility plays an important role in reducing the crystallisation tendency of the molecules (Yu et al., 2000). For molecules containing long alkyl chains, multiple conformations would be present in crystallising media and only one would fit into the crystal lattice (Yu et al., 2000). In the original report (Hursthouse et al., 2009), similar trends were observed for molecules i.e. crystallisation tendency was reduced on increasing the length of the alkyl chain. Propionanilide and butylanilide derivatives (Figure 5.11) had poorer crystallisation tendencies as compared to acetanilide and trimethylacetanilide derivatives.

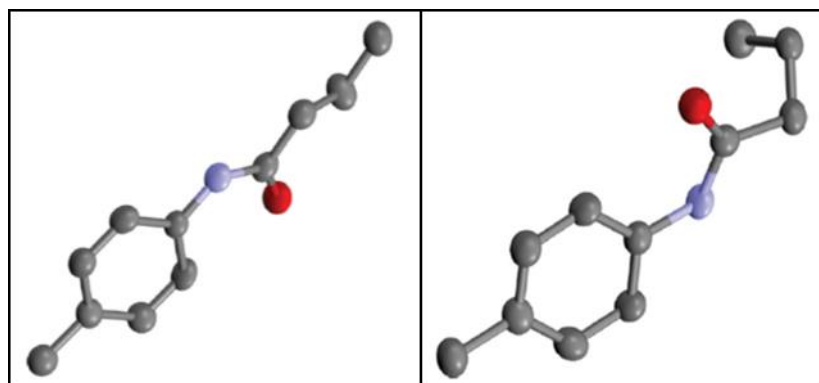


Figure 5.11. Two different conformation of propyl chains in 4-methylbutylanilide. Reproduced from reference (Hursthouse et al., 2009).

For some of the acylanilide derivatives, it was also observed that para-substituted derivatives were easier to crystallise than the ortho-substituted derivatives, which in turn were crystallised more easily than meta-substituted

derivatives. The possible explanation for this is that the para-substituted derivatives have only one major conformation as compared to two conformations for meta-substituted derivatives in the crystallising media (Figure 5.12) which in turn decreases the crystallisation tendency presumably due to the requirements of orientation of the molecules to fit into the growing lattice.

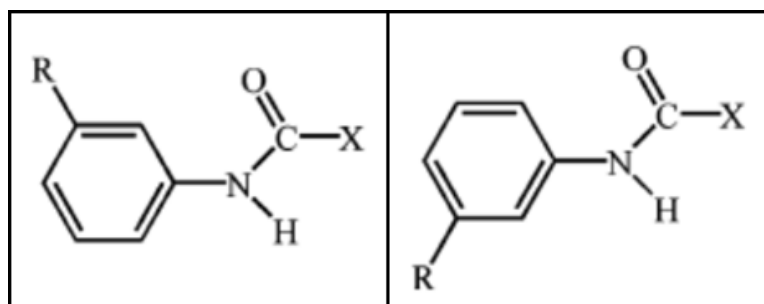


Figure 5.12. Alternate conformations of meta-substituted acylanilides. The rotation is around aromatic carbon to N bond. Reproduced from reference (Hursthouse et al., 2009).

Ortho-substituted acylanilide derivatives have a higher tendency of crystallisation compared to meta-substituted acylanilide derivatives. This is due to the steric effects of substituents at ortho position which tend to favour a particular conformation in solution and in the solid-state (Hursthouse et al., 2009). It is pleasing that the molecular descriptors identified by the RF methods are consistent with chemical and structural expectations and/or knowledge leading to confidence in this method.

### 5.3.5 Limitations of Random Forests Classification Model

This model is based on 2-D and 3-D molecular descriptors of the organic molecules and completely ignores the effect of impurities (Hursthouse et al., 2009), solvents and crystallisation conditions (RH, rate of solvent evaporation etc.). Unfortunately, the information about crystallisation conditions has not been provided in the original report and therefore, cannot be incorporated in the RF model in an attempt to improve its predictive ability.

The solubility of the compound in different solvents can vary significantly and affects the appearance of crystals. Very high solubility may lead to increased viscosity and possibly gums/oils whilst inadequate solubility may lead to extreme dilute solutions which rarely give large crystals. A limited range of solvents and crystallisation conditions have been used for the crystallisations of these molecules in the original report. The experimental results may vary after attempting the crystallisation in a wider range of solvents and crystallisation conditions.

This dataset is taken from the literature and given the aim of the original study was a structural systematics investigation, the systematic effort towards crystallisation may have been limited. The experiments were done under similar conditions but may not have been tightly controlled leading to changes in concentration, supersaturation due to temperature or evaporation rate fluctuations. All these factors might have an effect on the crystallisation outcome.

Impurities often play a role in inhibiting crystal growth and phase transformation (Blagden et al., 1998). The crystallisation behaviour may be different under different sets of conditions used for synthesis and crystallisation. These molecules are likely to have major/minor impurities and no information has been provided in the original report about the nature and amount of impurities and therefore is not considered in the RF classification model.

## 5.4 Summary

Statistical tools were applied to predict the crystallisability of a set of organic molecules. RF classification method has provided better results than PCA and a RF model with prediction accuracy of ~70% has been built. This kind of model can potentially identify the problematic systems with reasonable confidence, allowing early efforts towards improvement. This RF predictive model takes into account only the molecular descriptors of the organic molecules. Therefore, crystallisability prediction may fail when the other factors such as solvent, supersaturation, RH etc. are responsible for the crystallisation outcome.

Identification of physicochemical descriptors suitable for any statistical analysis is a persistent challenge. Crystallisation is known to be influenced by shape, surface and ease of conformational change. In this study, RF has correctly identified the descriptors responsible for varied crystallisation propensities, leading to confidence in the predictive model. Identification and selection of critical descriptors responsible for varied crystallisation outcome by RF model is a potentially useful tool.

## **6 Exploring the Crystal Structure Landscape of Olanzapine**



## 6.1 Introduction

Physical form screening of olanzapine (OZPN, C<sub>17</sub> H<sub>22</sub> N<sub>4</sub> S<sub>1</sub>, Figure 6.1) which belongs to group 1 (see Section 2.1) using HTCAA methodology has been described in Chapter 4. This chapter describes a comprehensive physical form screening of OZPN using various crystallisation techniques. In addition to a rigorous experimental search for solid-forms of OZPN, a CSP study was used as a complement to experimental screening and to interrogate the experimental results.

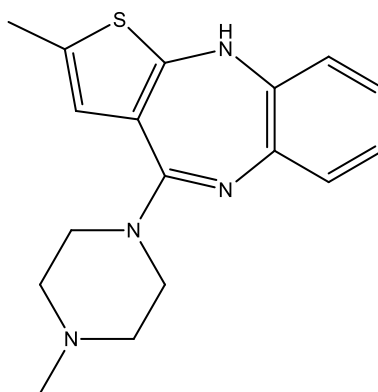


Figure 6.1. Molecular structure of olanzapine.

OZPN, is a benzodiazepine derivative with demonstrated efficacy in the treatment of schizophrenia and related psychoses (Fulton and Goa, 1997; Tollefson et al., 1997; Sanger et al., 2001). OZPN has been reported to display a number of solid-state forms including three polymorphic non-solvated forms; form I - CSD (Frank, 2002) refcode: UNOGIN01 (Reutzel-Edens et al., 2003; Wawrzycka-Gorczyca et al., 2004a); form II [form II in this work has been referred to as form IV (CSD refcode: UNOGIN02) previously by (Thakuria and

Nangia, 2011)]; and form III whose crystal structure has not been determined to date.

The crystal structure of form II was determined from a single crystal grown from the vapour phase as part of a larger crystallisation screen described in two published reports (Miglani et al., 2011; Bhardwaj et al., 2013) and in this thesis. This structure is consistent with the characteristic powder data for form II reported by Reutzel-Edens and co-workers (Reutzel-Edens et al., 2003) as evidenced by a Pawley type refinement ( $R_{wp} = 3.9$ ), confirming the assignment of this form as form II. During the course of our study, we were made aware of the structure report of form IV of OZPN obtained from a failed co-crystallisation attempt of OZPN with nicotinamide. This structure is isostructural with the form II structure here and a Pawley type refinement of the reported form IV unit cell against the form II powder data reported in research article by Reutzel-Edens and co-workers yielded an  $R_{wp} = 2.461$ . On this basis we conclude the reported form IV structure is actually form II, following the nomenclature of the published research article by Reutzel-Edens and co-workers.

Across the various reports to date on the polymorphism of OZPN, inconsistent nomenclature for the polymorphs of OZPN has been used (Wawrzycka-Gorczyca et al., 2004a; Polla et al., 2005; Tiwari et al., 2007). In this thesis, we are following the nomenclature of Reutzel-Edens and co-workers, specifically forms I, II and III are defined as per the characteristic data shown in Table 6.1.

Table 6.1. Nomenclature for non-solvated polymorphs of olanzapine and their characteristic data [this work and from reference (Reutzel-Edens et al., 2003)].

Form	Unit cell Parameters <sup>a</sup>	XRPD Peaks <sup>b</sup>	DSC <sup>c</sup>	Raman Bands (cm <sup>-1</sup> ) <sup>d</sup>	Solid State <sup>13</sup> C NMR Chemical Shifts <sup>e</sup> (Reutzel-Edens et al., 2003)
I	a = 10.328(3), b = 14.524 (3), c = 10.500(3), β = 100.606(14) vol =1548.2(7)	8.639, 12.463, 17.050, 19.834, 21.507, 22.302	Melt at 468 K	375.5, 1521, 1561, 1579, 1594,	16.7(C16), 46.0(C17), 48.4(C12,C15), 121.8(C11), 142.7(C9), 155.5(C3)
II	a = 9.8544(14), b = 16.314(2), c = 9.9754(12), β = 98.304(8), vol =1586.9(4)	9.002, 12.928, 13.975, 15.914,	Transforms to form I on heating at <i>ca.</i> 454 K prior to finally melting at 468 K	369.9, 391.8, doublet at 1510 and 1524, 1560, 1586	17.01(C16), 45.5(C17), 47.7, 50.1(C12,C15), 118.1(C11), 140.9(C9), 157.2(C3)
III	----	8.890, 14.306, 15.017, 19.269, 20.842, 22.117	-----	-----	14.9(C16), 45.2(C17), 47.6, 50.1,(C12,C15), 117.1(C11), 142.7(C9), 157.2(C3)

<sup>a</sup> Single crystal collected during this study at 123 K, <sup>b</sup> Characteristic low angle peaks identified using high resolution capillary powder diffraction data, <sup>c</sup> DSC measurements of pure samples of form I and II, <sup>d</sup> Raman data collected from single crystal samples of form I and II, <sup>e</sup> Chemical shifts are presented as ppm.

Other reported solid forms of OZPN include three polymorphic dihydrates referred to as D (AQOMAU), E (AQOMAU02), and B (AQOMAU01) (Reutzel-Edens et al., 2003), later being equivalent to disordered hydrate, AQOMEY01 (Wawrzycka-Gorczyca et al., 2007), a 'higher hydrate' (AQOMEY) (Reutzel-Edens et al., 2003), 6 single and mixed organic solvates [Dimethylsulfoxide (DMSO) hydrate: CAYTUS (Polla et al., 2005), methanol hydrate: ELEVOG (Capuano et al., 2003), methanol: UNOGOT (Wawrzycka-Gorczyca et al., 2004b), 2-butanol hydrate: WEXPUL, ethanol hydrate: WEXQEW and DCM solvate: WEXQAS (Wawrzycka-Gorczyca et al., 2007)] and 3 co-crystals (Clarke et al.,

2012). To date, anhydrous form III has only been obtained as a mixture with polymorphs I and II which has precluded structure determination attempts and the structure of this form remains unknown (Reutzel-Edens et al., 2003; Bhardwaj et al., 2013). The key aim of this work was to investigate the factors controlling the solid-state diversity of this important drug.

Calculating the crystal energy landscape has been used as a complement to experimental polymorph screening for small molecules, directing the search for specific new forms (Cross et al., 2002; Arlin et al., 2011), and providing complementary atomic-level insights and more confidence in the interpretation of the experimental data (Price, 2008a; Braun et al., 2010), including placing of protons in structures solved from XRPD data (Braun et al., 2012). The 2010 CCDC blind test of crystal structure prediction (Bardwell et al., 2011) included a highly flexible and drug-like molecule benzyl-(4-(4-methyl-5-(p-tolylsulfonyl)-1,3-thiazol-2-yl)phenyl)carbamate;  $C_{25}S_2N_2H_{22}O_4$  with 8 exocyclic torsion angles and forced the development of computational methodologies capable of successfully tackling such molecules (Kazantsev et al., 2011b). Hence, it is now timely to see how the crystal energy landscapes of pharmaceutical drugs can aid experimental solid form screening.

Larger molecules, with the conformational flexibility to adopt a range of nearly isoenergetic molecular shapes, introduce new questions in interpreting crystal energy landscapes, as well as being more challenging to calculate accurately. In particular, the ability to rearrange both the conformation and the packing in condensed phases is likely to be significantly more difficult than for small

molecules which can have many densely-packed structures with small barriers to rearrangement in condensed phases (Hulme et al., 2007). CSP study of the flexible OZPN molecule was aimed at providing a theoretical basis for physical form diversity in this important pharmaceutical compound and how the crystal energy landscapes of flexible pharmaceuticals can complement experimental approaches.

Molecular packing analysis of known structures in the CSD belonging to the OZPN family using XPac (Gelbrich and Hursthouse, 2005) and Mercury (Macrae et al., 2006) revealed that they all contain a common non-hydrogen bonded dimeric unit, with enantiomer pairs sitting on a symmetry centre (Figure 6.2).

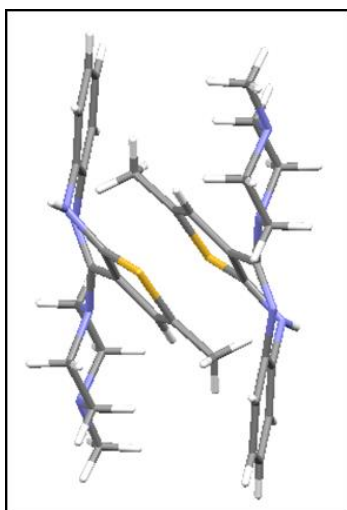


Figure 6.2. Centrosymmetric dimer observed in all the experimental crystal structures of olanzapine.

The centrosymmetric dimer is stabilised by the van der Waals interaction between thiophene of one OZPN molecule with piperazine ring of second OZPN molecule (Wawrzycka-Gorczyca et al., 2007). This motif is present in all the

reported OZPN structures and is believed to be the structural building block for OZPN crystal structures. Thus it is of interest to develop a better understanding of the factors controlling the solid-state packing of OZPN and to find out whether any alternative packing motifs are possible in other experimental forms found in this study or in the computed structures from CSP. Another aim of the study of this compound was to compare its overall extent of solid-state diversity with that of clozapine (CZPN), which is another structurally related compound in group 1, and the factors governing those findings (see Chapter 7).

## **6.2 Experimental Procedures**

### **6.2.1 Crystallisation**

The experimental screen was based on solution crystallisations using 71 solvents selected to cover a wide range of physicochemical properties. These were supplemented by a further extensive range of crystallisation techniques to maximise the range of conditions under which crystallisation of OZPN was affected (see Table 6.2). A total of 417 individual crystallisations were carried out. General crystallisation techniques have been described in Chapter 3; other additional techniques which have been used specifically for OZPN are mentioned here.

Following recrystallisation, each sample was analysed using XRPD/SXD, Raman and thermal analysis to confirm sample identity as described in the Chapter 3.

Table 6.2. Summary of manual crystallisation conditions used in the experimental search.

<b>Crystallisation Techniques</b>	<b>Crystallisation Method</b>	<b>Number of Crystallisations</b>
Solution crystallisation $T_{\text{sat}} = \text{RT}$	Evaporation in vials, on watch glass at RT and at 277 K	128
Solution crystallisation $T_{\text{sat}} = 343 \text{ K}/323 \text{ K}$	Cooling to 323 K/298 K with agitation 850 rpm	75
	Antisolvent crystallisation	60
	Neat/liquid-assisted grinding	35
	Crystallisation as function of pH	18
	Crystallisation from vapour phase	10
	Crystallisation on functionalised surfaces	12
Desolvation of solvates	Isothermal and <i>in-situ</i> desolvation at 333 K and 318 K Desolvation under vacuum at RT	66
Melt quenching		1
Recrystallisation from amorphous		5
Spray drying		3
Freeze drying		4

$T_{\text{sat}}$  = temperature at which the saturated solutions were prepared

### 6.2.1.1 Cooling Crystallisation of Olanzapine using Automated Platform

These crystallisations were carried out on Chemspeed Accelerator SLT100 Platform. Solutions were prepared by adding excess solid to each solvent to ensure that solutions were saturated at the appropriate temperature prior to being filtered automatically into a clean crystallisation vessel.  $T_{\text{sat}}$ , the temperature at which solutions were prepared was 343 K/323 K /298 K and the solutions were cooled to 323 K/298 K at a cooling rate of approximately 5 K  $\text{min}^{-1}$ . Crystallisation in filtered solutions was induced by either controlled cooling or evaporation. All the experiments were done in duplicate.

### **6.2.1.2 Antisolvent Crystallisation of Olanzapine**

Saturated solutions of OZPN were prepared by dissolving the starting material in the chosen solvent under agitation. After equilibration, solutions were filtered to remove seeds and three volume equivalents of the respective antisolvent (cyclohexane, heptane, diisopropyl ether) was added.

### **6.2.1.3 Crystallisation of Olanzapine as Function of pH**

Various salts were synthesised by grinding the compound with the respective base in molar ratio of 1:1 together with a drop of isobutyl acetate in a Retsch MM400 mixer mill (20 min at 20 Hz). Crystallisation of compound by disproportionation of each salt was carried out in acetone/water adjusted to pH 8-12 using ammonium hydroxide or aqueous or methanolic sodium hydroxide.

### **6.2.1.4 Crystallisation of Olanzapine on Functionalised Surfaces**

Saturated solution of OZPN was prepared in ethyl acetate at RT and the derivatised glass slides were kept in the vials at 45° angle completely submerged in the saturated OZPN solution.

Derivatised glass slides were prepared by following procedure: Glass slides were kept in Pyrana solution (H<sub>2</sub>SO<sub>4</sub> and H<sub>2</sub>O<sub>2</sub>) for 45 minutes and washed using distilled water and dried under nitrogen air. These slides were then placed in 5, 10, 20 and 20% v/v solution of APTES ((3-Aminopropyl)triethoxysilane), MPTMS ((3-Mercaptopropyl)trimethoxysilane), TPS (trimethoxy(propyl)silane) and GPTMS ((3-glycidoxypropyl)



trimethoxysilane) respectively in toluene for 12 hours. The slides were sonicated in toluene, dimethylformamide and distilled water for 20 minutes each. After sonication, these slides were dried under nitrogen. The uniformity of self-assembled monolayer was assessed by contact angle measurement of small drops (2-3  $\mu\text{L}$ ) of water (Millipore, 18.2 M  $\Omega$  cm) using Krüss DSA30B goniometer (Krüss, GmbH, Germany).

#### **6.2.1.5 *Spray Drying of Olanzapine***

Preliminary investigation using spray drying techniques was carried out for physical form screen of OZPN. Spray-dried samples of OZPN were prepared at various concentrations (0.25, 0.5 and 1%) in different solvents (dichloromethane (DCM) and methanol). Solutions were spray dried under nitrogen atmosphere, 600 psi/4 kbar using a Buchi Mini Spray Dryer B-290 (BUCHI Labortechnik AG, Flawil, Switzerland). Resulting samples were analysed using XRPD within 1 hour of their preparation.

#### **6.2.1.6 *Recrystallisation from Amorphous Olanzapine***

Amorphous sample obtained from melt quenching was left as it is (on the same slide which was used for melt quenching) or was scratched gently from the slide and slightly triturated before filling into a 0.7 mm borosilicate capillary or in vials. These capillaries were then analysed using XRPD at constant temperature or VT.

## 6.2.2 Variable Temperature-X-ray Powder Diffraction

Temperature was controlled using an Oxford Cryosystem 700 Plus series device, with a cryostream cold head with a nozzle suspended at the top of, and parallel to, the capillary. Gas at the required temperature flows through the isothermal nozzle out to the capillary. In some cases, the capillary was shifted directly to the required temperature, while in other cases, especially *in-situ* desolvation, temperature was increased at a rate of 5 K min<sup>-1</sup> following each data collection.

### 6.2.2.1 Desolvation of Olanzapine Solvates

Isothermal desolvation of various OZPN solvates was carried out at 333 K and 318 K in a convection oven. Selected solvates were subjected to desolvation in vacuum at RT and *in-situ* desolvation at 313 K using XRPD.

## 6.2.3 XPac Analysis of Crystal Structures of Olanzapine

The molecular shape of the OZPN was parameterised using six common ordered sets of points (COSPs) shown in Figure 6.3. The calculations were performed using standard cut-off parameters. The use of additional COSPs, including the selection of the whole molecule, was also tried but did not result in an appreciable improvement in the comparison. The base vectors correspond to a 1-D arrangement of molecules and are defined by their direction relative to the three crystallographic axes along with the repeat distance. Supramolecular construct (SC) can be defined as a recurring periodic or discrete arrangement of molecules with unique spatial characteristics.

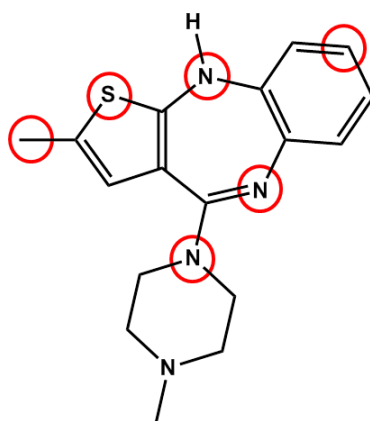


Figure 6.3. Common ordered sets of points selected for XPac analysis.

### 6.2.4 Random Forests Classification Model of Olanzapine Solvates

Random Forests (RF) (see section 5.1.4) classification was performed using the 'RandomForest library package' available in statistical computing and graphics environment R (version 2.11.1) (R Development Core Team, 2006) denoted as RF/R and a commercially available RandomForest® package (Salford Systems) represented as RF/Salford. The training data comprised of 250 numerical physicochemical solvent descriptors calculated using MOE (MOE, 2002). Three known experimental outcomes (three packing types) were defined as 1, 2 and 3 corresponding to SC<sub>31</sub>, SC<sub>32</sub> and SC<sub>33</sub> (described later in Section 6.7). The RF model was trained using all the 250 numerical descriptors and the molecular packing type outcome using the following parameters: ntree = 15,000, mtry = 15, nodesize = 1, seed = 45. 24 out of 28 solvates were taken for random forest classification as there were no calculated descriptors for 3 solvents (1,4-butanediol, t-butanol and ethylene glycol (EG)) and OZPN hydrate was considered only once in the molecular packing type SC<sub>32</sub>.

### 6.3 Calculation and Analysis of the Crystal Energy Landscape (Crystal Structure Prediction) of Olanzapine

The crystal energy landscape was generated using CSP methods (Day, 2011) based on quantum mechanical calculations on the single molecule, to provide conformational energy differences  $\Delta E_{\text{intra}}$ , and the molecular charge distribution. This molecular charge distribution is analysed to give the model for the intermolecular electrostatic forces, which in combination with an empirical exp-6 repulsion-dispersion model is used (Price et al., 2010) to evaluate the intermolecular lattice energy  $U_{\text{inter}}$ . The relative stability of the different crystal structures is estimated from the lattice energy,  $E_{\text{latt}}=U_{\text{inter}}+\Delta E_{\text{intra}}$ , which is the energy of the static crystal relative to a gas of the molecules in their lowest energy conformation.

The search for possible crystal structures has to consider the entire range of conformations which the molecule could adopt to give a favourable crystal structure, balancing specific intermolecular interactions such as hydrogen bonding with the dispersion stabilisation from a densely packed crystal structure. *Ab initio* calculations on OZPN showed that the accessible conformational space naturally separated into two low energy regions as defined in Figure 6.4.

Conformations outside these two regions (Figure 6.5) are too high in energy to be compensated for by favourable intermolecular interactions in the crystal and so need not be considered in the search. Whether solvent molecules could lower

the conformational energy barrier so that the molecules could pass between regions A and B in solution, is a different question (Abraham et al., 1974).

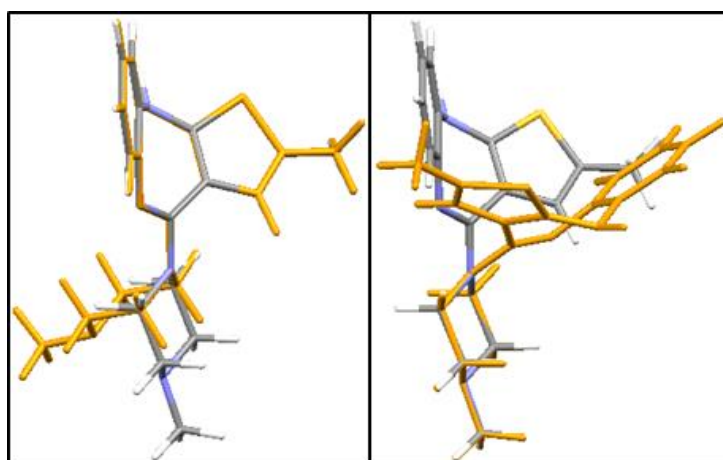


Figure 6.4. The difference between the observed conformation (region A, coloured by element) and the alternative low energy conformation (region B, coloured by orange) depicted by overlaying (left) the benzodiazepine or (right) the piperazine rings, of the gas phase optimised conformations.

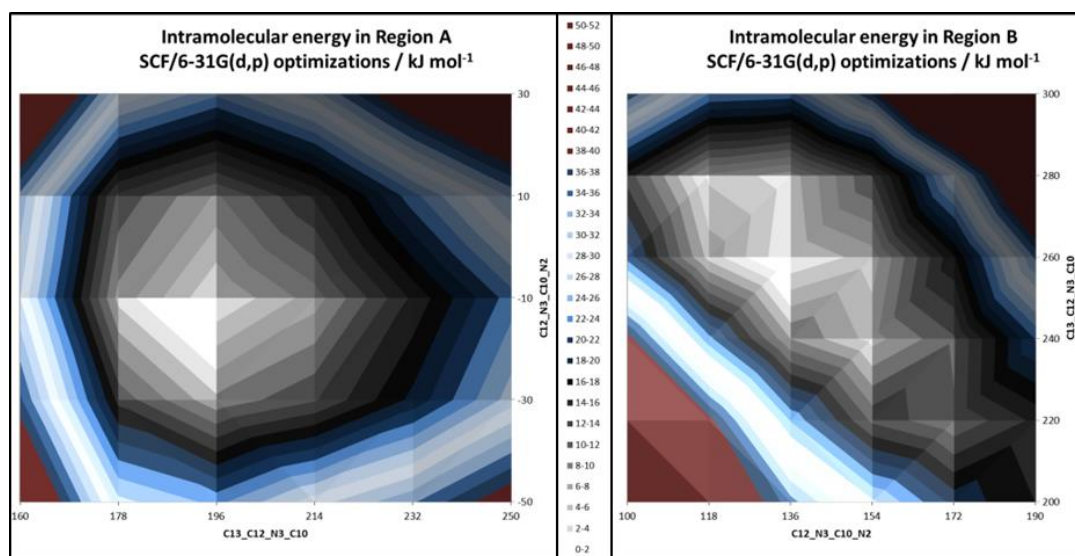


Figure 6.5. The CrystalPredictor grid of intramolecular energies  $\Delta E_{\text{intra}}$ , in  $\text{kJ mol}^{-1}$  relative to the lowest energy point in each region shown in Figure 6.4.

Two separate grid calculations were carried out with the angle ranges described in Table 6.3. These were done with full optimisation at the SCF 6-31G(d,p) level, with a single point charge calculation at the MP2 6-31G(d,p) level.

Table 6.3. The angles considered in the CrystalPredictor grid generation.

Angle	Range in A region	Range in B region
C12_N3_C10_N2	-50° to +30° (5 points / 20° steps)	+100° to +190° (6 points / 18° steps)
C13_C12_N3_C10	+160° to +250° (6 points / 18° steps)	+200° to +300° (6 points / 20° steps)

Approximately a million crystal structures in each conformational region were generated using CrystalPredictor (Karamertzanis and Pantelides, 2007) in the space groups *P1*, *P-1*, *P2<sub>1</sub>*, *P2<sub>1</sub>/c*, *P2<sub>1</sub>2<sub>1</sub>2*, *P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>*, *Pna2<sub>1</sub>*, *Pca2<sub>1</sub>*, *Pbca*, *C2/c*, *Cc* and *C2*, the 12 most common space groups (observed for molecules of this size (Kazantsev et al., 2011b) in the CSD). This includes chiral as well as centrosymmetric space groups, as although OZPN does not have a chiral centre and is observed in both mirror images, we do not wish to assume that the molecule could not invert in solution or gas phase to crystallise in just one hand. The relative energies of the resulting *ca.* 13,000 unique structures were improved by using a distributed multipole model for the electrostatic interactions. The lowest 1150 unique low energy structures, within 30 kJ mol<sup>-1</sup> of the global minimum for each region, were optimised to minimise their lattice energy,  $E_{\text{latt}}$ , with CrystalOptimizer allowing all the torsion angles shown in Figure 6.6 to change along with the crystal cell parameters and relative positions and orientations of the molecules. The intramolecular conformational

energy was calculated at the PBE0/6-31G(d,p) (Kazantsev et al., 2011a; Kazantsev et al., 2011c) level, and these charge densities provided the electrostatic model for  $U_{\text{inter}}$ .

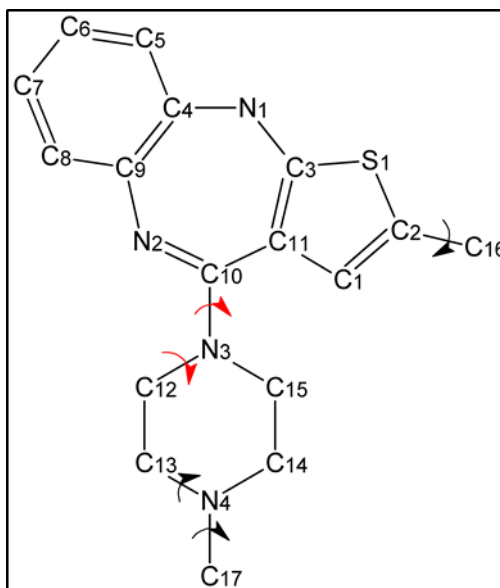


Figure 6.6. The molecular structure of olanzapine, showing all degrees of flexibility considered in final CrystalOptimizer refinements (C<sub>11</sub>-C<sub>10</sub>-N<sub>3</sub>-C<sub>12</sub>, C<sub>10</sub>-N<sub>3</sub>-C<sub>12</sub>-C<sub>13</sub>, H-C<sub>17</sub>-N<sub>4</sub>-C<sub>13</sub>, C<sub>17</sub>-N<sub>4</sub>-C<sub>13</sub>-C<sub>12</sub> and H-C<sub>16</sub>-C<sub>2</sub>-C<sub>1</sub>).

All other terms in the intermolecular energy were estimated by empirically fitted isotropic exp-6 atom-atom potential with the FIT parameters (Price et al., 2010) choosing the sulphur parameters fitted to thioether and thiocarbonyl groups. As a final refinement of the lattice energy, the effect of an average organic crystalline environment (Cooper et al., 2008) was approximated by using a Polarizable Continuum Model (PCM) (Cossi et al., 2002) with  $\epsilon=3$ , to calculate the molecular charge density and conformational energy  $\Delta E_{\text{intra}}$ .

## 6.4 Results and Discussion

Across all crystallisation conditions, three non-solvated polymorphs, 56 solvates and an amorphous form were obtained (Table 7.3). No evidence for non-solvated polymorphs other than forms I, II and III was observed.

### 6.4.1 Forms I, II and III of Olanzapine

Form I was most frequently observed under all conditions and the only non-solvated form which could be obtained directly from solution recrystallisation as a pure phase. Form II single crystals were obtained through vapour phase via sublimation of OZPN and XRPD analysis of the experimental samples revealed a mixture of form II and III. No suitable single crystals of form III were obtained. Form II, in presence of small amount of form III, was also obtained by neat grinding of form I in a Retsch mixer mill MM400. Form III was also observed, albeit as a mixture with form II, from desolvation of DCM, ethanol and 2,2,2-trifluoroethanol solvates, and the 1-butanol solvate hydrate.

An overlay of the XRPD data for form I, II (simulated from RT single crystal structure) and a mixed phase sample of III and II (obtained by desolvation of the DCM solvate) in which form II is a minor component is shown in Figure 6.7. All three polymorphs can be readily distinguished by the low-angle reflection positions in their XRPD patterns, however detailed comparison of forms II and III is complicated by the lack of a pure form III sample. Since the low angle peak at  $10.84^\circ$   $2\theta$  appears in XRPD data previously reported (Polla et al., 2005; Tiwari



et al., 2007) for form II, Figure 6.7 suggests that the samples used were actually mixtures of form II and III.

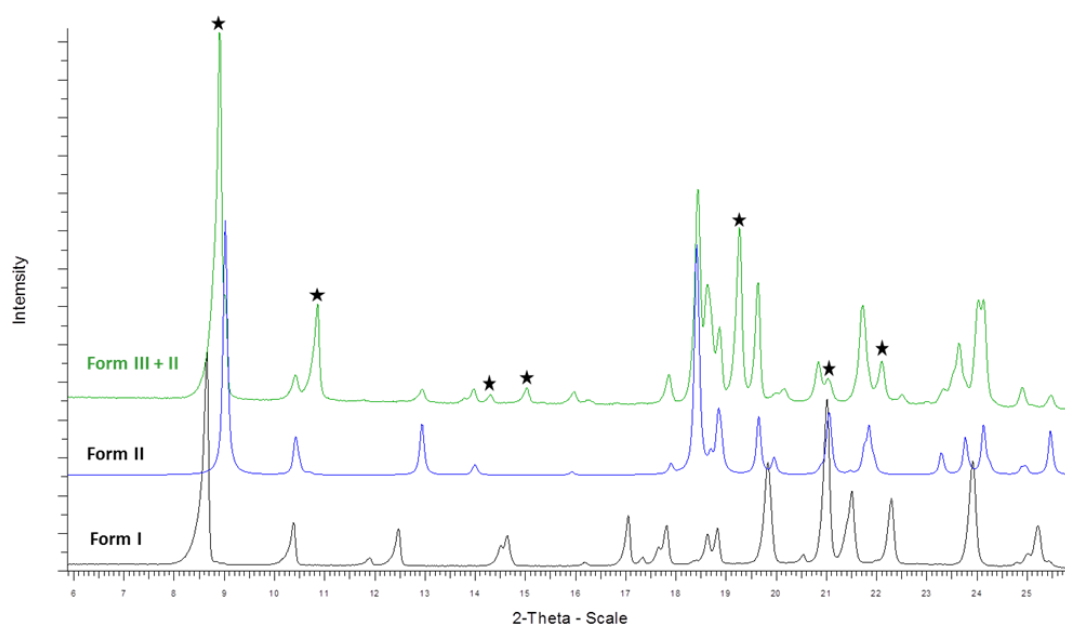


Figure 6.7. Overlay of XRPD patterns of form I (experimental, black), form II (simulated from room temperature single crystal structure, blue) and III + II (experimental, green) of olanzapine. Characteristic peaks of form III are marked.

DSC data collected at  $10 \text{ K min}^{-1}$  shows that form I melts at 468 K and form II exhibits a transformation to form I via melt/recrystallisation with an onset temperature of 454 K prior to melting at 468 K (Figure 6.8). The mixed phase form III produces a similar DSC trace except that the endotherm corresponding to the transformation of the sample to form I has a lower onset at *ca.* 443K. Due to the presence of form II in the form III sample, it was not possible to accurately estimate the transition temperature for form III to form I based on this measurement alone.

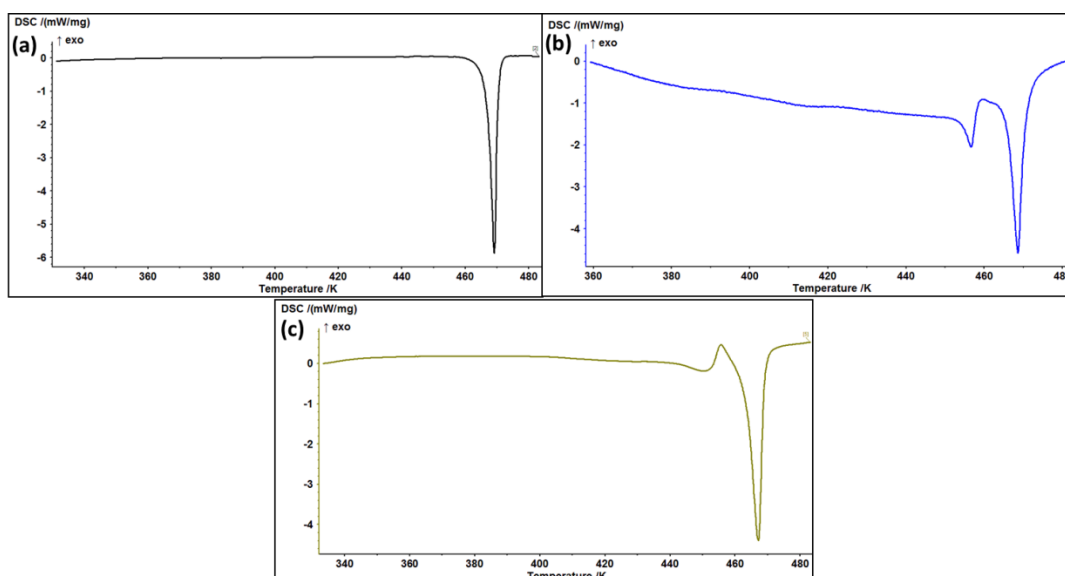


Figure 6.8. DSC thermograms from samples of olanzapine (a) form I, (b) form II, sample comprising of several single crystals, (c) form III + II.

Whilst metastable to form I, forms II and III are relatively stable in the solid-state at room temperature, with no evidence of transforming after 10 months. At 313 K and above, forms II and III transform rapidly to form I, typically within 48 hours. Slurries of mixed phase samples of II and III in non-solvate forming solvents transform completely to form I within two hours at RT.

Raman spectroscopy allows unambiguous identification of phase pure form I and form II. Differences were observed between C-H bands in high frequency zone ( $2700\text{-}3100\text{ cm}^{-1}$ ), C-C (aromatic) band in medium frequency zone ( $1500\text{-}1700\text{ cm}^{-1}$ , peak splitting at  $1525\text{ cm}^{-1}$ ), C-S band ( $1080\text{-}1100\text{ cm}^{-1}$ ), C-C alicyclic bands in  $600\text{-}1300\text{ cm}^{-1}$  region and in the low wavenumber region, below  $400\text{ cm}^{-1}$ , which are due to the differences in crystal packing and lattice vibrations (Figure 6.9), whereas the spectra collected from a mixture of forms III (major) and II show subtle differences compared with pure form II.

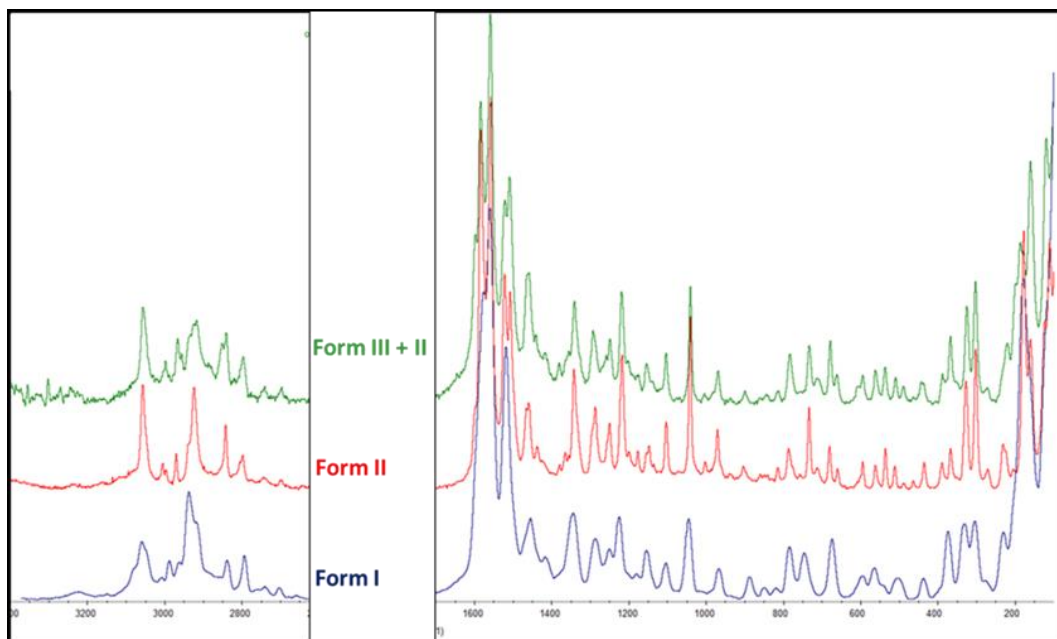


Figure 6.9. Raman spectra of form I (blue), form II (red) and form III + II (green).

FT-IR spectra of form I vs. form II (major) + III or form III (major) + II show significant differences whereas form II (major) + III or form III (major) + II show subtle differences (Figure 6.10). Raman and FT-IR data suggest that the structures of forms II and III are closely related which was also observed using SS-NMR (Reutzel-Edens et al., 2003).

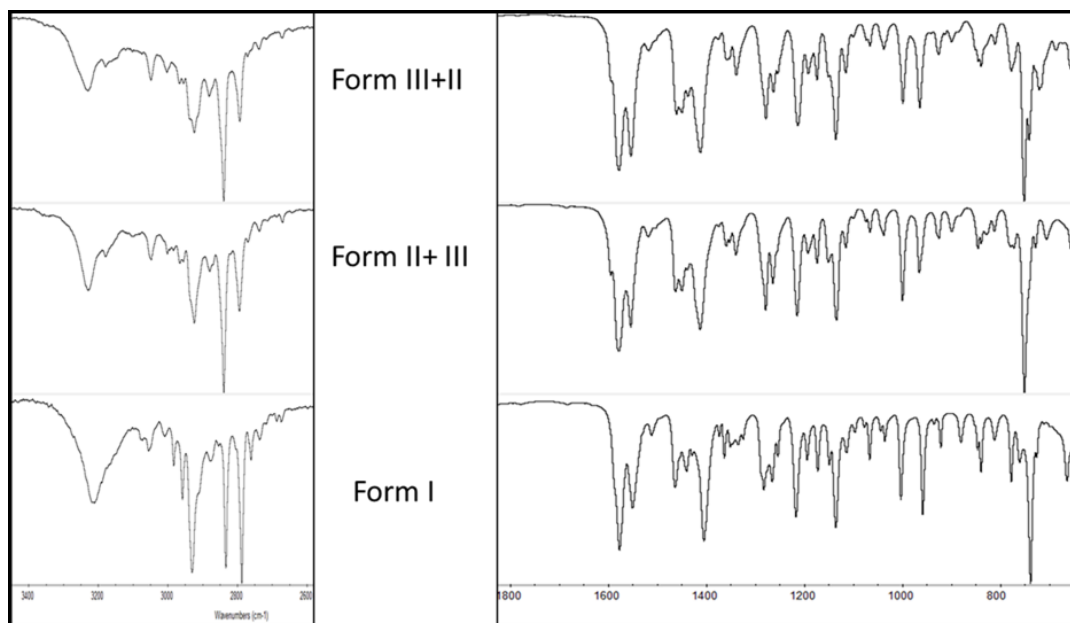


Figure 6.10. FT-IR spectra of form I, II (major) + III (minor), III (major) + II (minor) and amorphous phase of olanzapine.

#### 6.4.2 Solution Crystallisation of Olanzapine

The solvent screen includes slow evaporation in vials at RT and 277 K, faster solvent evaporation, cooling crystallisation on automated platform, antisolvent crystallisation, and crystallisation on various surfaces and as a function of pH using 71 diverse solvents (see Section 6.2.1 and Appendix Table Apx 1.1- Table Apx 1.6). The results of various solution crystallisations are reported in Table 6.4.

Table 6.4. Summary of crystallisation techniques used and summary of solid forms produced in the experimental search for physical forms of olanzapine.

Crystallisation Techniques	Crystallisation Method	Number of Solid Form(s) (See Further Details for Identity)			Further Details
		Non-solvated Polymorph (s)	Solvate (s)	Amorphous	
Solution crystallisation	T <sub>sat</sub> = RT, Evaporation at RT, on watch glass and at 277 K (Appendix Table Apx 1.1- Table Apx 1.2)	1	56 (see further details for identity)		<b>Crystal Structure Determined in This Work</b> see Table 6.5 <b>Crystal Structure Known from Literature</b> dihydrate D, dihydrate E, DMSO hydrate, methanol hydrate, 2-butanol hydrate , ethanol hydrate <b>No Crystal Structure Determined<sup>a</sup></b> Section 6.4.2
	T <sub>sat</sub> = 343K, Cooling to 298 K with agitation 850 rpm (Appendix Table Apx 1.3)	1			
	Antisolvent crystallisation ( Appendix Table Apx 1.4)	1			
	Crystallisation on functionalised surfaces ( Appendix Table Apx 1.5)	1			Form I
	Crystallisation by pH swing ( Appendix Table Apx 1.6)			1	dihydrate B
Neat / liquid-assisted grinding (Appendix Table Apx 1.8)		1	20		form II and solvates as obtained for solution crystallisation except ethanol, water, DCM and DCM:acetonitrile:water (1:1:1)
Desolvation of solvates (Appendix Table Apx 1.9)	Isothermal and <i>in-situ</i> desolvation at 333 K and 318 K Desolvation under vacuum at RT)	3			form I, or II and III <sup>b</sup>
Crystallisation from vapour phase		3			Single crystals of form I and II, II, only concomitant with polycrystalline form III
Melt quenching				1	
Recrystallisation from amorphous form (Appendix Table Apx 1.10)		3			form I, or II and III <sup>b</sup>
Spray drying		3			form I, or II and III <sup>b</sup>
Freeze drying		3			form I, or II and III <sup>b</sup>

<sup>a</sup>hydrated solvates of 1-methylnaphthalene, 2-butanone, 2-methoxyethylether, 1,1,2-trichloroethylene, N,N,N-triethylamine, 2,2,2-trifluoroethanol, benzene, bromoform, butyl ether, cyclohexane, carbon tetrachloride, chloroform, DCM, dodecane, furfural, N,N-dimethylacetamide, N-methyl pyrrolidone, pentyl acetate, tetrachloroethylene and solvates of formic acid, 1-bromo-2-chloro ethane, 1-octanol, <sup>b</sup> Mixture of forms I, II and III in varying ratios

Of the 293 solution crystallisations carried out, 30 (10.2%) yielded either no sample or insufficient solid for XRPD analysis due to poor solubility and/or chemical decomposition. Of the 263 samples giving measurable diffraction, 180 were identified as crystalline solvates and 31 were identified as mixtures comprising form I and the respective solvate. The remaining 52 single phase samples obtained from solution crystallisation were form I (starting material). Crystallisation carried out at higher temperature (323 K) to avoid solvate formation on automatic platform yielded form I. No solution crystallisations yielded forms II or III. In all salt disproportionation experiments the OZPN dihydrate B (CSD refcode: AQOMAU01) (Reutzel-Edens et al., 2003) was obtained regardless of pH used. This reflects the rapid formation of dihydrate B in the presence of water.

In total, 56 different solvates were identified from solution crystallisation. Overall, whilst OZPN shows a prolific capacity to form solvates with a wide range of solvent molecules, under the conditions tested it did not form solvates with several esters (ethyl acetate, isopropyl acetate, butyl acetate, and isobutyl acetate), diethyl ether, diisopropyl ether, toluene or fluorotoluene. In some cases, dihydrate B/D was obtained from esters (Appendix Table Apx 1.1- Table Apx 1.6).

Of the 56 organic solvates and hydrates identified in this work, 10 crystal structures were available in the literature. Twenty four previously unreported solvate crystal structures were determined by SXD and the crystal structures of the remaining 22 solvates were not determined but were identified on the basis

of XRPD and/or DSC/TGA analyses. Lattice parameters of 30 crystal structures (24 novel solvates, form I, form II and crystal structure redeterminations of 4 solvates) are tabulated in Table 6.5. From the solvate crystal structures, there are three groups of similar unit cell parameters (Table 6.5) highlighting the essentially isostructural relationships within each group with OZPN forming a common network in which different solvent molecules can be accommodated. Apart from the higher hydrate, all the 11 other structures that crystallise in space group *C2/c* have the organic solvent molecule residing on a special position and are highly disordered. In this group of solvates, OPZN molecules play as host, where variety of guest molecules can be accommodated. Crystal data, data collection and refinement details are provided in Table Apx 1.7 of the appendix.

Table 6.5. Lattice parameters for olanzapine crystal structures found in the experimental search (standard uncertainties in parentheses). All parameters were determined from single-crystal diffraction collected at 123(2) K.

Structure ID	Form (OZPN:solvate)	Space Group	Stoichiometry <sup>a</sup>	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\alpha$ (°)	$\beta$ (°)	$\gamma$ (°)	Cell Volume (Å <sup>3</sup> )
1	Form I <sup>b</sup>	<i>P2<sub>1</sub>/c</i>		10.3411(13)	14.521(2)	10.5314(14)	90	100.291(4)	90	1555.9(4)
2	Form II	<i>P2<sub>1</sub>/c</i>		9.8544(14)	16.314(2)	9.9754(12)	90	98.304(8)	90	1586.9(4)
3	2-butoxyethanol	<i>P2<sub>1</sub>/c</i>	1:1	10.4420(10)	17.7328(15)	12.5720(12)	90	95.755 (5)	90	2316.17(12)
4	2,2,2-trifluoroethanol	<i>P2<sub>1</sub>/c</i>	1:2	12.72 44(9)	11.7538(8)	16.006 7(11)	90	94.782(2)	90	2385.63(5)
5	Acetic acid	<i>P2<sub>1</sub>/c</i>	1:1	10.0427(11)	12.7430(13)	14.3886(15)	90	92.436(4)	90	1839.70(5)
6	Ethylene glycol	<i>P2<sub>1</sub>/c</i>	1:1	9.839(2)	13.278(2)	14.588(2)	90	96.328(7)	90	1894.37(16)
7	Ethanol	<i>P2<sub>1</sub>/c</i>	1:1	10.349(2)	13.003(2)	13.870(2)	90	92.469(6)	90	1864.66(7)
8	Tertiary butyl methyl ether (TBMe)_ hydrate	<i>P2<sub>1</sub>/c</i>	1:1:2	14.2070(1)	12.2210(9)	14.5500(1)	90	109.239(3)	90	2385.2(2)
9	2-pentanol_hydrate	<i>P2<sub>1</sub>/c</i>	1:1:1	13.4466(10)	12.3833(9)	14.5522(9)	90	104.839(2)	90	2342.3(3)
10	2,2,2-triethylamine _ hydrate	<i>P2<sub>1</sub>/c</i>	1:0.5:2	14.817(1)	12.623(1)	14.461(1)	90	113.359(2)	90	2483.1(2)
11	t-butanol_hydrate	<i>P2<sub>1</sub>/c</i>	2:2:5	14.2491(9)	12.3838(8)	27.1418(17)	90	99.052(2)	90	4729.7(5)
12	3-methyl-1-butanol_hydrate	<i>P-1</i>	1:1:1	12.34(2)	13.55(2)	14.63(2)	105.88(6)	90.29(6)	92.00(6)	2317.9(4)
13	1,4-butanediol_hydrate	<i>P-1</i>	2:1:2	12.2634(1)	14.115(1)	14.178(1)	63.98(3)	65.65(4)	65.44(4)	1924.9(4)
14	Dimethylformamide (DMF)_ hydrate	<i>P2<sub>1</sub>/c</i>	2:1:2	14.9885(9)	12.7558(7)	19.8567(12)	90	93.348(2)	90	3789.93(6)
15	Pyridine_hydrate	<i>P2<sub>1</sub>/c</i>	1:0.5:1	14.8824(7)	12.7289(7)	20.3218(11)	90	93.465(2)	90	3842.65(6)
16	Tetrahydrofuran (THF) _ hydrate	<i>C2/c</i>	1:0.5:1	24.2355(10)	12.5913(5)	15.0767(6)	90	124.999(1)	90	3768.76(3)
17	1,4-dioxane_hydrate	<i>C2/c</i>	1:0.5:1	24.447(3)	12.566(2)	14.959(2)	90	125.42(5)	90	3744.9(16)
18	Acetone_hydrate	<i>C2/c</i>	1:0.5:1	24.226 (2)	12.527(12)	15.0737(15)	90	124.98(4)	90	3748.3(12)
19	1-butanol_hydrate	<i>C2/c</i>	1:0.5:1	24.575(6)	12.429(3)	15.007(3)	90	125.26(7)	90	3743.1(3)
20	Acetonitrile_hydrate	<i>C2/c</i>	1:0.7:1	24.307(1)	12.350(1)	15.031(1)	90	125.31(5)	90	3682.7(16)
21	Nitromethane_hydrate	<i>C2/c</i>	1:0.5:1	24.352(3)	12.371(1)	15.018(1)	90	125.531(5)	90	3681.9(16)
22	1,2-dimethoxyethane_hydrate	<i>C2/c</i>	n.d.	24.467(13)	12.229(8)	15.107(8)	90	125.67(2)	90	3672.2(7)
23	2-methoxyethanol_hydrate	<i>C2/c</i>	1:0.5:1	25.0869(8)	12.3821(4)	15.1983(5)	90	126.872(1)	90	3776.7(3)
24	Methyl acetate_hydrate	<i>C2/c</i>	1:0.5:1	24.899(4)	12.532(2)	15.243(2)	90	126.99(5)	90	3798.9(18)



25	1,2-propanediol_hydrate	<i>C2/c</i>	1:0.5:1	24.704(2)	12.188(1)	15.124(1)	90	125.20(3)	90	3721.0(5)
26	1-propanol_hydrate	<i>C2/c</i>	1:0.5:1	24.475(1)	12.351(6)	15.111(8)	90	125.65(1)	90	3712.0(3)
27	Higher hydrate <sup>a</sup>	<i>C2/c</i>	1:2.5	24.940(6)	12.156(3)	14.867(3)	90	124.928(6)	90	3695.2(14)
28	Methanol <sup>a</sup>	<i>P2<sub>1</sub>/c</i>	1:1	10.1493(9)	12.2895(10)	14.0944(13)	90	91.905(5)	90	1757.0(3)
29	Dihydrate B <sup>a</sup>	<i>P2<sub>1</sub>/c</i>	1: 2	9.846(2)	12.672(3)	14.384(3)	90	92.724(9)	90	1792.7(7)
30	Dichloromethane(DCM) <sup>a</sup>	<i>P-1</i>	2:1	9.6671(15)	11.5070(17)	16.697(2)	103.991(6)	96.989(6)	99.825(6)	1749.8(4)

<sup>a</sup> stoichiometry of OZPN:solvent:water, determined using crystal structure and/or mass loss in TGA curve,

<sup>b</sup> crystal structure redetermined in this study at 123(2) K

n.d.-not determined

Twenty two solvates, for which crystal structures were not determined, were obtained during cooling crystallisation on automated platform (Appendix Table Apx 1.3) except formic acid solvate/salt which was obtained from liquid-assisted grinding (Appendix Table Apx 1.8). As mentioned earlier, these solvates were identified on the basis of XRPD and/or DSC/TGA analyses. 18 out of 22 solvates were found to be isostructural with the group of OZPN solvates which crystallise in space group  $C2/c$  (Table 6.5) based on solvent loss in TGA and similarity in XRPD data (Figure 6.11). The XRPD data of remaining 4 solvates did not match with any of the known solvate structures and are shown in Figure 6.12. The majority of solvate crystal structures (38/56) were mixed solvates containing the organic solvent of recrystallisation and at least one water molecule in the asymmetric unit in addition to OZPN. Water was not deliberately used during crystallisation but instead was either absorbed from the atmosphere or introduced in the organic solvent used for crystallisation.

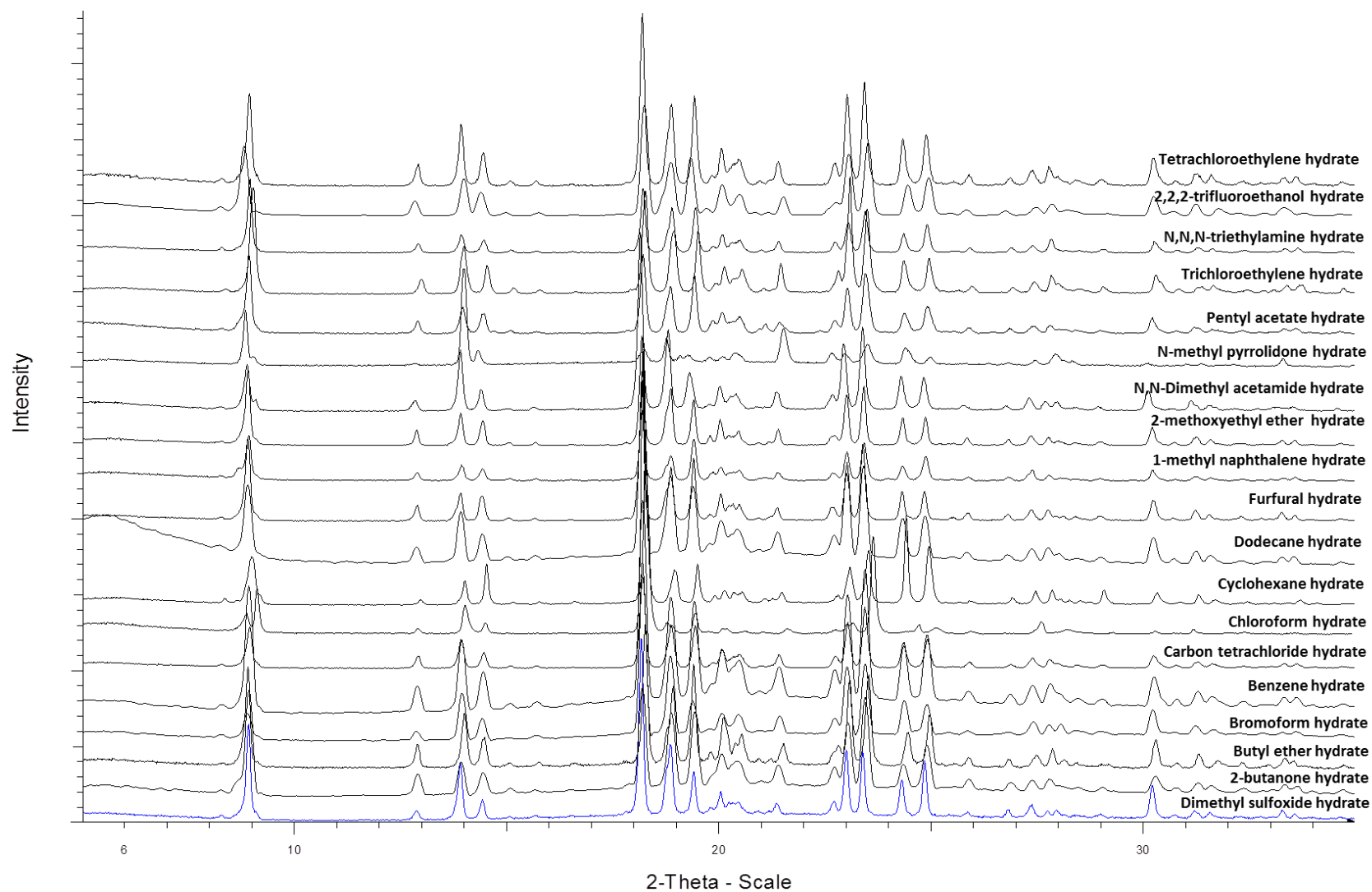


Figure 6.11. XRPD patterns of the 18 novel isostructural solvates (belong to crystal structures in space group  $C2/c$ ) observed during experimental screen and a reference pattern of the known olanzapine DMSO hydrate.

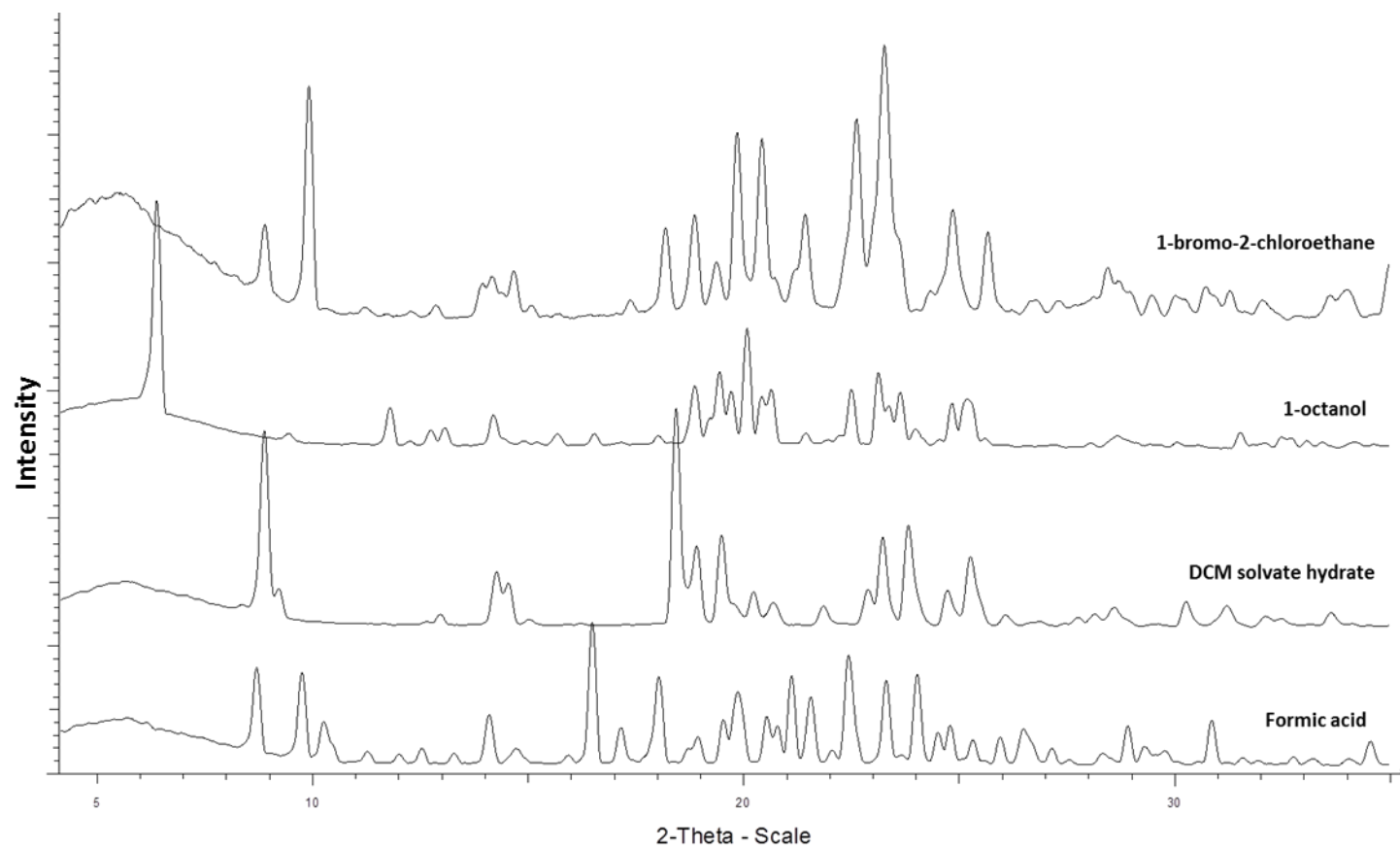


Figure 6.12. XRPD patterns of the four novel olanzapine solvates obtained from crystallisation on the automated platform.

### 6.4.3 Neat and Liquid-Assisted Grinding of Olanzapine

Neat grinding in a Retsch MM400 mixer mill (see Section 3.2.1.3) of form I produced form II with traces of form III. During the experiment, XRPD peaks of form II started appearing after 10 min of grinding at a frequency of 30 Hz with increased background, peak broadening and poor signal to noise ratio. Form II with traces of form III was obtained after 60 minutes of grinding at 30 Hz. Further grinding was not found to change the results. A Pawley fit of the single crystal unit cell and space group of form II against the XRPD data for the sample obtained by grinding is shown in Figure 6.13. The good agreement confirms that the single crystal structure is representative of the bulk sample, although some unfitted peaks of form III e.g.  $19.3^\circ 2\theta$  are present.

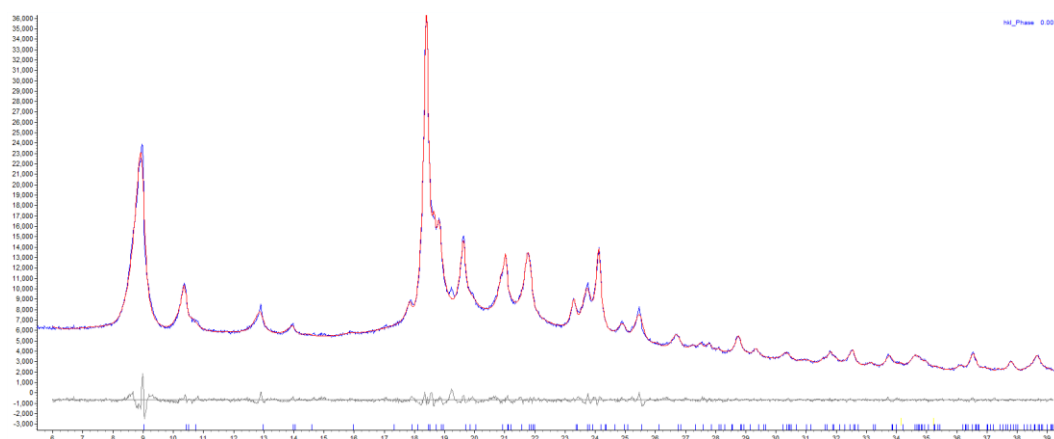


Figure 6.13. Pawley fit in TOPAS Academic (version 4.1), observed profile (blue), calculated profile (red) and difference plot (black) of the Pawley fit for form II in the range of  $6\text{-}39^\circ 2\theta$ .  $R_{wp} = 2.461$ .

Solvent assisted grinding (see Section 3.2.1.3) of form I OZPN was carried out using 33 solvents at 25 Hz for 30 minutes. Form I was observed with 13

solvents tested, and the remaining 20 samples returned a solvate, either as a pure sample (18 samples) or as a mixture with form I (2 cases; Appendix Table Apx 1.8). In the majority of cases (16/20), grinding produced the same solvated form as solution recrystallisation. For the remaining four solvent-assisted grinding experiments using ethanol, water, DCM, and DCM:acetonitrile:water (1:1:1), different solvated forms were obtained. These were mixed ethanol hydrate, higher hydrate which transforms to mixture of dihydrate E and B, DCM solvate and DCM solvate hydrate respectively (Appendix Table Apx 1.8).

#### **6.4.4 Desolvation of Olanzapine Solvates**

With the exceptionally large number of OZPN crystalline solvates produced, desolvation was assessed as a potential route to obtain new polymorphic forms. The majority of the solvates for which crystal structures were determined in this study were desolvated isothermally at 333 K and 318 K. After over 24 hours at 333 K, most samples transformed to form I, while desolvation at 318 K yielded mixtures of varying composition of forms I, II and III (in case of ethanol, 2-butoxyethanol, 2,2,2-trifluoroethanol, 1-butanol, dichloromethane, dihydrate B and chloroform) (Appendix Table Apx 1.9).

In an attempt to obtain form III, those solvates that showed evidence of forms II and III upon desolvation were also subjected to vacuum at RT as well as *in-situ* desolvation at 313 K monitored using XRPD. The lower temperatures were used in an attempt to isolate forms II and/or III and minimise the transformation to form I, which occurs rapidly at temperatures over 333 K. However, no

conditions were identified that yielded a pure sample of form III suitable for structure determination from powder diffraction data or further characterisation. Desolvation of the DCM solvate either under vacuum or at 318 K (Figure 6.14) gave the best yield of form III from any of the experimental conditions tested as part of the screen and samples prepared from this method were used in subsequent characterisations of this form. Desolvation of methanol and dihydrate B yielded form II as the major product.

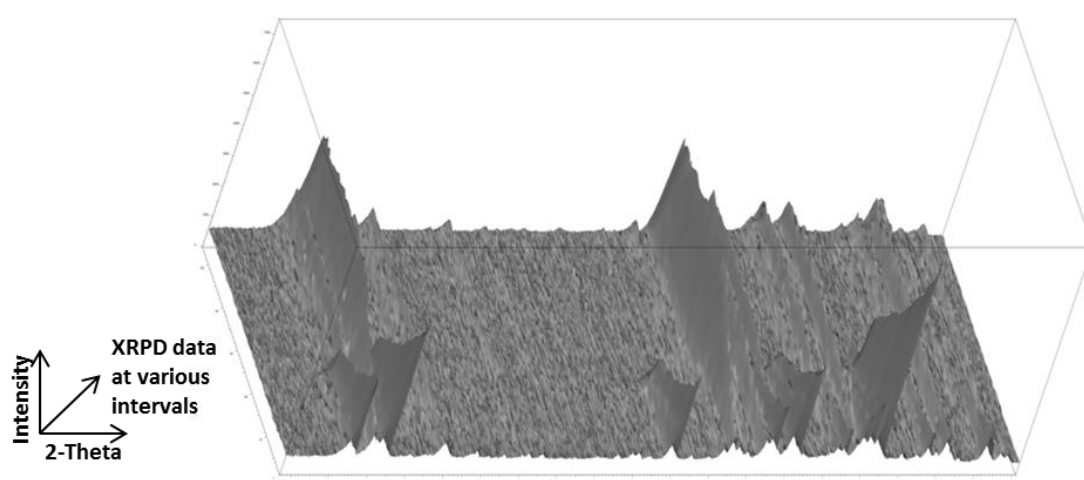


Figure 6.14. 3-D overlay of the XRPD patterns during *in-situ* desolvation of olanzapine DCM solvate at 318 K, showing appearance of peaks of form III and II and disappearance of peaks of DCM solvate.

#### 6.4.5 Melt Quenching and Recrystallisation of Amorphous Olanzapine

XRPD data of an amorphous form of OZPN showed the characteristic broad halo and absence of Bragg diffraction peaks (Figure 6.15). STA data show a glass transition  $\sim 343$  K after which the material crystallises with an onset temperature of 381 K to form II, III or a mixture thereof. On heating, form II/III

transform to form I at a temperature ca. 453 K, prior to finally melting at ca. 468 K (Figure 6.16).

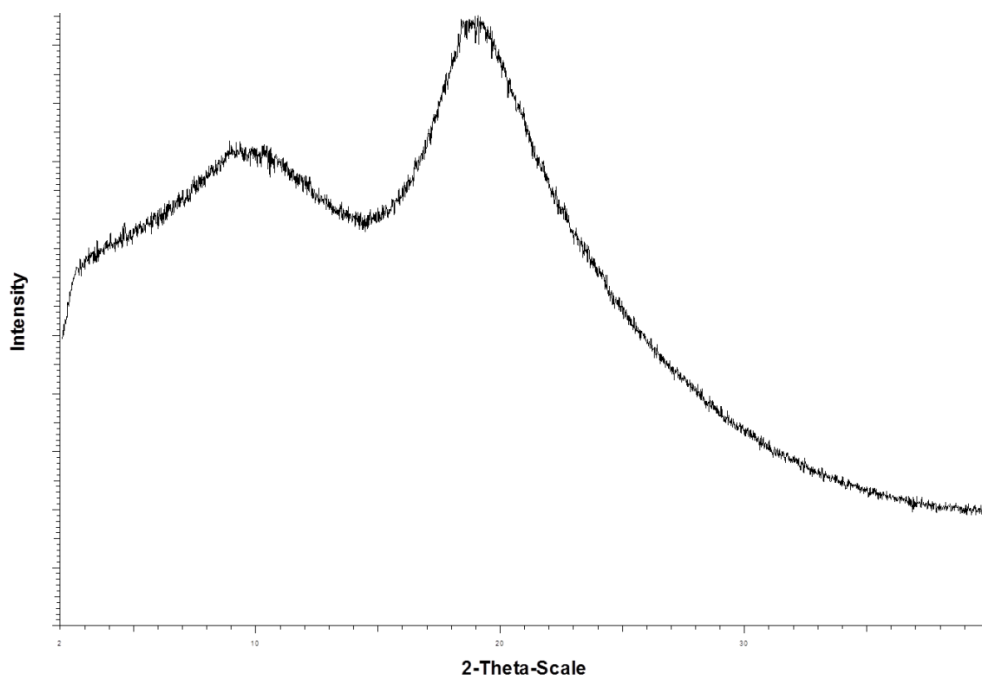


Figure 6.15. XRPD data of amorphous olanzapine in the region of 2-40° 2 $\theta$ .

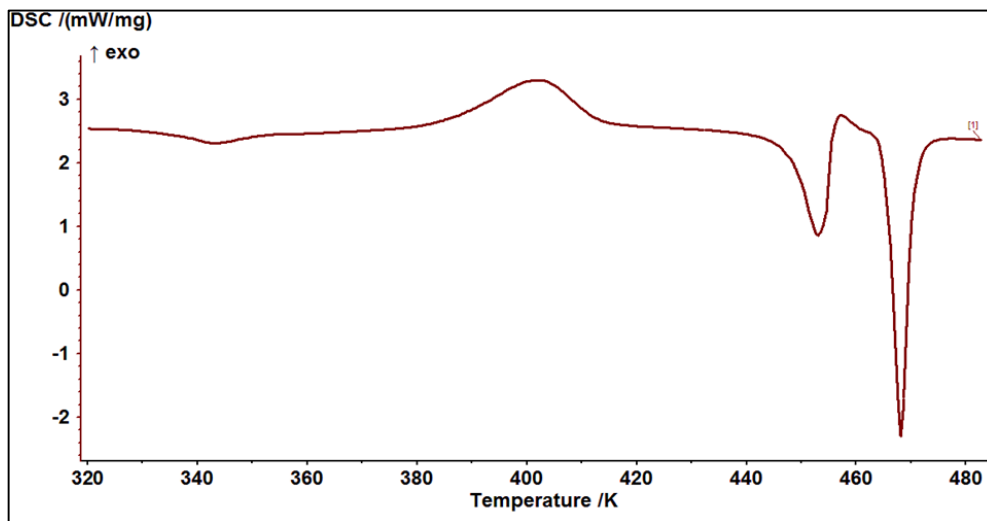


Figure 6.16. DSC data of amorphous olanzapine. Data was collected from 263 K to 483 K at a heating rate of 20 K min<sup>-1</sup>.

Further experiments to study the recrystallisation of OZPN from amorphous samples held under different conditions (for a minimum of 24 hours) show a



range of outcomes (Appendix Table Apx 1.10). A melt-quenched amorphous phase prepared on a glass slide, if left undisturbed, was stable for weeks. The amorphous form reproducibly crystallised to mixtures of form II and form III (in varying ratios) within a 0.7 mm borosilicate glass capillary within 24 hours (Figure 6.17). Also a mixture of form I, II and III was obtained from an amorphous sample left in a vial under ambient conditions (Figure 6.17). When the amorphous sample was heated at 313 K, the ratio of form I was increased in the recrystallised product (Figure 6.17). No phase transformation was observed when sample was stored below 283 K.

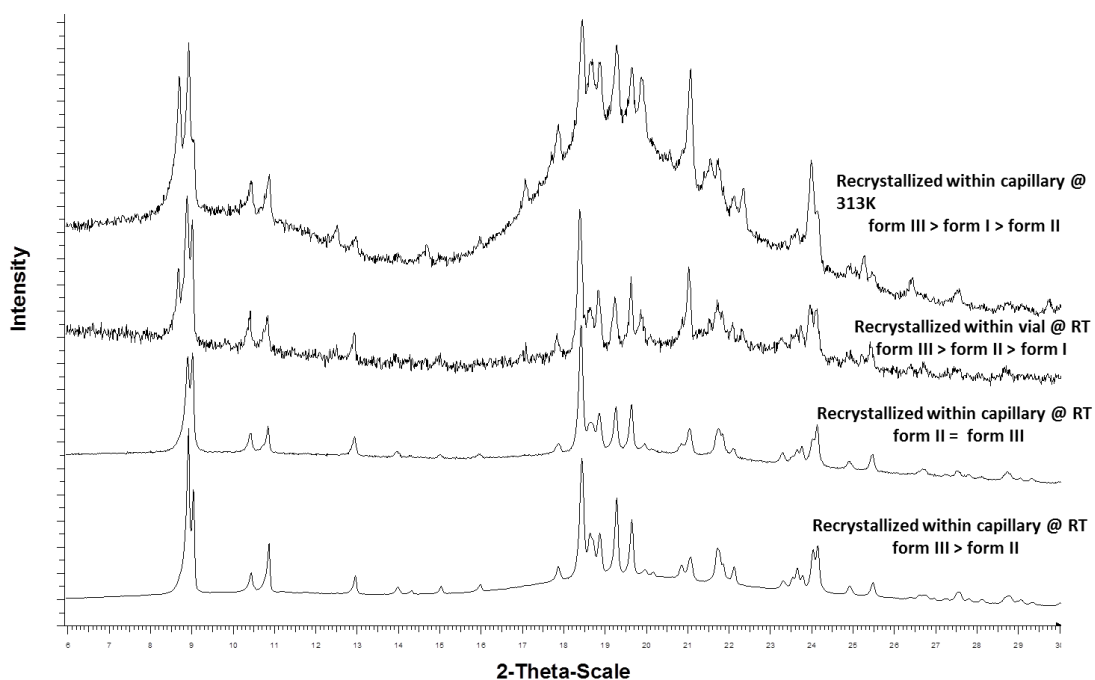


Figure 6.17. XRPD data in the range 6–30° 2 $\theta$  collected from melt-quenched sample of olanzapine recrystallised within boro-silicate capillary or in a vial held at RT or at 313 K.

Based on the DSC, OZPN is present as form II or III after recrystallisation of amorphous phase at 393 K. As guided by DSC, to get pure form II/III, the

amorphous sample was directly heated at 393 K, but the thermodynamically stable form I was observed within 5 minutes. No traces of form II or III were observed when heated above 353 K. Phase pure sample of form II or III could not be obtained in a capillary under the conditions tried for recrystallisation of amorphous form.

#### 6.4.6 Spray Drying and Freeze Drying of Olanzapine

Spray drying (see Section 6.2.1.5) can produce different polymorphs, particle morphologies and particle sizes (Yu and Ng, 2002; Vehring, 2008). All three known polymorphs were obtained from preliminary assessment of spray drying (Table 6.6) and the polymorphic form(s) produced vary according to the experimental conditions used. A further systematic study is required to explore the possibility of obtaining pure form III in particular. Limited freeze drying experiments (see Section 3.2.1.6) provided mixture of form I and form II. XRPD data revealed some amount of amorphous content in the sample and were confirmed by STA.

Table 6.6. Results of spray drying experiments for olanzapine.

S. No	OZPN Solution	PXRD	DSC
1	1% DCM	Form I (major) + form III (minor) + form II (traces)	Melting endotherm corresponds to form I
2	0.5% DCM	Form III (major) + form I (minor) + form II (traces)	Melting endotherm corresponds to form I
3	0.25% CH <sub>3</sub> OH	Form II + form I + methanol solvates (traces) Observed halo of amorphous, but material crystallised within minutes	Mass loss and Melting endotherm corresponds to form I

## 6.5 Molecular Packing Analysis of Olanzapine Crystal Structures using XPac

Packing arrangements of all 35 known OZPN crystal structures (30 crystal structures from this work as per Table 6.5 and 5 solvate crystal structures from the CSD) were compared using XPac to identify the common packing features that underpin the large number of crystal structures. In all the structures, in the solid-state, the diazepine ring of OZPN exists in a puckered conformation between the planes of the thiophene and benzene rings and the piperazine ring adopts a chair conformation with the methyl group in an equatorial position. All structures show packing motifs involving the centrosymmetric dimer, hereafter referred to as  $SC_0$ . This uniform appearance of  $SC_0$  in all crystal structures confirms  $SC_0$  as the key structural building block from which crystalline structures of OZPN are assembled (Reutzel-Edens et al., 2003). The head-to-tail arrangement of OZPN molecules in the dimer affords the opportunity for the single hydrogen-bond donor, N1 and the two hydrogen-bond acceptors (imine N2 and piperazine N4) to form hydrogen bonds with adjacent dimers or solvent molecules, resulting in a variety of extended arrangements. The XPac analysis also identified a range of 1-D, 2-D and 3-D similarities across the OZPN structures that illustrate repeated occurrences of common packing arrangements in the series. The supramolecular constructs (SC) (Gelbrich and Hursthouse, 2006) identified are summarised in Table 6.7 and the associated structural inter-relationships are illustrated in Figure 6.18 and Figure 6.19.

Table 6.7. Highest-dimensionality SCs observed in the 35 olanzapine structures and forms obtained after desolvation. The number in square brackets corresponds to each structure in Figure 6.18.

SC	Dimensionality	Structures	Desolvation Product
SC <sub>0</sub>	0D	2-butoxyethanol[3], 2,2,2-trifluoroethanol[4]	Form II and III
SC <sub>11</sub>	1D	Form II[2], DCM [30]	Form II and III
SC <sub>12</sub>	1D	2-butanol hydrate[35] <sup>a</sup>	
SC <sub>22</sub>	2D	Form I[1], dihydrate D[34] <sup>a</sup>	
SC <sub>31</sub>	3D	1,4-butanediol hydrate[13], DMF hydrate[14], pyridine hydrate[15], THF hydrate[16], 1,4-dioxane hydrate[17], acetone hydrate[18], 1-butanol hydrate[19], acetonitrile hydrate[20], nitromethane hydrate[21], 1,2-dimethoxyethane hydrate[21], methoxy ethanol hydrate[23], methyl acetate hydrate[24], 1,2-propanediol hydrate[25], 1-propanol hydrate[26], higher hydrate[27], DMSO hydrate[31] <sup>a</sup> , methanol hydrate[32], ethanol hydrate[33] <sup>a</sup>	Form I except 1-butanol hydrate which provides form II and III
SC <sub>32</sub>	3D	Acetic acid[5], ethylene glycol[6], , ethanol[7], methanol solvate[28], dihydrate B[29],	Form II + III except acetic acid solvate which provides form I
SC <sub>33</sub>	3D	TBMe hydrate[8], 2-pentanol hydrate[9], N,N,N-triethylamine hydrate[10] <sup>a</sup> , isoamylalcohol hydrate[12] t-butanol hydrate[11]	Form I

<sup>a</sup> – no desolvation data.

SC<sub>11</sub> is the most frequently observed 1-D OZPN SC (10/35 structures) and is of particular relevance given that it is a shared feature of OZPN form I and II as well as a number of solvates. It comprises a 1-D stack of SC<sub>0</sub> dimers and can

pack in a variety of ways. The form I and II crystal structures differ in the relative orientations of the SC<sub>11</sub> stacks about their  $t_4$  translation vector (defined in Figure 6.19) such that in form I they are approximately parallel whilst in form II they adopt an offset “corrugated” arrangement (Figure 6.20).

The DCM solvate (‘30’, Figure 6.20) shows a significantly different arrangement, in which alternating layers of SC<sub>11</sub> pack along a second axis with alternate layers lying perpendicular to those of its neighbours. This solvate shares 1-D similarity with only one other polymorph structure, form II, via SC<sub>11</sub>.

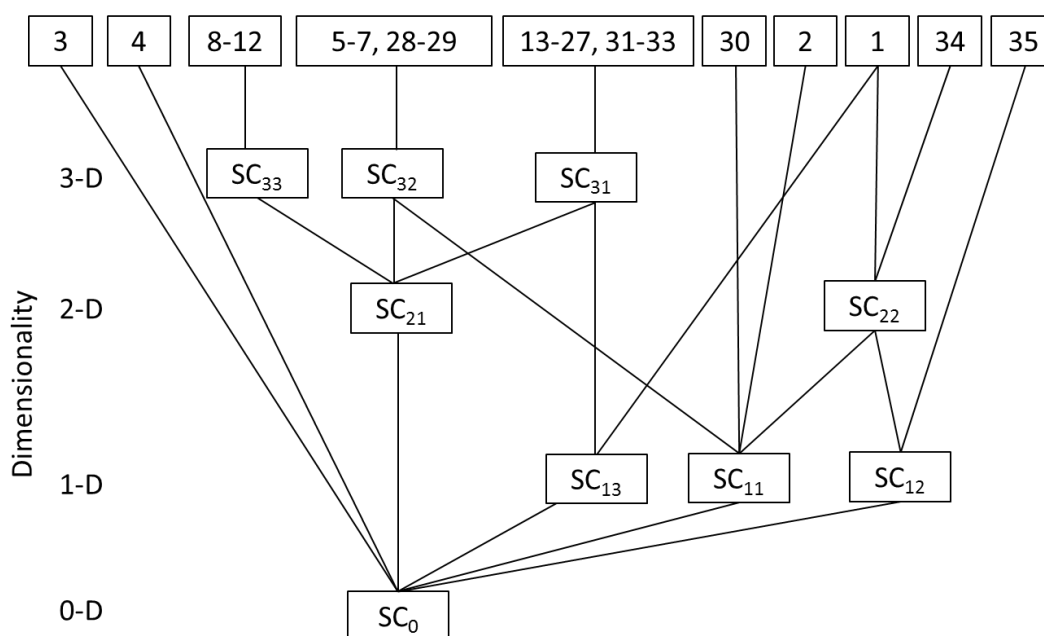


Figure 6.18. Diagram showing the similarity relationship between 35 structures of olanzapine with structures identified by the crystal structure numbers used in Table 6.7.

In 28 structures, the dimer SC<sub>0</sub>, adopts a common arrangement along two axes to form the SC<sub>21</sub> layer, the most frequently occurring 2-D SC. The high frequency

of occurrence (80% of all OZPN structures) highlights the very robust nature of SC<sub>21</sub>. SC<sub>21</sub> sheets can also stack with different inter-layer spacing and translations to produce three distinct 3-D SCs, namely SC<sub>31</sub>, SC<sub>32</sub> and SC<sub>33</sub> (Figure 6.18). SC<sub>22</sub> is present in only 2 structures, form I and dihydrate D ('1' and '34', Figure 6.18) with the major difference between the structures resulting from the relative orientations of neighbouring SC<sub>22</sub> sheets due to the incorporation of water molecules in dihydrate D (Figure 6.20).

Of the 3-D SCs, the most common, SC<sub>31</sub>, is observed in 18 structures which are all solvates (Figure 6.18). This high dimensionality SC comprises SC<sub>21</sub> plus the 1-D SC<sub>13</sub> which is also shared with OZPN form I. This structural relationship may contribute to the facile nature of the transformation for the 18 solvates that contain SC<sub>13</sub> to form I upon desolvation (Table 6.7). SC<sub>32</sub>, is observed in five structures in which the arrangement of SC<sub>21</sub> is defined by the 1-D SC<sub>11</sub>. The corrugated arrangement of SC<sub>11</sub> is preserved in the SC<sub>32</sub> arrays but the inclusion of solvent molecules, illustrated in Figure 6.20 for the acetic acid solvate, leads to an increase in the separation between SC<sub>11</sub> columns and breaks the 3-D similarity between form II and this group of solvates. Given the shared SC<sub>11</sub> link between the SC<sub>32</sub> solvates and form II it is interesting to note that three of the four solvates in this group produced mixtures of forms II and III upon desolvation. SC<sub>33</sub> is shared by 5 solvates in which the arrangement of molecules along the third axis shows a larger inter-layer spacing in order to accommodate the bulky solvent molecules in this group (2-pentanol, TBMe, isoamyl alcohol, N,N,N-triethylamine and t-butanol).

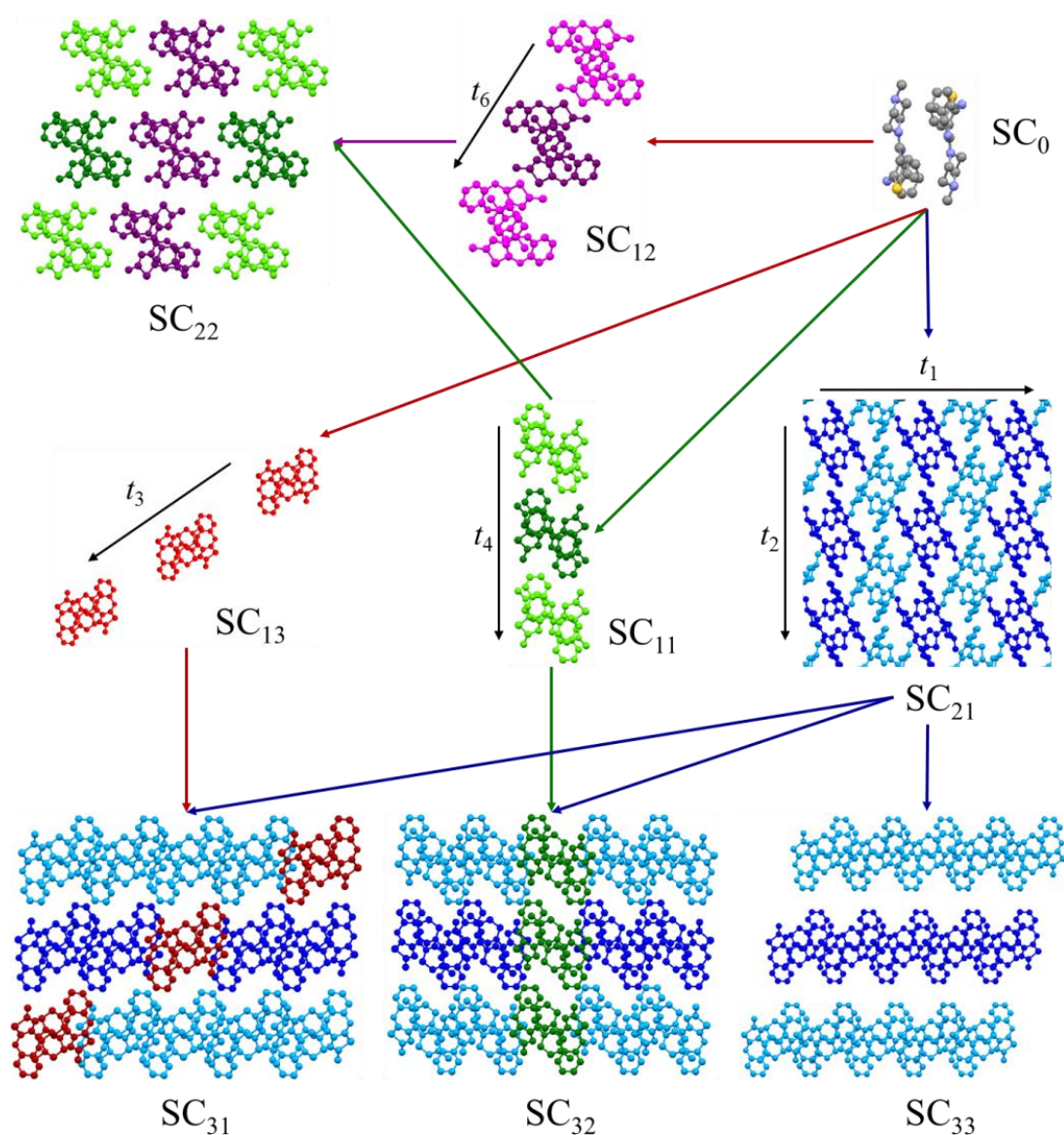


Figure 6.19. Packing diagrams highlighting the supramolecular constructs (SCs) and their relationships highlighted in Figure 6.18. The directions of the primary vectors are indicated by arrows and in the  $SC_{21}$  layer the  $SC_0$  pairs are differentiated by light or dark blue shading. As the three-dimensional arrays correspond to three distinct packing modes of the  $SC_{21}$  layers these are all viewed along the direction of the  $t_2$  vector with the constituent  $SC_{21}$  layers differentiated by dark and light blue shading. The three one-dimensional SCs are coloured green ( $SC_{11}$ ), purple ( $SC_{12}$ ) and red ( $SC_{13}$ ) respectively and are viewed perpendicular to the axis of their translation vectors. In  $SC_{31}$  and  $SC_{32}$ , the  $SC_{21}$  layers are related by the  $SC_{13}$  and  $SC_{11}$  constructs respectively and a single occurrence of each, viewed perpendicular to its translation vector, is highlighted with the corresponding colour. A single  $SC_{22}$  layer is shown with the  $SC_{12}$  component running approximately vertical, highlighted in purple, with three  $SC_{11}$  constructs running horizontally and differentiated by dark and light green shading.

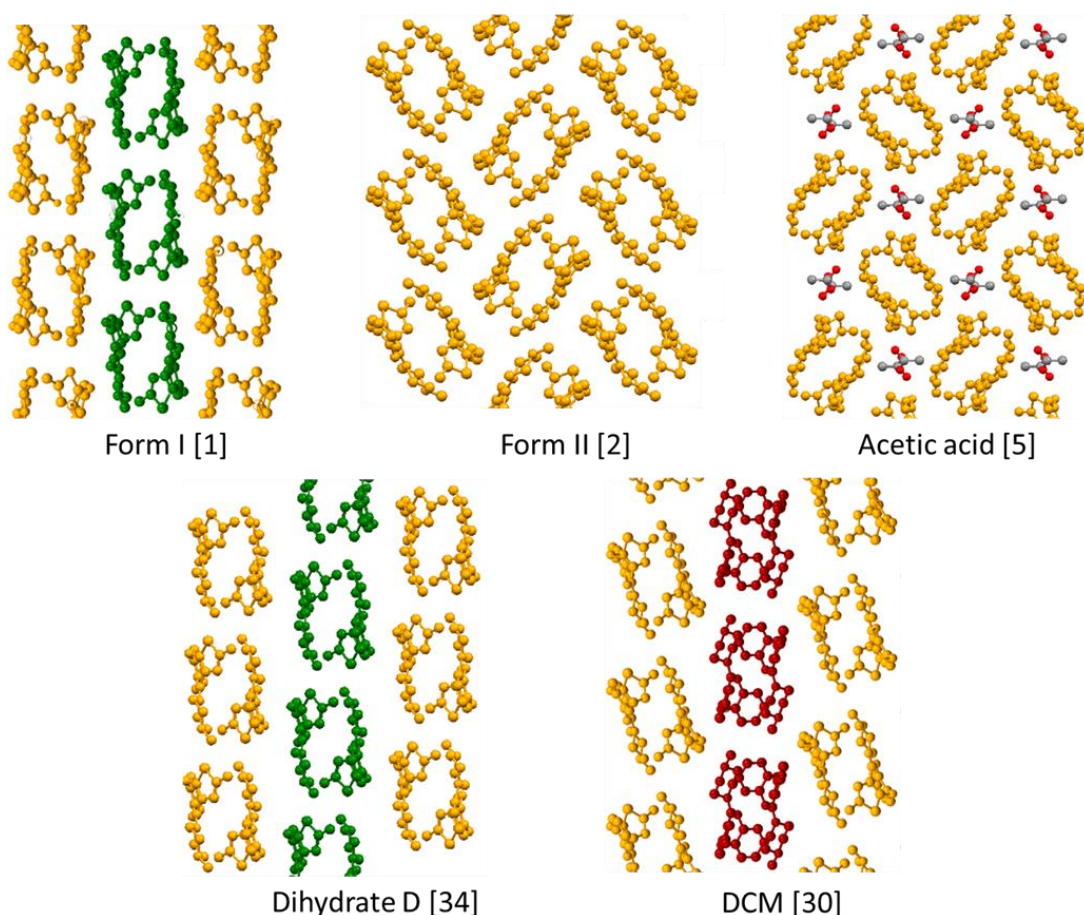


Figure 6.20. The five different packing arrangements adopted by the  $SC_{11}$  stacks, these are viewed along the axis of the  $t_4$  vector (which run perpendicular to page) with each pair of face-to-face molecules (coloured yellow) representing a single  $SC_{11}$  stack. In the dichloromethane solvate the dark red molecules represent  $SC_{11}$  columns in which the  $t_4$  vector runs perpendicular to those in the neighbouring layers, in form I and dihydrate D, a  $SC_{22}$  layer is shown by the green molecules.

## 6.6 Hydrogen-Bonding Analysis of Olanzapine Crystal Structures

In all the experimental structures, the arrangement of OZPN in  $SC_0$  exposes the H-bond donor and acceptors to take advantage of available H-bonding sites. Both form I and II have the same hydrogen bonding i.e. N1-H of the azepine ring of every OZPN forms H-bond with N2 of the azepine ring of OZPN in adjacent



centrosymmetric dimer, leaving behind two unsatisfied hydrogen bond acceptors (N3 and N4) on the piperazine ring (Figure 6.21).

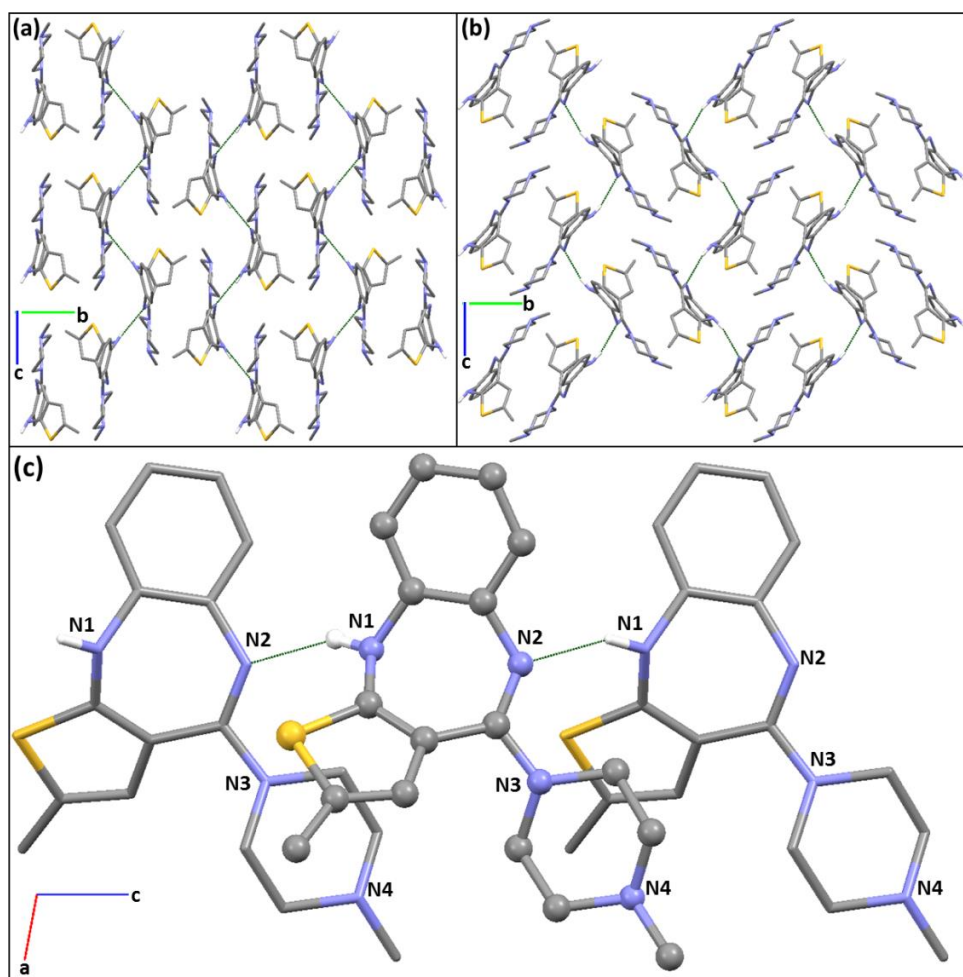


Figure 6.21. Different arrangements of SC<sub>11</sub> in (a) form I and (b) II of olanzapine, which share the same H-bonding (shown by green dotted line). (c) A closer view at the olanzapine atoms involved in H-bonding. Other hydrogen atoms are omitted for clarity.

In majority of the solvate structures, hydrogen bonding is mediated by the solvent molecule(s) and, in some cases, the solvent molecule(s) fulfil(s) the hydrogen bonding sites of OZPN in place of other OZPN contacts. As a water molecule is present in 36 out of 56 solvates, special emphasis was given to find out the environments of water molecules in these structures. Water can be

easily incorporated into crystal lattices because of its small size and multi-dimensional H-bonding capabilities (Gillon et al., 2003). As such, water molecules can stabilise crystal structures when there is an imbalance in the number of acceptors and donors (Desiraju, 1991b). In OZPN solvates, three environments of water were observed (Figure 6.22), **Type 1, DD**: where water molecule acts as donor for two H-bonds, **Type2, DDA**: where water molecule acts as a donor for two H-bonds and acceptor for one H-bond [indeed this is the most common environment found for water molecules in organic crystal structures in the CSD (Gillon et al., 2003)], **Type 3, DDAA**: where water molecule acts as donor as well as acceptor for two H-bonds each, thus forming four H-bonds. This is the 2<sup>nd</sup> most common environment for water molecule in organic crystal structures in the CSD (Gillon et al., 2003) .

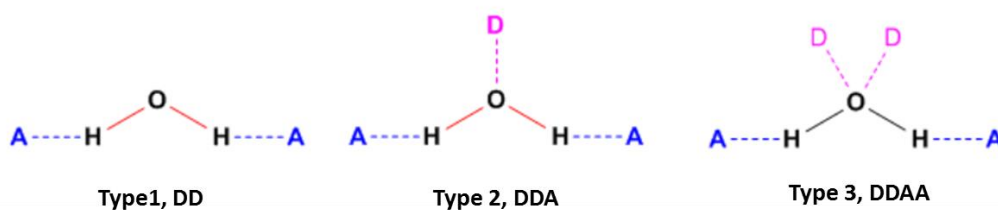


Figure 6.22. Three water H-bonding environments observed in family of olanzapine solvate crystal structures, (D-Donor, A-acceptor). Weak interactions such as C-H...O were not included in the analysis. Reproduced from reference (Clarke et al., 2010).

The packing analysis of solvates has been covered in Section 6.5. Here the specific hydrogen bonding interactions of OZPN dimers with water molecules and/or solvent in various packing types (based on SC<sub>0</sub>, SC<sub>11</sub>, SC<sub>22</sub>, SC<sub>31</sub>, SC<sub>32</sub> and SC<sub>33</sub>, see Section 6.5) are described.

## 6.6.1 Hydrogen Bonding in Olanzapine Solvates Based on Packing

### Type SC<sub>31</sub>

Mixed OZPN solvates that exist in group SC<sub>31</sub> have a minimum of one water molecule with the organic solvent in the asymmetric unit. OZPN dimers are connected by N1H-N4 hydrogen bond (instead of N1H-N2 in form I and II) mediated by water molecule present within the SC<sub>21</sub> sheets. Each OZPN molecule participates in two hydrogen bonds as in anhydrous forms but involving different acceptor site (N4) in H-bonding. Water molecules are present in type 2 and/or type 3 environments. OZPN 1,4-dioxane solvate hydrate is shown as an example of observed type 2 environment for water molecules in Figure 6.23.

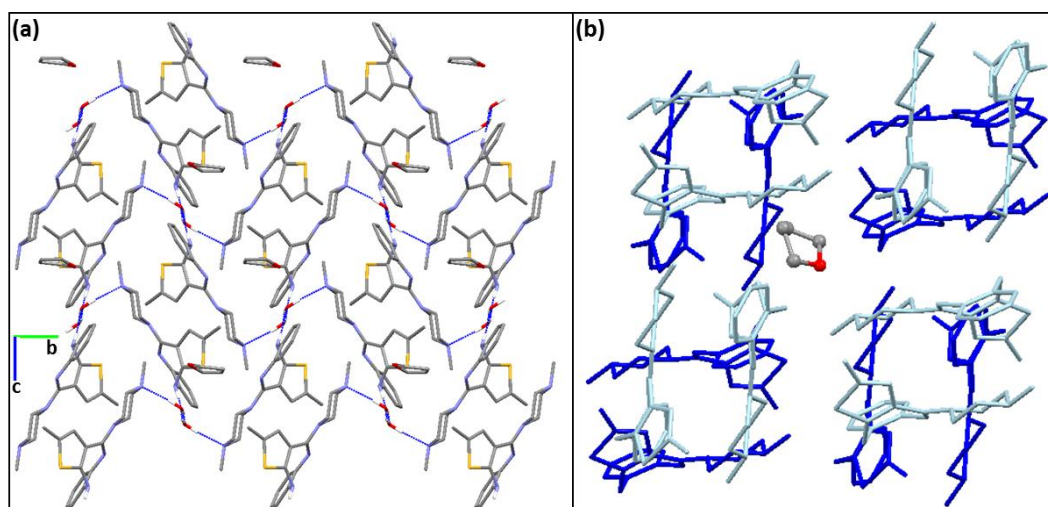


Figure 6.23. (a) Hydrogen bonds (N1H...N4) between centrosymmetric dimers, conciliated by water molecules present within SC<sub>21</sub> sheet of olanzapine 1,4-dioxane solvate hydrate. H-bonds are shown by blue dotted lines (b) organic solvent molecule surrounded by 4 pairs of centrosymmetric dimers from adjacent SC<sub>21</sub> sheets (coloured by blue and cyan), arranged in a staggered manner. Other hydrogen atoms are omitted for clarity.

Each organic solvent molecule is surrounded by four pairs of centrosymmetric dimers of OZPN (four from two different SC<sub>21</sub> sheet), similar to a host-guest complex, where the solvent molecule acts as a guest. The solvent molecule is involved in hydrogen bonding and/or hydrophobic interactions with water and OZPN molecules (Figure 6.23).

## 6.6.2 Hydrogen Bonding in Olanzapine Solvates Based on Packing Type SC<sub>32</sub>

All OZPN solvates present in this group are anhydrous organic solvates. Methanol, ethanol and acetic acid solvates showed similar H-bond interactions (N1H...N4) to those of mixed solvates structures in SC<sub>31</sub> but mediated by a solvent molecule (Figure 6.24).

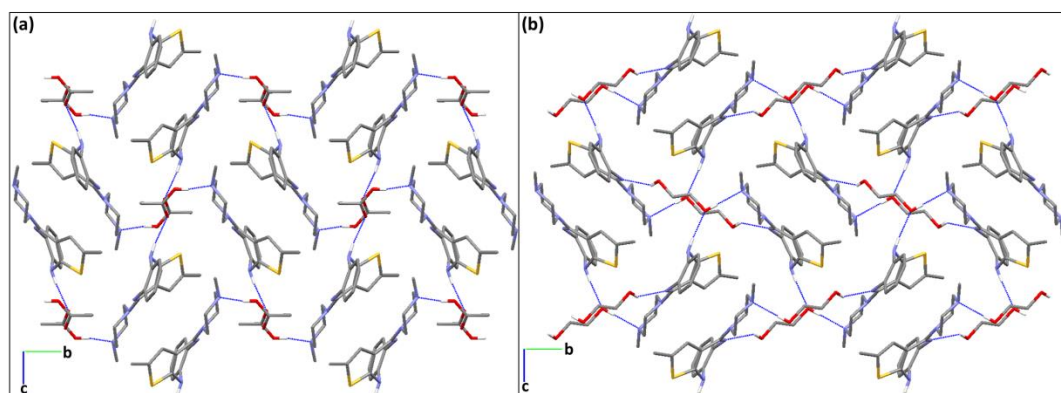


Figure 6.24. (a) Olanzapine acetic acid solvate in which acetic acid links the centrosymmetric dimers through N1H...N4 H-bonds. (b) Olanzapine ethylene glycol solvate, where each olanzapine molecule is involved in three H-bonds (N1H, N2 and N4) with solvent molecules. H-bonds are shown by blue dotted lines. Other hydrogen atoms are omitted for clarity.

In case of EG solvate and dihydrate B, each OZPN molecule takes advantage of one additional hydrogen bonding site (N1H, N2 and N4) (Figure 6.24). In case of dihydrate B, water molecules are arranged as a catemer.

### 6.6.3 Hydrogen Bonding in Olanzapine Solvates Based on Packing Type, SC<sub>33</sub>

All olanzapine solvates in this group are mixed solvates, i.e. organic solvent exists with at least one water molecule in the asymmetric unit. Water molecules were observed in a type 2 environment in all crystal structures except t-butanol solvate hydrate. In the t-butanol solvate hydrate, water molecules are arranged as catemer which shows all the three types of water environments (Figure 6.25).

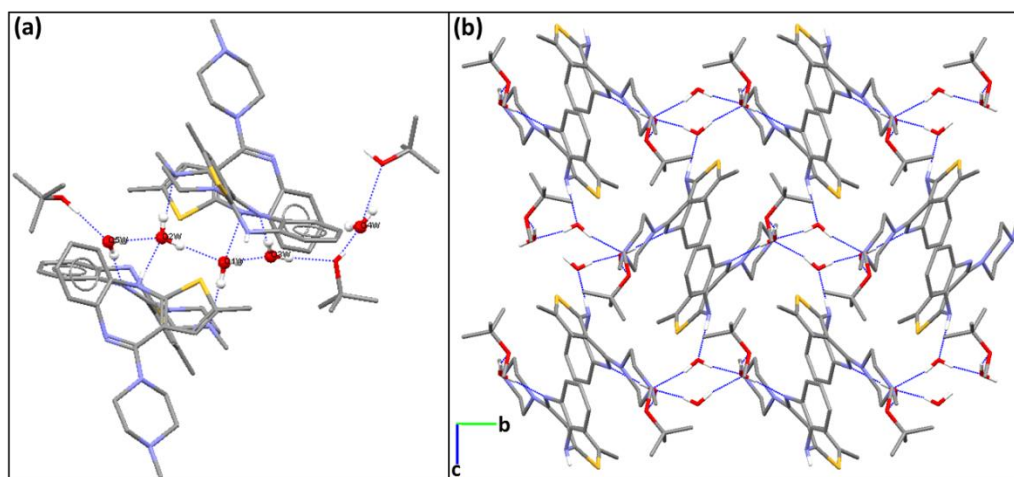


Figure 6.25. (a) Catemer arrangement of water molecules in olanzapine t-butanol solvate hydrate. (b) Water molecules exist in type 2 environment in olanzapine TBMe solvate hydrate. H-bonds are shown by blue dotted lines. Other hydrogen atoms are omitted for clarity.

TBMe solvate hydrate is shown as an example of observed type 2 environment in Figure 6.25. In each structure, OZPN molecule takes advantage of one additional hydrogen bonding site (N1H, N2 and N4) compared to anhydrous forms.

#### 6.6.4 Hydrogen Bonding in Other Olanzapine Solvates Based on SC<sub>0</sub>, SC<sub>11</sub> and SC<sub>22</sub>

In 2-butoxyethanol solvate, the SC<sub>0</sub> dimers arrange themselves to accommodate the large 2-butoxyethanol molecules which connect OZPN dimers through N1H...N4 H-bonds (Figure 6.26). In 2,2,2-trifluoroethanol solvate, in addition to the solvent mediated N1H...N4 hydrogen bond, each solvent molecule also forms H-bond with the N2 of OZPN molecule. In case of 2-butanol hydrate, water molecules are present in a type 2 environment and each OZPN participates in three H-bonds. In Dihydrate D, each OZPN forms three H-bonds (N1H, N2 and N4) with water molecules, which are arranged as a cyclic tetramer and are present in type 2 and 3 environments (Figure 6.26).

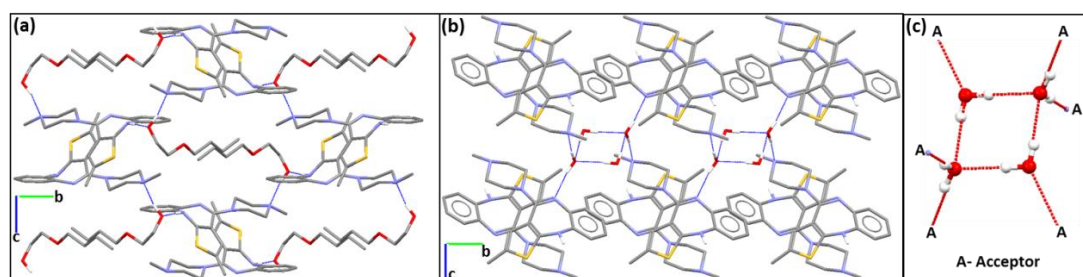


Figure 6.26. H-bonding (N1H...N4) in (a) olanzapine 2-butoxyethanol solvate (b) olanzapine dihydrate D, where water is present as cyclic tetramer (c) cyclic tetramer, where water is present in type 2 and type 3 environments. H-bonds are shown by blue colour. Other hydrogen atoms are omitted for clarity.

## 6.7 Prediction of Olanzapine Solvate Formation using Random Forests Classification Model

XPac analysis (see Section 6.5) revealed that the SC<sub>21</sub> sheets are present in 28 solvates out of 35 and gives rise to three 3-D molecular packings based on SC<sub>31</sub>, SC<sub>32</sub> and SC<sub>33</sub> (Figure 6.27). SC<sub>32</sub> and SC<sub>33</sub> differ in the separation distance between neighbouring SC<sub>21</sub> sheets. In the case of SC<sub>31</sub>, the distance between SC<sub>21</sub> sheets is similar to that of SC<sub>32</sub>, but the neighbouring SC<sub>21</sub> sheets are shifted by one molecule relative to each other along the  $t_1$  axis. In other words, SC<sub>21</sub> sheets pack in either a staggered or an eclipsed manner in SC<sub>31</sub> and SC<sub>32</sub>, respectively.

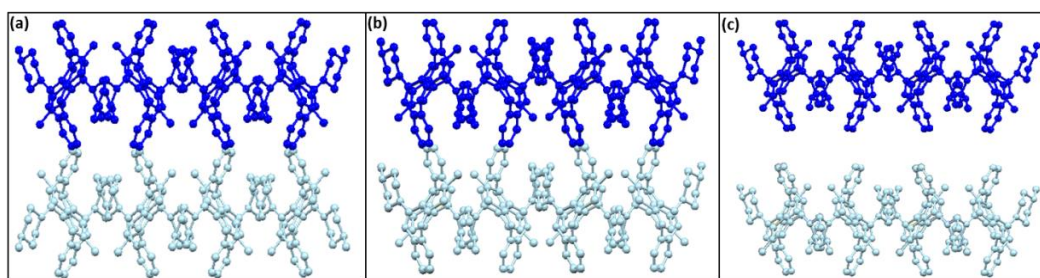


Figure 6.27. Three 3-dimensional supramolecular constructs, (a) SC<sub>31</sub>, (b) SC<sub>32</sub> and (c) SC<sub>33</sub>, formed by SC<sub>21</sub> sheets (coloured by cyan and blue colour) in olanzapine solvates.

The solvent molecules occupy the void spaces between the SC<sub>21</sub> sheets in solvates based on SC<sub>31</sub> and SC<sub>33</sub> and within the SC<sub>21</sub> sheet in case of solvates based on SC<sub>32</sub> (Figure 6.28). Therefore, it was of interest to explore the effect of the solvent molecules on the three observed 3-D molecular packing types in the 28 experimental crystal structures and to confirm if the solvent molecules are directing the crystal packings of the solvates. To achieve this, RF (see Section

5.1.4) technique was employed to develop a classification model to correlate calculated solvent physicochemical properties with the three 3-D molecular packing types observed in the crystal structures of OZPN solvates.

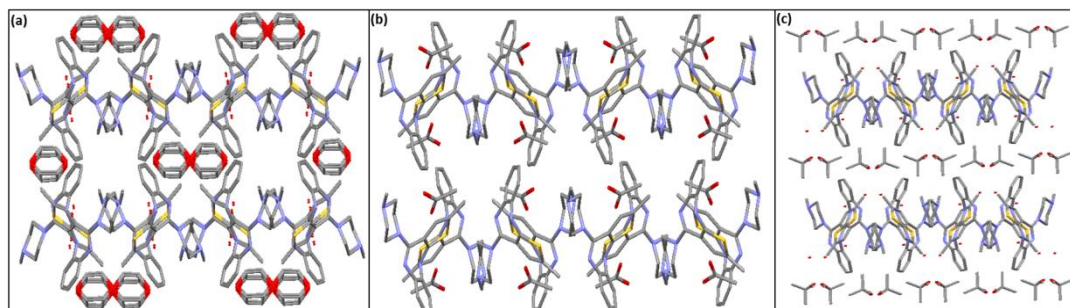


Figure 6.28. Representative crystal structures of olanzapine solvates based on (a) SC<sub>31</sub>-olanzapine-1,4-dioxane solvate hydrate; (b) SC<sub>32</sub>-Olanzapine-acetic acid solvate and (c) SC<sub>33</sub>-Olanzapine-TBMe hydrate; showing the position of solvent molecules within the crystal lattice.

### 6.7.1 Random Forests Classification Model – Prediction Results

RF classification model (see Section 5.1.4 and 6.2.4) trained using all 250 physicochemical descriptors of solvents, showed an overall error rate of 25% with 100% accurate prediction rates for OZPN solvates based on SC<sub>33</sub> (Figure 6.29). The confusion matrix summarising the prediction error associated with each 3-D molecular packing type (class) is shown in Table 6.8.



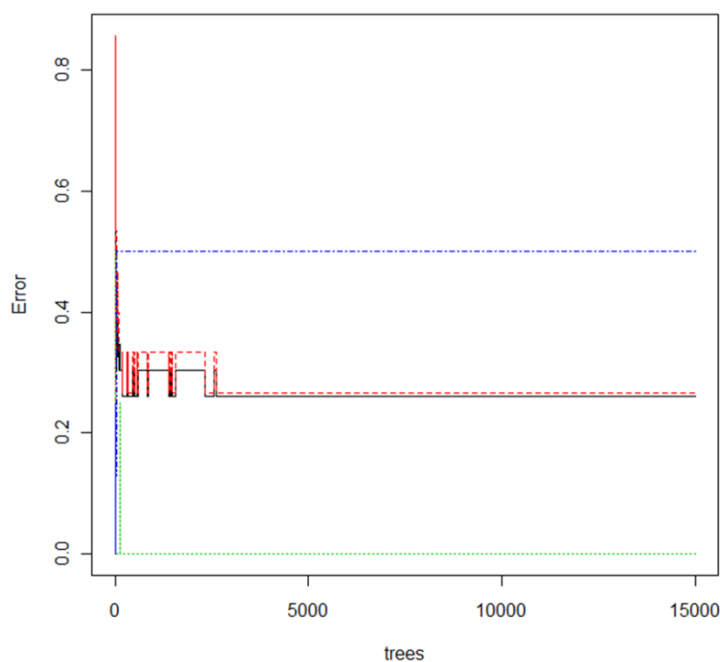


Figure 6.29. Plot of Random Forest classification error for molecular packing types of olanzapine solvates. The black line represents the error associated with the overall model. Red, blue and green lines represent the error associated with the molecular packing type classes;  $SC_{31}$ ,  $SC_{32}$  and  $SC_{33}$ .

Table 6.8. Confusion matrix generated from Random Forest classification for the three molecular packing types of olanzapine solvates.

Predicted $\rightarrow$	Class 1 ( $SC_{31}$ )	Class 2 ( $SC_{32}$ )	Class 3 ( $SC_{33}$ )	Class Error (%)
Observed $\downarrow$				
Class 1 ( $SC_{31}$ )	12	4	0	25
Class 2 ( $SC_{32}$ )	2	2	0	50
Class 3 ( $SC_{33}$ )	0	0	4	0

The MDS plot (see Section 5.1.4) (Figure 6.30) obtained from the RF classification proximity matrix represents the molecular packing types as two orthogonal clusters. Points represented by red, green and blue colour correspond to OZPN solvates based on  $SC_{31}$  (class 1),  $SC_{32}$  (class 2) and  $SC_{33}$  (class 3), respectively.

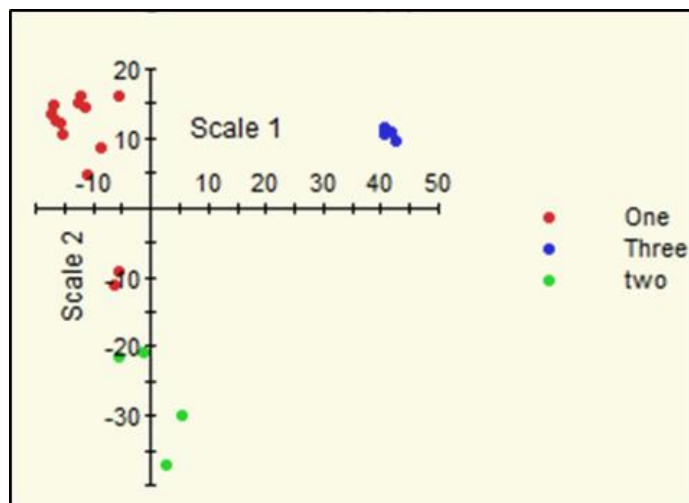


Figure 6.30. The MDS plot obtained from the Random Forest classification proximity matrix. Each of the 24 points represents a olanzapine solvate and is coloured according to crystal packing type: class 1 (red), class 2 (green) and class 3 (blue).

These results indicate that the physicochemical properties of the solvents are directing the crystal packing in OZPN solvates. The OZPN solvates in class 3 have been classified with 100% accuracy, although there is an overlap in the classification of solvates in class 1 and class 2. Four OZPN solvates present in class 1 (methanol solvate, dihydrate E, acetonitrile solvate hydrate and nitromethane solvate hydrate) are classified into class 2 and two OZPN solvates present in class 2 (acetic acid and ethanol solvates) are classified in class 1. It is worth considering why the model failed to predict accurate classes for 6 OZPN solvates.

All the data corresponding to these 6 solvents were removed from the training dataset and the model was trained using the remaining training dataset. Subsequent predictions of molecular packing type of OZPN solvates with these 6

solvents were carried out using the new RF model. Nitromethane and acetonitrile have been given a prediction probability slightly in favour of class 2 (54.5% and 56.1% for nitromethane and acetonitrile respectively) compared to class 1 (44.9% and 43.6% for the nitromethane and acetonitrile, respectively). This model suggests that acetonitrile and nitromethane have a slightly higher probability of forming solvates with OZPN in class 2.

According to this model, methanol and water have been given prediction probabilities for solvate formation with OZPN in favour of class 2 as compared to class 1. Experimentally, methanol and water form solvates with OZPN in class 1 as well as in class 2 depending upon the crystallisation conditions (see Section 6.4.2). The developed RF classification model does not take account of crystallisation conditions and particularly, for methanol and water, crystallisation conditions e.g. residual water in the solvent of crystallisation, % RH, and temperature dictate the crystallisation outcome i.e. molecular packing type in resulting OZPN solvates.

The RF model suggests that ethanol and acetic acid have a higher tendency of forming solvates with OZPN in class 1 (with a prediction probability of 69.5% and 72.6% for ethanol and acetic acid respectively) but still have some probability of forming solvates with OZPN in class 2 (with a prediction probability of 30.5% and 27.4% for ethanol and acetic acid respectively). Experimentally, OZPN ethanol solvate exhibits both types of molecular packing (see Section 6.4.2).

This model was then further tested to predict the packing type for solvents next to ethanol in terms of alkyl chain length and branching. In experimental screening, 1-propanol, 1-butanol and 1,2-propanediol formed solvates with OZPN in the class 1 ( see Section 6.4.2). Similar probability ratios as that of ethanol were obtained for 1-propanol and 1,2-propanediol for both the classes (1 and 2) which suggests that 1-propanol and 1,2-propanediol might also form solvates with OZPN based on SC<sub>32</sub>. The prediction probability for 1-butanol was 90% for class 1 which indicates that it has a higher probability of forming solvate in class 1 than in class 2.

Based on the results obtained so far, it can be hypothesised that because of the different interplanar spacing between SC<sub>21</sub> sheets of OZPN, solvent molecules bigger than 1,2-propanediol cannot be accommodated within the SC<sub>21</sub> sheets or between eclipsed SC<sub>21</sub> sheets to give rise to packing based on SC<sub>32</sub>. Alternatively, the driving force would be towards forming packing based on SC<sub>31</sub>, where alternate SC<sub>21</sub> sheets are arranged in a staggered manner. This packing arrangement leaves a bigger space to accommodate larger solvent molecules than the other two packings based on other SCs and is also stabilised by water molecules located within the SC<sub>21</sub> sheets. Solvent molecules smaller than 1,2-propanediol can adopt any of the two packing types (SC<sub>31</sub> and SC<sub>32</sub>) depending on the crystallisation conditions.

Further experiments are desirable to confirm whether acetonitrile, nitromethane, 1-propanol and 1,2-propanediol would form solvates with OZPN based on SC<sub>32</sub> perhaps by crystallisation using anhydrous solvents. Similarly

further investigations are required to confirm whether acetic acid can form a solvate with OZPN based on SC<sub>31</sub> perhaps by crystallisation in presence of water.

OZPN solvates based on SC<sub>31</sub> and SC<sub>33</sub> have water molecules in the asymmetric unit and so the effect of incorporation of water on the predictive capability of the RF classification model was also investigated. The same results were obtained after taking account of water as a categorical descriptor (present/not present) and even by taking it as a numerical descriptor whose value was equivalent to the number of water molecules in the asymmetric unit of the respective OZPN solvate.

### **6.7.2 Important Solvent Physicochemical Descriptors for Random Forests Classification Model of Olanzapine Solvates**

With specific reference to molecular packing type, the RF classification also provided a rank dependence of OZPN solvate classification on solvent descriptors. The most important descriptors were found to be SMR (molecular refractivity which is a measure of the total polarisability of the molecule), apol (sum of the atomic polarisabilities), vdW\_vol (van der Waals volume) and b\_count (number of covalent bonds).

The variable dependency plots (Figure 6.31, Table 6.9) indicate that the packing based on SC<sub>33</sub> is favoured by large solvent molecules and these can be easily distinguished from other solvent molecules forming solvates in other two classes (numerical values of descriptors are provided in Table 6.9). Smaller

solvent molecules like water and methanol favour the packing based on SC<sub>32</sub>. Although there is some overlap in values of descriptors of the solvents forming solvates based on SC<sub>31</sub> and SC<sub>32</sub> (Table 6.9). Depending on the crystallisation conditions (RH, residual water in crystallisation solvent, temperature and supersaturation) some solvents that form solvates based on SC<sub>32</sub> can form solvates based on SC<sub>31</sub> with OZPN.

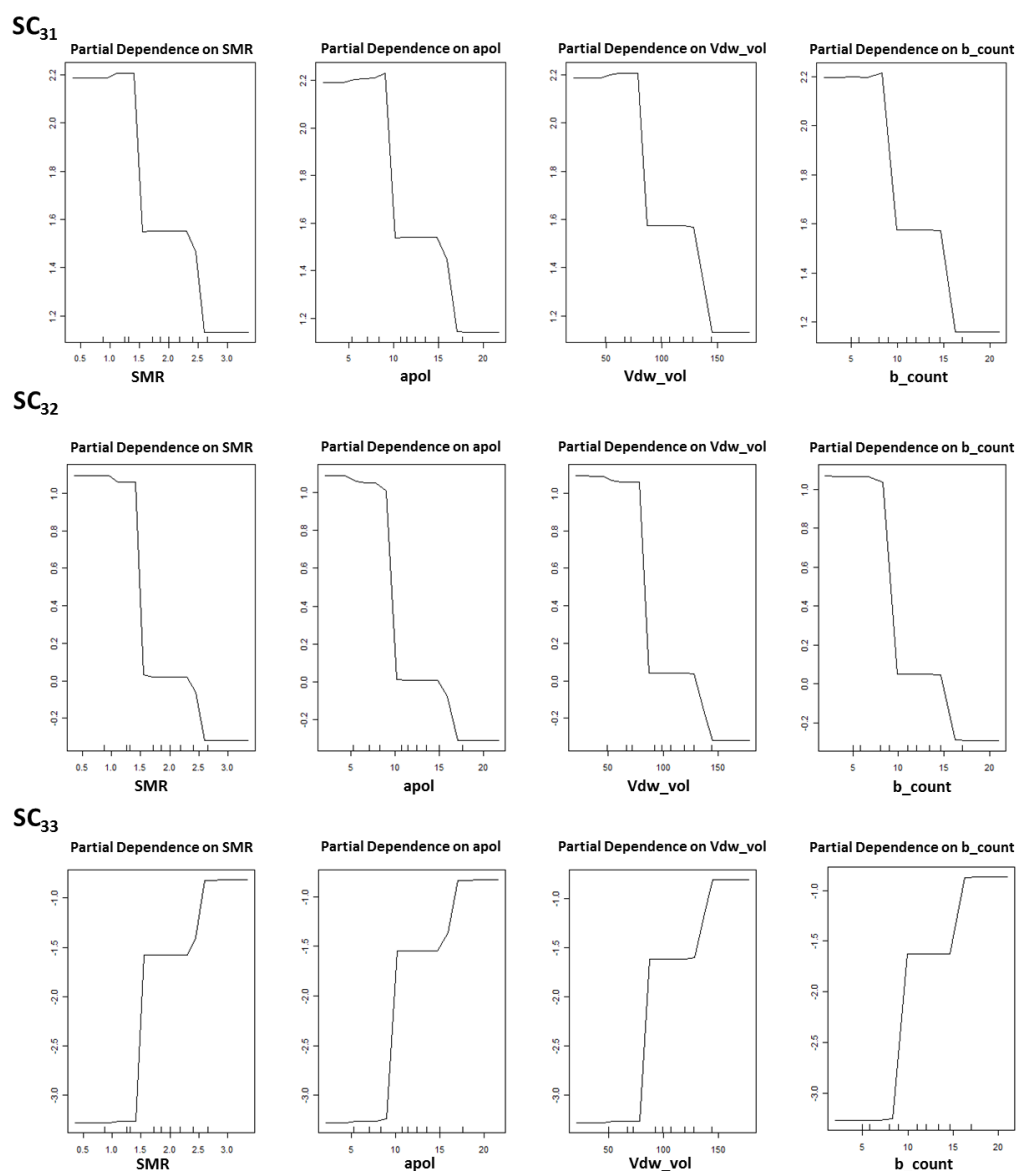


Figure 6.31. Variable dependency plots (SMR, apol, VdW\_vol and b\_count) for molecular packing types, SC<sub>31</sub>, SC<sub>32</sub> and SC<sub>33</sub>. Numerical values for solvent descriptors are summarised in Table 6.9.

Table 6.9. Numerical values of solvent descriptors obtained from variable dependency plots generated during Random Forests classification of olanzapine solvates.

<b>Packing type</b>	<b>SMR</b>	<b>apol</b>	<b>VdW_vol</b>	<b>b_count</b>
SC <sub>31</sub>	0.8-2.4	5.2-14.5	59.0-132.9	5-15
SC <sub>32</sub>	0.3-1.3	<8.3	<72.5	<8
SC <sub>33</sub>	>2.6	>17.6	>144.6	>17

To summarise, the developed RF classification model suggests that the crystal packing of OZPN solvates is dependent on the physicochemical properties of the solvent molecules incorporated in the crystal lattice. Largest solvent molecules tend to form solvates based on SC<sub>33</sub> and using this classification model, these can be predicted more accurately than the solvents which tend to form solvates based on SC<sub>31</sub> and SC<sub>32</sub>. This model also suggests that the solvent molecules bigger than 1,2-propanediol have a higher tendency to exhibit packing based on SC<sub>31</sub>. Solvent molecules smaller than 1,2-propanediol can exhibit packing based on both SC<sub>31</sub> and SC<sub>32</sub> depending on the crystallisation conditions. Further crystallisation experiments are required to confirm this.

## 6.8 Crystal Energy Landscape of Olanzapine

The computational search (Figure 6.32) found both the experimentally observed non-solvated polymorphs, reproducing the structures well (Figure 6.33). Both are present in the low energy region below about  $-125 \text{ kJ mol}^{-1}$ , in the crystal energy landscape.

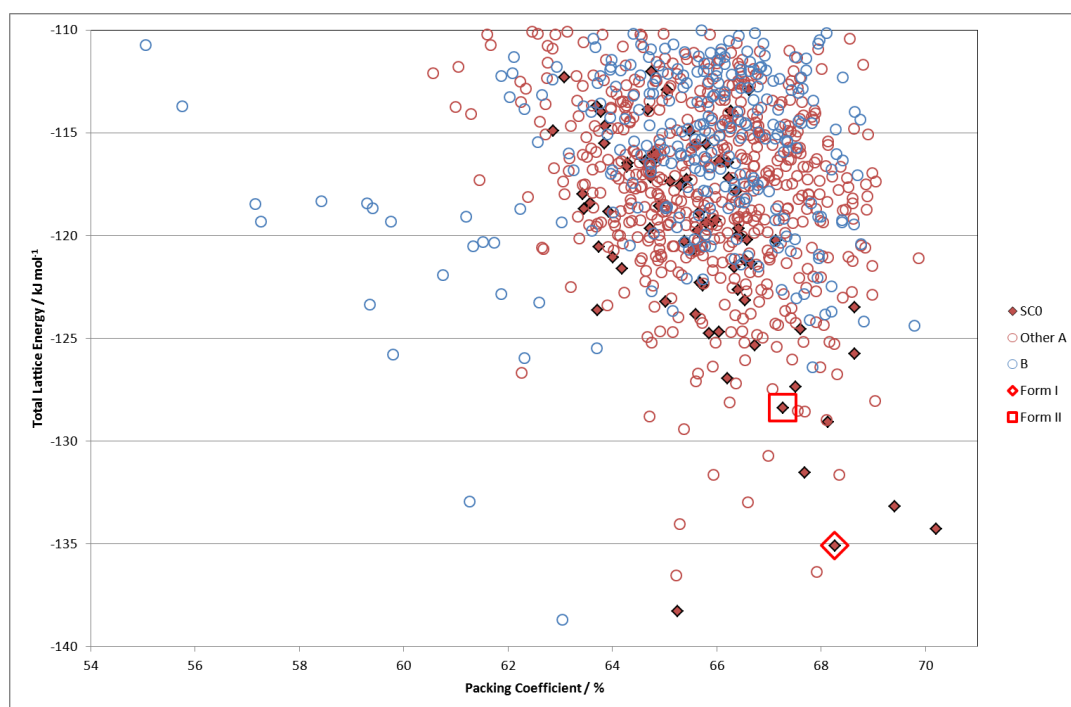


Figure 6.32. Summary of crystal structure search for olanzapine. Each point represents a crystal structure classified by the conformational region, A or B, as defined in Figure 6.4, and whether it contains the centrosymmetric dimer SC<sub>0</sub>. The open red diamond and square symbols denote lattice energy minima corresponding to experimental structures, forms I and II respectively. The crystal energy landscape is comprised of the structures below -125 kJ mol<sup>-1</sup>. A11 and A45 are the computed structures corresponding to form I and II respectively.

Above this value there are a plethora of crystal structures whose relative lattice energies are sufficiently high to suggest that they are less likely to be viable experimental forms. Hence this crystal energy landscape is reasonably realistic, though there is some uncertainty in the relative energies.



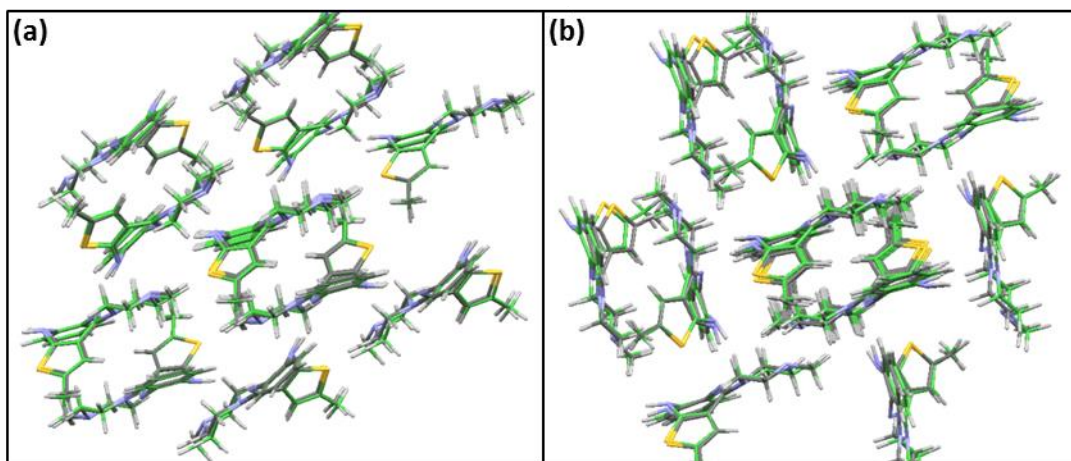


Figure 6.33. Overlay of (a) experimental form I (colour by element) with A11 (green) and (b) experimental form II (colour by element) with A45 (green). The  $\text{RMSD}_{15}$  of these overlays are (a) 0.170 Å and (b) 0.225 Å.

The crystal energy landscape shows that there are other structures which are competitive in energy with the known forms and this is generally consistent with the observation of polymorphism and multiple solvates of this molecule. The global minimum corresponds to a structure with a different conformation (conformation B, Figure 6.4) from the experimental structures, although low energy computed structures based on the A-conformer do tend to be more efficiently packed than those based on the alternative, B. The majority of the structures on the landscape do not contain the  $\text{SC}_0$  motif that is found in all of the experimental polymorphs and solvates observed to date. The predictions therefore provide evidence that alternative packing arrangements are thermodynamically competitive. This suggests that kinetic effects are responsible for all the observed structures being based on the centrosymmetric dimer.

Another notable feature of the landscape is the occurrence of low energy structures that display 2-D similarity with the different layers found in form I and form II but with alternative stackings and consequent differences in interlayer hydrocarbon-hydrocarbon interactions. For example, the hypothetical structure A1392 shows 2-D similarity to form I (SC<sub>23</sub>) and the different arrangement of layers in these two structures is shown in Figure 6.34.

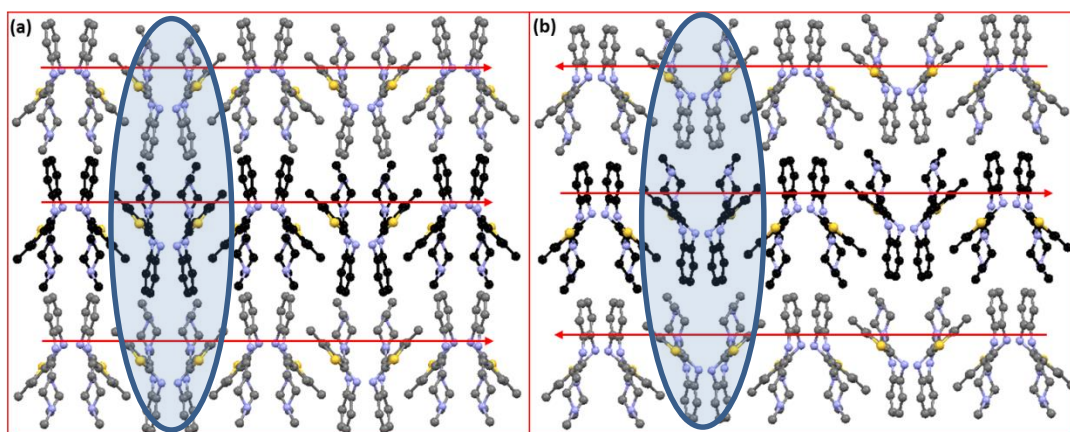


Figure 6.34. Comparison of (a) form I (calculated structure A11) and (b) calculated structure A1392, showing the arrangements of the common 2-D SCs; SC<sub>23</sub> (based on stacking of SC<sub>12</sub>). The constructs are viewed along the axis of the  $t_6$  vector with the  $t_8$  axis (red arrows) running left-right across the page for the central layers, which are shown in black. In form I the upper and lower layers (grey) stack in parallel to the middle layer (black) while in 1392 they are rotated such that the top and bottom layers pack with opposite direction to that of the central layer. SC<sub>23</sub> is not defined in Figure 6.18, as it only occurs in one of the experimental structures, i.e. form I.

Similarly, A162 and form II share a 2-D SC, SC<sub>24</sub> however; the structures differ in the relative orientations of this layer (Figure 6.35). The lattice energy difference between form I and A1392 (3.56 kJ mol<sup>-1</sup>) is relatively large, producing a significant thermodynamic driving force favouring growth based on the stacking

of form I. In contrast, the small difference of 1 kJ mol<sup>-1</sup> between form II and A162 suggests this change in stacking may be thermodynamically competitive.

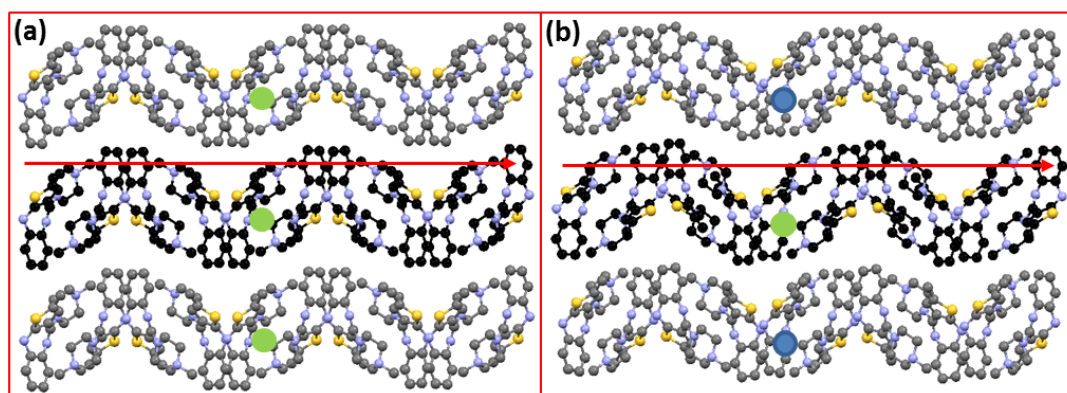


Figure 6.35. Comparison of (a) form II and (b) A162, showing the arrangement of the common 2-D SC, SC<sub>24</sub>. The central constructs (black) are viewed along the axis of the  $t_{11}$  vector (shown by the green circles) with the  $t_{10}$  axis running left-right across the page as indicated by the red arrows. The upper and lower layers in form II (grey) run parallel with  $t_{11}$  axis of the central construct while in A162 alternate layers pack in opposite directions ( $-t_{11}$  indicated by the blue circles). SC<sub>24</sub>, not shown in Figure 6.18 as this 2-D construct was only observed in form II of the experimental structures included in the XPac analysis.

The lack of availability of a suitable single-crystal or phase-pure polycrystalline samples of form III has precluded attempts to define in detail the structural relationship between forms II and III. In an alternative approach to identify a reliable structural model for form III, the unit cell and space group of each of the 50 lowest-energy predicted structures were used along with the form II unit cell in a series of two-phase Pawley-type refinements [using TOPAS version 3 in batch mode, (Coelho, 2003b)] against the best available mixed-phase form III pattern (desolvated DCM solvate; ~20% form II). The best visual and statistical fit was obtained with A162 ( $R_{wp}=6.152$ , Figure 6.36). The observed fit suggests a

degree of similarity, however the refinement did not account for several observed peaks attributed to form III. Thus, A162 is not a reliable structural model for form III and either the unit cell or space group is incorrect. Thus, assuming that form III is a  $Z'=1$  ordered crystal structure, as indicated by the SS-NMR (Reutzel-Edens et al., 2003), it was either missed in the search, or was sufficiently similar to A162 for the clustering algorithms to have not considered the structures unique.

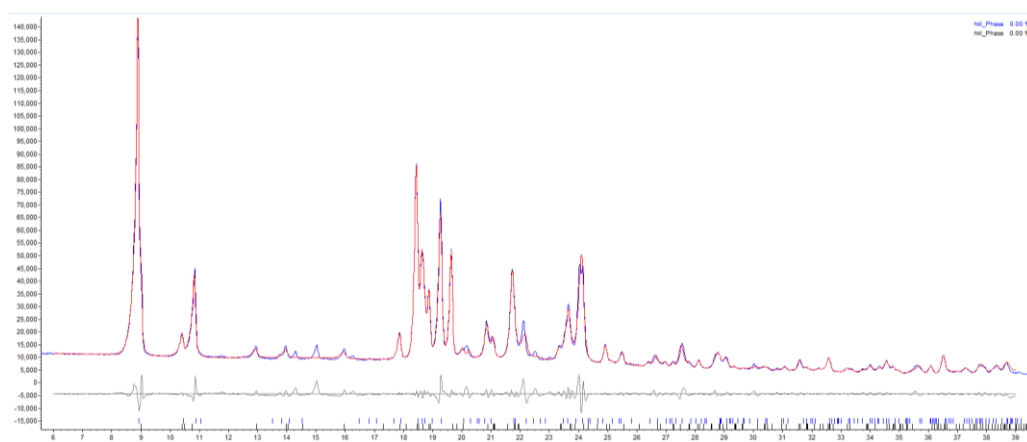


Figure 6.36. Multi-phase Pawley fit of unit cell of A162 against XRPD data of mixture of form III (major) and II (minor), observed profile (blue), calculated profile (red) and difference plot (black) in the range of 6-39° 2 $\theta$ .  $R_{wp} = 6.152$ .

## 6.9 PIXEL Calculations on Olanzapine

The PIXEL calculations provide an alternative method of estimating the intermolecular contributions to the lattice energy, based on the interactions between the electronic density in a model of its charge distribution. The intermolecular lattice energy difference between forms I and II varies from 0.7 kJ mol<sup>-1</sup> to 5 kJ mol<sup>-1</sup> depending on whether 298 K or 123 K experimental

structures or the 0 K computational lattice energy minimum is used in the calculations (Table 6.10), emphasising the sensitivity of this energy to the precise geometry of the crystal. However, the PIXEL calculations clearly show that by far the largest contribution to crystal stabilisation comes from dispersion energy ( $E_d$ ).

Table 6.10. Partitioned intermolecular interaction energies ( $\text{kJ mol}^{-1}$ ) from the PIXEL calculations of form I and II structures at 298 K, 123 K and 0 K.

Structure	Temperature (K)	$E_c^a$	$E_p^b$	$E_d^c$	$E_r^d$	$E_t^e$	Density $\text{g cc}^{-1}$
Form I	298	-67.0	-32.3	-180.7	139.0	-141.0	1.296
	123	-75.5	-38.3	-200.3	170.3	-143.8	1.340
	0	-57.3	-27.6	-168.4	124.1	-129.2	1.263
Form II	298	-64.6	-40.6	-179.7	139.2	-145.7	1.279
	123	-74.9	-42.3	-194.1	168.1	-143.1	1.308
	0	-47.1	-25.4	-165.3	113.6	-124.2	1.241

<sup>a</sup> Electrostatic (Coulombic) energy, <sup>b</sup> Polarisation term, <sup>c</sup> Dispersion term, <sup>d</sup> Repulsion term, <sup>e</sup> Total intermolecular interaction energy,  $E_t = E_c + E_p + E_d + E_r$

The PIXEL calculations on form I, II identified several dimer arrangements within the  $2.5 \text{ kJ mol}^{-1}$  threshold. Analysis of the crystal packing in terms of the individual pairwise molecular interactions (Table 6.11, Figure 6.37) shows that the overall intermolecular energy in each form is dominated by the contribution from  $SC_0$ , which is predominantly stabilised by the  $E_d$ .

Table 6.11. Intermolecular interaction energies ( $\text{kJ mol}^{-1}$ ) for the top three ranked pairwise interactions in experimental form I and II at 298 K, 123 K and in predicted form I and II at 0 K.

Structure	Temperature (K)	Rank <sup>a</sup>	$E_c$	$E_p$	$E_d$	$E_r$	$E_t$	COM ( $\text{\AA}$ ) <sup>b</sup>
Form I	298	1 <sup>st</sup>	-19.9	-11.3	-93.5	63.7	-60.9	4.019
	123	1 <sup>st</sup>	-26.4	-15.7	-104.3	81.7	-64.7	3.929
	0	1 <sup>st</sup>	-19.0	-9.9	-87.5	57.4	-59.0	4.127
	298	2 <sup>nd</sup>	-44.4	-20.1	-52.7	66.7	-50.5	6.440
	123	2 <sup>nd</sup>	-45.7	-21.6	-57.1	75.9	-48.5	6.376
	0	2 <sup>nd</sup>	-34.8	-15.5	-47.9	53.4	-44.7	6.552
	298	3 <sup>rd</sup>	-3.2	-1.4	-18.8	9.3	-14.0	10.388
	123	3 <sup>rd</sup>	-7.0	-3.7	-18.5	18.5	-10.8	9.739
	0	3 <sup>rd</sup>	-2.0	-0.9	-14.9	5.6	-8.7	10.400
Form II	298	1 <sup>st</sup>	-22.3	-13.7	-95.9	70.0	-61.9	4.042
	123	1 <sup>st</sup>	-26.4	-17.2	-104.4	85.0	-62.9	3.977
	0	1 <sup>st</sup>	-20.0	-10.2	-91.2	62.3	-59.0	4.145
	298	2 <sup>nd</sup>	-35.1	-16.8	-50.0	56.3	-45.6	6.593
	123	2 <sup>nd</sup>	-39.6	-19.4	-53.2	66.6	-45.6	6.553
	0	2 <sup>nd</sup>	-21.5	-11.5	-45.5	43.8	-34.7	6.617
	298	3 <sup>rd</sup>	-5.3	-2.9	-22.9	16.2	-14.8	9.913
	123	3 <sup>rd</sup>	-6.7	-3.8	-24.2	20.1	-14.6	9.854
	0	3 <sup>rd</sup>	-3.7	-1.7	-19.5	10.4	-14.5	10.000

<sup>a</sup> Rank contribution of a pair of molecules to the overall intermolecular interaction energy in the structure studied, <sup>b</sup> Distance between the centres of mass of the two molecules ( $\text{\AA}$ )

This centrosymmetric dimer, which is a top contributor towards stabilising energy may be regarded as one of the fundamental structure units directing the assembly of OZPN molecules in crystal lattice (Dunitz and Gavezzotti, 2005b). The 2<sup>nd</sup> ranked pairwise intermolecular interaction, which includes contributions from the N1H...N4 hydrogen bond, is over  $10 \text{ kJ mol}^{-1}$  less

favourable in forms I and II at all temperatures. These two molecule-molecule interactions are significantly more favourable than the others within the coordination shell.

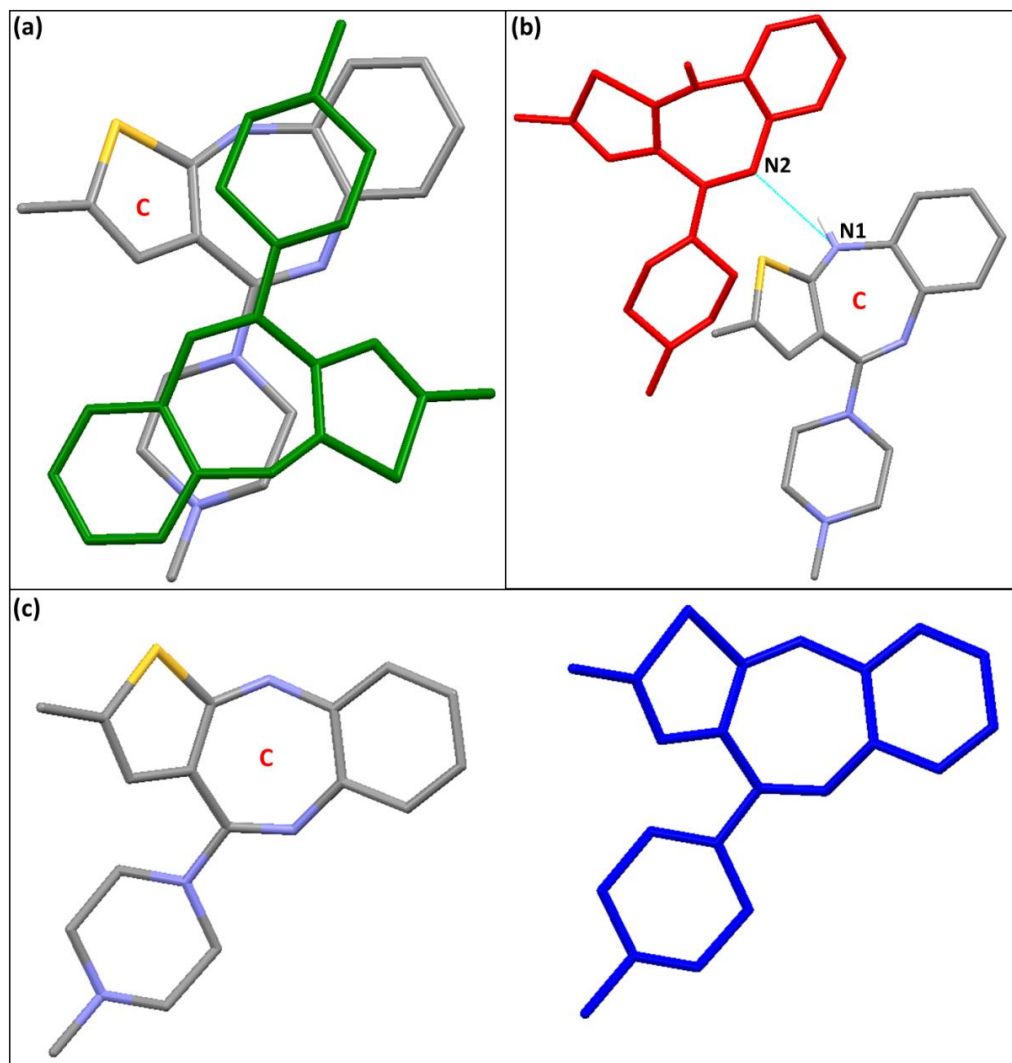


Figure 6.37. Top three pairwise intermolecular interactions in form I and form II, quantified in Table 6.11. The central molecule (colour by atom type and labelled as 'C') interacts with neighbouring molecule to form a) centrosymmetric dimer, SC<sub>0</sub>, b) molecular pair involving hydrogen-bond (N1-H...N2), and c) molecular pair involving weaker molecule-molecule interaction.

Dispersion forces clearly have a key role in directing the pairwise assembly of OZPN molecules to form the SC<sub>0</sub> unit that underpins all of the experimental

structures. PIXEL calculations on two non-dimer based low energy computed structures (A1 and A30) were carried out to explore the relative contributions of the pairwise intermolecular interactions in the absence of the dispersion stabilised SC<sub>0</sub>. As with forms I and II, the total interaction energy is dominated by the E<sub>d</sub> (Table 6.12).

Table 6.12. Partitioned intermolecular interaction energies (kJ mol<sup>-1</sup>) from the PIXEL calculations of hypothetical structures A1 and A30.

Structure	E <sub>c</sub>	E <sub>p</sub>	E <sub>d</sub>	E <sub>r</sub>	E <sub>t</sub>
A1	-57.8	-28.7	-167.7	121.2	-133.0
A30	-78.9	-40.0	-158.2	157.6	-119.4

However the largest interaction in the coordination sphere is significantly smaller than that in the SC<sub>0</sub> dimer in forms I and II (by a factor of 0.55 for A1 and 0.67 for A30) and is dominated by the contribution of electrostatic energy (E<sub>c</sub>) contribution associated with alternative the N1-H...N4 hydrogen bonds (Table 6.13).

Table 6.13. Intermolecular interaction energies (kJ mol<sup>-1</sup>) for the top two ranked dimer interactions in computed structures, A1 and A30 at 0 K.

Structure	Rank	E <sub>c</sub>	E <sub>p</sub>	E <sub>d</sub>	E <sub>r</sub>	E <sub>t</sub>	COM (Å)
A1	1 <sup>st</sup>	-26.9	-13.9	-34.6	42.8	-32.6	8.594
A30	1 <sup>st</sup>	-58.1	-31.5	-51.3	101.4	-39.5	7.184
A1	2 <sup>nd</sup>	-11.4	-4.5	-34.4	19.1	-31.2	8.105
A30	2 <sup>nd</sup>	-13.2	-4.8	-47.1	33.4	-31.7	6.536



## **6.10 Concomitant Appearance of Form III with other Polymorphs of Olanzapine**

It seems highly likely that the crystal structure of form III is also based on SC<sub>0</sub>. SS-NMR (Reutzel-Edens et al., 2003), Raman and IR provides supporting evidence that the structures of forms II and III are closely related. Although the hypothetical structure A162 is not a reliable structural model for form III, the similarities observed from comparison with the powder data and the small energy and structural difference from form II, suggest that form III is based on an alternative packing of the layers in form II (Figure 6.35). This type of structural similarity implies that it would be challenging to find conditions to nucleate them separately, which is supported by current and previous (Reutzel-Edens et al., 2003) experimental results. Indeed the small energy difference between the two ways of stacking the layers of form II implies that there could be a range of structures, differing in the sequence of the two types of stacking that would be competitive in energy; however, no evidence of such disorder in stacking is apparent in XRPD patterns. Although single-crystals of form II have been produced by growth from the vapour phase, form III has never been obtained in pure form and is always observed to co-exist with form II, though never with form I alone. There are several explanations for the concomitant occurrence of these two OZPN polymorphs that include the promotion of nucleation of one form by another through structural templating at the face of the seed form (Tao et al., 2007) or as a form of cross-nucleation at errors in the formation of the interlayer interactions, as seen for ROY (Chen et al., 2005) or

via a nucleation mechanism that gives rise to polymorphic domains as observed for aspirin (Bond et al., 2007). In the experiments conducted in this study that produced form III (*in-situ* desolvation, vapour phase crystallisation and grinding) it has not been possible to establish even a qualitative distinction between the growth kinetics of forms II and III, with both forms appearing simultaneously within the temporal resolution of the measurements.

### **6.11 Prolific Solvate Formation of Olanzapine**

The experimental search has highlighted the extensive tendency of OZPN to form crystalline solvates and mixed solvates forming a total of 56 different solvated forms. In a study of solvate formation based on entries in the CSD, molecules with non-coplanar aromatic rings were found to be one of three categories of molecules with a higher than average tendency to form solvates (van de Streek, 2007). Thus OZPN solvate formation appears to be related to the shape of the benzodiazepine moiety with the  $SC_0$  dimers able to form a range of extended structures that can host a wide range of solvent molecules, stabilising the structure through filling void space and/or hydrogen bonding to surrounding OZPN molecules. The majority of solvates (38/56) are mixed solvates that include at least one water molecule as well as the organic solvent within the asymmetric unit. In four cases, methanol, ethanol, DCM and 2,2,2-trifluoroethanol, both solvate and solvate hydrate forms have been obtained depending on the crystallisation method used. In all the mixed solvate structures water is hydrogen bonded to adjacent OZPN molecules, presumably due to its ability to act as a small orientational flexible probe for favourable

electrostatic interactions i.e. it can fill a void and fulfil H-bond interactions inaccessible to the larger and more constrained organic solvent.

The crystal energy landscape (Figure 6.32) also highlights that OZPN molecules have no one good way of packing to give a dense structure, which is consistent with the prolific tendency to form solvates. A large number of the experimental solvates (28/35 structures available for analysis) are based on SC<sub>21</sub> with the space between SC<sub>21</sub> layers able to accommodate a broad range of solvent molecule sizes, shapes and physicochemical properties.

Computational desolvation calculations were used to investigate the role of solvent in structures based on SC<sub>31</sub> (1-butanol hydrate), SC<sub>32</sub> (dihydrate B; AQOMAU01) and SC<sub>33</sub> (t-butanol hydrate) which all share 2-D similarity via SC<sub>21</sub>. “In-silico desolvation” involved removing all the atoms relating to the solvent and relaxing the structure with the CrystalOptimizer model described previously in Section 6.3 (in which the first step, with the molecule held rigid, will usually eliminate much of the void space formerly occupied by solvent) and subsequently refining within the Polarisable Continuum Model. This is not a simulation of desolvation, but more an analysis of the effect of the solvent on the OZPN-OZPN interactions in the solvates. The calculations minimised the structures obtained after removal of the solvent molecule and resulted in structures that are quite high up on the energy landscape confirming that the interactions between SC<sub>21</sub> layers are relatively unfavourable in the absence of solvent. The solvent molecules are actively stabilising the solvate structures, by hydrogen bonding and other favourable interactions to OZPN molecules. This is

reflected in the stability of the majority of the solvates to desolvation under ambient conditions.

The stabilisation of the solvates by solvent molecules is not just a density packing effect as observed in some published reports (Price et al., 2006; Roy et al., 2012), Inspection of the packing efficiency for the various known structures of OZPN (Figure 6.38) shows that the solvates do not generally have a significantly better packing coefficient than the non-solvated forms.

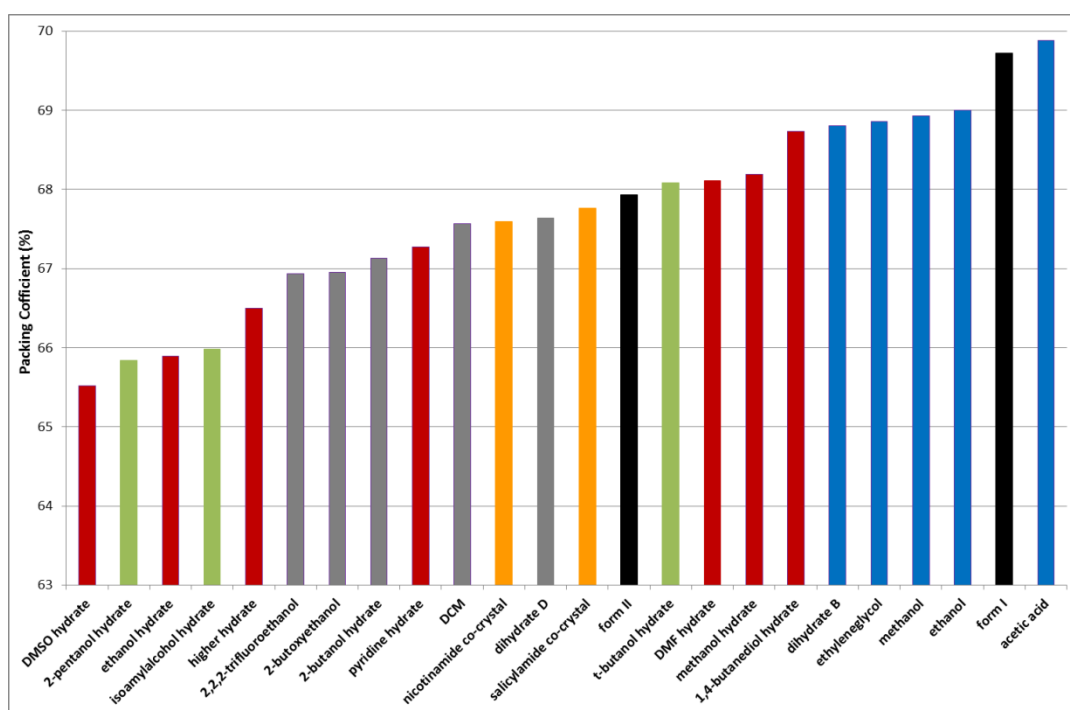


Figure 6.38. Packing coefficient (%) for the non-solvated polymorphs (black), co-crystals (orange) and solvates of OZPN. Solvate structures are coloured according to structural relationships. Specifically, SC<sub>31</sub>, SC<sub>32</sub> and SC<sub>33</sub> are shown by red, blue and green bars.

## 6.12 Challenges with Crystal Structure Prediction of Olanzapine and Unobserved Calculated Structures

Considering the uncertainty in the relative energies of the structures on the crystal energy landscape (Figure 6.32), the results are consistent with the experimental observations that form I is the thermodynamically most stable single component phase. PIXEL calculations show that the intermolecular lattice energy is dominated by the dispersion contribution. This universal attractive force is particularly challenging for computational methods to evaluate reliably, and its contribution to the lattice energy increases quite markedly with density (as shown by the PIXEL results in Table 6.10). The problem with evaluating the dispersion contribution between the different parts of the OZPN molecule, contributes in part to the uncertainty in the conformational energy penalty between type A and B structures. Whilst the CSP accurately predicts the structures of both experimental polymorphs amongst the most thermodynamically stable structures, specific alternative predicted structures remain unobserved despite extensive experimental screening. Thus, under the experimental conditions tested, OZPN apparently favours only a subset of all the feasible packings (molecules in conformation A) arranged in the SCs ( $SC_0$  dimer exclusively observed) and lattices identified in the predicted crystal energy landscape. This could be because this calculated crystal energy landscape underestimates the stabilising role of the  $SC_0$  dimer in the formation of experimental structures. Alternatively, during crystallisation the barrier to rearranging to form the dimers in the form I, II or III packing is sufficiently small

that the alternative structures cannot be kinetically trapped as metastable polymorphs. Many of the low energy structures have packing features in common with forms I and II, which may mean that there is no mechanism for nucleating and growing these hypothetical structures in competition with form I. Alternatively, the experimental searches to date may simply not have included conditions that can promote the formation of alternative structures. Another possibility is that solvate formation offers a highly competitive alternative to the formation of many non-solvated predicted structures.

## 6.13 Summary

In this chapter, systematic experimental screening in conjunction with the calculation of the crystal energy landscape using high level theoretical models has been applied to the antipsychotic drug OZPN. An extensive experimental search for solid forms of OZPN identified 60 distinct solid forms including three non-solvated polymorphs (thermodynamically stable form I and metastable forms II and III), 56 crystalline solvates and an amorphous phase. However, no conditions could be identified to produce phase pure samples of the metastable polymorph III and its crystal structure remains elusive. It seems that form III is based on  $SC_0$  and SS-NMR, IR and Raman provide evidence that it is closely related to form II and further work is required to confirm this.

XPac analysis of the 35 experimental crystal structures (30 from this study and 5 from the CSD) containing OZPN showed that they all contain a specific, dispersion-bound, dimer structure,  $SC_0$ , which can adopt various arrangements that accommodate diverse solvents to produce structures with a similar, moderate, packing efficiency to form I. PIXEL calculations showed that the overall intermolecular energy in OZPN polymorphs is dominated by the contribution from  $SC_0$ . Thus,  $SC_0$  can be regarded as the fundamental structural unit which directs the assembly of crystalline aggregates in OZPN crystal formation. Structural similarity analysis using XPac provides a rationale of desolvation products of various OZPN solvates by highlighting the close relationships between the forms and desolvated 'end product'. RF classification revealed that the calculated molecular refractivity, polarisabilities, number of

covalent bonds and van der Waals volume of the solvent molecules are responsible for the molecular packing types in OZPN solvates based on SC<sub>21</sub>. Solvents which tend to form molecular packing type SC<sub>33</sub> can easily be distinguished from solvents which tend to form molecular packing types SC<sub>31</sub> and SC<sub>32</sub>.

The CSP studies found and reproduced both the experimentally observed non-solvated polymorphs on the crystal energy landscape. CSP studies identified a hypothetical structural type that offers an explanation for the inability to obtain the metastable forms II and III separately. It is plausible that form III contains the same SC<sub>24</sub> layers as form II, but stacked in the alternative manner similar to the hypothetical structure, A162, which is only slightly less stable than form II.

Crystal energy landscape shows that the prolific solvate formation of OZPN is due to the inability of OZPN to pack with an efficiency of more than 70% and explains the role of solvent in stabilising solvate structures using computational desolvation studies. H-bonding analysis also revealed that the solvent molecules actively stabilise the crystal lattice of OZPN solvate via hydrogen bonding and other favourable interactions and solvate formation is not solely based on space filling, as shown by the packing efficiency calculations. Due to their small size and orientational freedom, water molecules play a specific role in stabilising the OZPN solvate structures and are present in various H-bonding environments.

The CSP calculations also found thermodynamically feasible structures that do not contain the SC<sub>0</sub>. The inability to obtain these forms despite a diverse range of experimental studies suggests that the rapid kinetics of dimer formation in



solution are responsible for all the observed structures based on the dimer. The experimental challenge remains to achieve sufficient control over molecular self-assembly to promote the formation of alternative growth units that lead to a specific, thermodynamically feasible structure. Whilst comprehensive screening has enhanced the ability to identify multiple forms more completely and in a shorter period of time, CSP methods showed that there may be an opportunity to move beyond physical form discovery based on ease of formation, to target the selection of specific crystal structures using the information that they appear thermodynamically feasible.

## **7 Exploring the Physical Form Landscape of Clozapine, Amoxapine and Loxapine**

## 7.1 Introduction

A detailed investigation of solid form diversity of olanzapine (OZPN) which belongs to group 1 (see Section 2.1, Figure 7.1) using various experimental and computational approaches has been described in Chapter 6. This chapter describes the results of physical form screening of clozapine (CZPN, group 1), loxapine (LXPN, group 2) and amoxapine (AXPN, group 2) (see Section 2.1, Figure 7.1) using multiple crystallisation techniques. CSP studies were used as a complement to experimental screening for these three molecules. Investigation of the experimental crystal structures using PIXEL calculations has identified the dominant pairwise intermolecular interactions and provided a basis to investigate the effect of substituents and solvent incorporation on the dominant intermolecular interactions in the crystal structures of these molecules. In this context it is of particular interest to contrast the different solid-state landscapes and the respective crystal structures for molecules within each group to understand the structural factors that may influence similarities and differences in the packing of these closely related molecules (Figure 7.1).

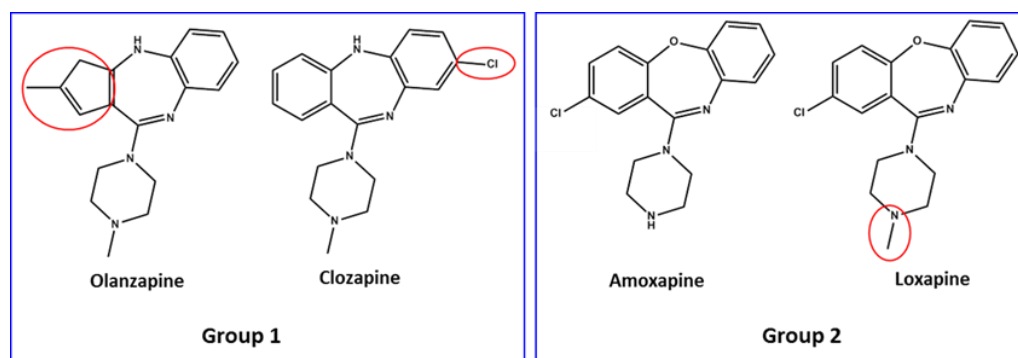


Figure 7.1. The four compounds studied in this thesis. Differences in the chemical structures within a group are highlighted in red.

While it may be expected that small changes in substituents lead to changes in solid-state diversity or crystal packing, the magnitude of these changes and resultant physicochemical properties remains unpredictable (see Section 1.10). Hence it is of continued interest to study solid form diversity of structurally related pharmaceutical compounds. Each of the molecules mentioned above (Figure 7.1) exhibit different reported levels of solid form diversity in terms of the reported numbers of polymorphs and solvates. Given the extent of previous crystallisation studies is not always apparent there is a need to carry out further investigations to confirm whether reported structures are representative of the actual solid-state properties of the molecule rather than any limitation in the scope of previous crystallisation searches.

### **7.1.1 Background of Molecule in Group 1-Clozapine**

CZPN, [8-chloro- 11-(4-methyl- 1-piperazinyl)-3H-dibenzo (b,e)(1,4) diazepine], is a benzodiazepine derivative and an atypical antipsychotic drug used in the treatment of schizophrenia, which affects approximately 1% of the world's population (Jablensky et al., 1987). It is effective against treatment-resistant schizophrenia (Ashby and Wang, 1996). Crystal structures of CZPN [CSD refcode: NDNHCL01 (Petcher and Weber, 1976) and NDNHCL10 (Fillers and Hawkinson, 1982b)] have been reported. The crystal structure of CZPN monohydrate was independently determined as a part of this doctoral research at the same time as a published report (Siva Lakshmi Devi et al., 2011). Crystal structures of four arylmethyl analogues of CZPN (Capuano et al., 2000; Capuano et al., 2005; Capuano et al., 2010), a monohydrate of 2-chloro isomer of CZPN

(Petcher and Weber, 1976) and a dihydrobromide salt (Fillers and Hawkinson, 1982b) are also available in the CSD.

### **7.1.2 Background of Molecules in Group 2-Amoxapine and Loxapine**

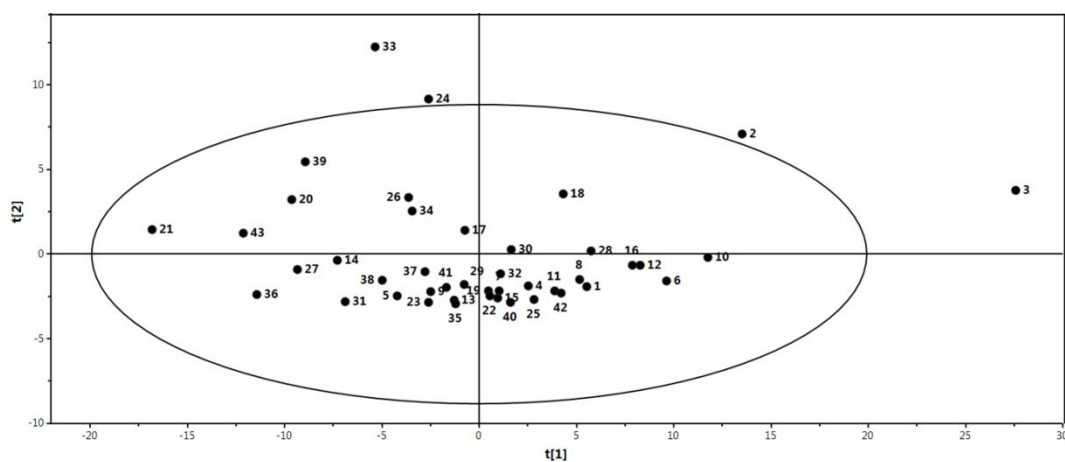
AXPN and LXPN are benzoxazepine derivatives, and differ structurally from one another by a methyl group at position 4 on the piperazine ring. AXPN [2-chloro-11-(1-piperazinyl)-dibenz(b,f)(1,4)oxazepine] has anti-depressant properties (Greenblat and Osterber, 1968) with only one reported crystal structure (CSD refcode: AMOXAP) (Cosulich and Lovell, 1977). LXPN [2-chloro-11-(4-methyl-1-piperazinyl)dibenz(b,f)-(1,4)oxazepine] acts as a tranquilising agent (Latimer, 1969) and is used in the treatment of schizophrenia or bipolar disorder. It is reported to exist in two polymorphic forms (CSD refcode: NDOCLH10 and NDOCLH01 respectively) (Petcher and Weber, 1976; Cosulich and Lovell, 1977) with no reported hydrates/solvates. It is marketed as a succinate salt whose crystal structure has been determined (Fillers and Hawkinson, 1982a).

The methyl group at 4 position of piperazine ring is responsible for the different spectrum of biological activity i.e. anti-schizophrenic vs. anti-depressant for LXPN and AXPN respectively. Comparison of the crystal forms and their structures therefore also allows the effect of this methyl group on the solid-state diversity of these two molecules to be studied.

## 7.2 Experimental Details

### 7.2.1 Principal Component Analysis of Solvent Properties

In order to select a diverse range of solvent properties for an efficient and effective crystallisation search a solvent classification approach was utilised. A PCA (see Section 5.1.2) of solvent property diversity was carried out using the SIMCA package (version 13.0) (Umetrics) (SIMCA, 2012) by inputting calculated physicochemical properties of the solvents. 2-D and 3-D physicochemical descriptors for solvents were calculated using MOE (MOE, 2002). Descriptors which showed zero variance and covariance (threshold of Pearson correlation coefficient >90 %) were removed. All the data were reduced to two principal components (t[1] and t[2]) that described 69.8% of the total variance in solvent properties (observations, n = 43 and descriptors = 250) (Figure 7.2). The 2-D scatter plot from PCA provides a convenient means of cluster identification and/or visualisation of the crystallisation results as a function of solvent identity (Florence et al., 2006).



ID	Solvent	ID	Solvent	ID	Solvent
1	1,5-pentanediol	16	Isobutyl acetate	31	Ethanol
2	1-methyl naphthalene	17	Pyridine	32	Ethyl acetate
3	Dodecane	18	Toluene	33	Carbon tetrachloride
4	N-methyl-2-pyrrolidone	19	1,4-dioxane	34	2,2,2-trifluoro ethanol
5	Dimethyl sulphoxide	20	Nitromethane	35	Tetrahydrofuran
6	2-butoxy ethanol	21	Water	36	Methanol
7	N,N-dimethyl acetamide	22	2-butanol	37	Methyl acetate
8	Cyclohexanol	23	1-propanol	38	Acetone
9	N,N-dimethyl formamide	24	Trichloroethylene	39	Dichloromethane
10	Pentyl acetate	25	1,2-dimethoxyethane	40	Diethyl ether
11	3-methyl-1-butanol	26	1,2-dichloroethane	41	1,2-propanediol
12	Butyl acetate	27	Acetonitrile	42	2-methoxy-2-methylpropane
13	2-methoxy ethanol	28	Cyclohexane	43	Methanoic acid
14	Acetic acid	29	2-butanone		
15	1-butanol	30	1-chlorobutane		

Figure 7.2. Principal component analysis score plot of solvent physicochemical properties, shown with an ellipse that represents the Hotelling T2 with 95% confidence. 1-methylnaphthalene, dodecane, trichloroethylene and carbon tetrachloride were outliers from the 95% confidence limit. Each number in the plot represents one solvent as mentioned below the score plot.

## 7.2.2 Crystallisation Experiments of Clozapine

The experimental screen was based on solution crystallisations using 43 solvents selected from the PCA (see Section 7.2.1). In addition to solution

crystallisation, other crystallisation techniques were also used (see Section 3.2). A total of 157 individual crystallisations were carried out (Table 7.1).

Table 7.1. Summary of the manual crystallisation techniques used in the experimental search for physical forms of clozapine.

<b>Crystallisation Technique</b>	<b>Crystallisation Method</b>	<b>Number of Crystallisations</b>
Solution crystallisation, $T_{\text{sat}} = \text{RT}$	Solvent evaporation in vials and on watch glass at RT	73
Solution crystallisation, $T_{\text{sat}} = \text{B.P.}-10 \text{ K}$	Cooling to 298 K	32
Neat / liquid-assisted grinding (using Retsch MM400 mixer mill or cryo mill)		36
Crystallisation from vapour phase		6
Melt quenching		4
Recrystallisation from amorphous		4
Freeze drying		2

$T_{\text{sat}}$  = temperature at which the saturated solutions were prepared, B.P.= boiling point of the solvent

### 7.2.3 Crystallisation Experiments of Amoxapine

The experimental screen was based on solution crystallisations using 47 solvents selected from the PCA (see Section 7.2.1). Other crystallisation techniques (see Section 3.2) were also used to maximise the crystallisation search space. A total of 138 individual crystallisations were carried out (Table 7.2).



Table 7.2. Summary of the manual crystallisation techniques used in the experimental search for physical forms of amoxapine.

<b>Crystallisation Technique</b>	<b>Crystallisation Method</b>	<b>Number of Crystallisations</b>
Solution crystallisation $T_{\text{sat}} = \text{RT}$	Solvent evaporation in vials at RT and at 277 K	70
Solution crystallisation $T_{\text{sat}} = \text{B.P.}-10 \text{ K}$	Cooling to 298 K	44
Neat/liquid-assisted grinding (using Retsch MM400 mixer mill)		10
Crystallisation from vapour phase		6
Melt quenching		4
Recrystallisation from the amorphous		4

$T_{\text{sat}}$  = temperature at which the saturated solutions were prepared, B.P. = boiling point of the solvent

#### 7.2.4 Crystallisation Experiments of Loxapine

A limited experimental screening was carried out for LXPN due to material and time constraints. The experimental screen was based on solution crystallisations using 32 solvents selected from the PCA (see Section 7.2.1). A total of 96 individual crystallisations were carried out (Table 7.3).

Table 7.3. Summary of the manual crystallisation conditions used in the experimental search for physical forms of loxapine.

<b>Crystallisation Techniques</b>	<b>Crystallisation Method</b>	<b>Number of Crystallisations</b>
Solution crystallisation $T_{\text{sat}} = \text{RT}$	Solvent evaporation in vials and on watch glass at RT	64
Solution crystallisation $T_{\text{sat}} = \text{B.P.} - 10 \text{ K}$	Cooling to 298 K	32

$T_{\text{sat}}$  = temperature at which the saturated solutions were prepared, B.P. = boiling point of the solvent

### 7.2.5 Preliminary Crystal Structure Prediction Studies for Clozapine, Amoxapine and Loxapine

Preliminary CSP studies on CZPN, AXPN and LXPN were carried out to provide a view of thermodynamically feasible packing arrangements. A simple rigid molecule search using MOLPAK/DMACRYS (Holden et al., 1993; Price et al., 2010) with the molecular conformation optimised at the PBE0/6-31G(d,p) level (Figure 7.3), with multipoles calculated from the charge density at the same level. Searches were carried out in space groups  $P1$ ,  $P-1$ ,  $P2_1$ ,  $P2_1/c$ ,  $P2_12_12$ ,  $P2_12_12_1$ ,  $Pna2_1$ ,  $Pca2_1$ ,  $Pbca$ ,  $C2/c$ ,  $Cc$  and  $C2$ , the 12 most common space groups observed for molecules of this size (Kazantsev, Karamertzanis et al. 2011) in the CSD.

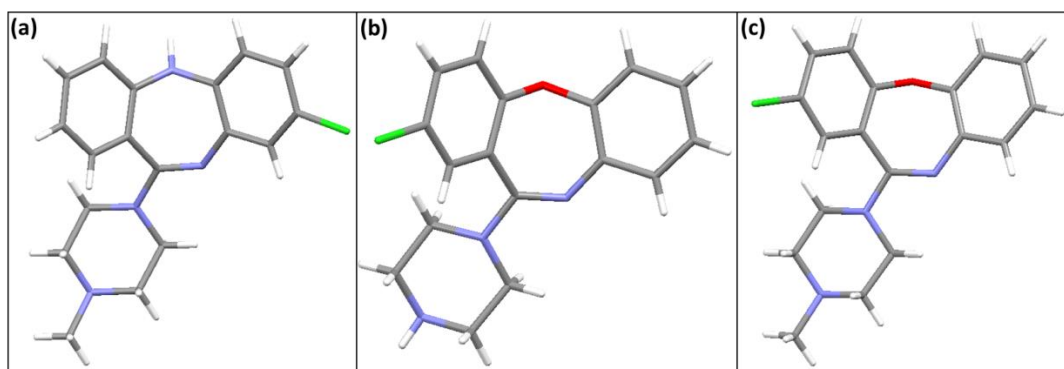


Figure 7.3. Conformation 'A' of (a) clozapine, (b) amoxapine and (c) loxapine used for preliminary crystal structure prediction study and is also present in the experimental crystal structures (CSD refcodes: NDNHCL01, AMOXAP and NDOCLH01 for clozapine, amoxapine and loxapine respectively).

### 7.2.6 Solid-State Calculations using CASTEP

Density Functional Theory (DFT) calculations were carried out using CASTEP (version 4.3 and 5.0)(Clark et al., 2005) on the LXPN hydrogen succinate crystal structure determined from XRPD data. The input files were prepared and results were visualised using Material Studio (MS) software suite (©Accelrys) (version 5.5) interface to CASTEP. Geometry optimisation calculations were carried out using the Broyden-Fletcher-Goldfarb-Shannon (BFGS) algorithm. The lattice parameters and space group symmetry were kept fixed during all the calculations. Exchange-correlation potential was described using a generalised gradient approximation (GGA-PBE). Medium convergence criteria (tolerances of energy per atom, maximum forces and maximum displacement were 0.00002 eV/atom, 0.05 eV/Å and 0.002 Å respectively) were used. Ultrasoft pseudopotentials were used for core electrons as they allow calculations to be performed at the lowest possible cut-off energy for the plane wave basis set. Depending on the size of the unit cell, the number of k-points employed for

sampling was automatically calculated by the MS interface. CASTEP calculations were carried out by distributing on the National Grid Server (UK) facilities and results were analysed using the MS visualiser.

### **7.2.7 Structure Analysis**

All the samples obtained from the crystallisation screens were analysed using XRPD/SXD, thermal analyses and Raman spectroscopy as described in Sections 3.2.1-3.2.4 of Chapter 3. PIXEL calculations and molecular packing analyses were also carried out using the techniques outlined in Sections 3.2.5-3.2.6 of Chapter 3.

## 7.3 Results and Discussion

### 7.3.1 Physical Form Screening of Clozapine

Across all crystallisation conditions (Section 7.2.1), the known anhydrous form (same as the starting material), one monohydrate, two solvates [N,N-dimethylacetamide (DMA) and ethylene glycol (EG)] and an amorphous form were obtained. No polymorphs of CZPN were observed.

Of the 105 solution crystallisations carried out, 5 (4.8%) yielded no sample for XRPD analysis due to poor solubility. Of the remaining 100 samples giving measurable diffraction, 6 were identified as monohydrate, 6 were mixtures of anhydrous form and monohydrate, 3 were recognised as solvates, and 3 were non-crystalline (i.e. they gave no diffraction). All the remaining 82 single phase samples were the known anhydrous form.

After neat grinding of CZPN either at RT or at 77 K, no phase transformation was observed. Solvent assisted grinding of CZPN at RT was carried out using 32 solvents. CZPN monohydrate was obtained after grinding with water and the remaining 31 solvents returned the known anhydrous form. No change was observed after grinding at 77 K in the presence of methanol.

The outcomes of each crystallisation technique are represented by a series of scatter plots (Figure 7.4 and Figure 7.5). Each point in the plot is positioned according to the value of  $t[1]$  and  $t[2]$  from the solvent property PCA score plot (Figure 7.2, see Section 7.2.1) and coloured according to the crystallisation

outcome. A detailed listing of the form(s) identified from each recrystallised sample is provided in the Appendix (Table Apx 1.11- Table Apx 1.14).

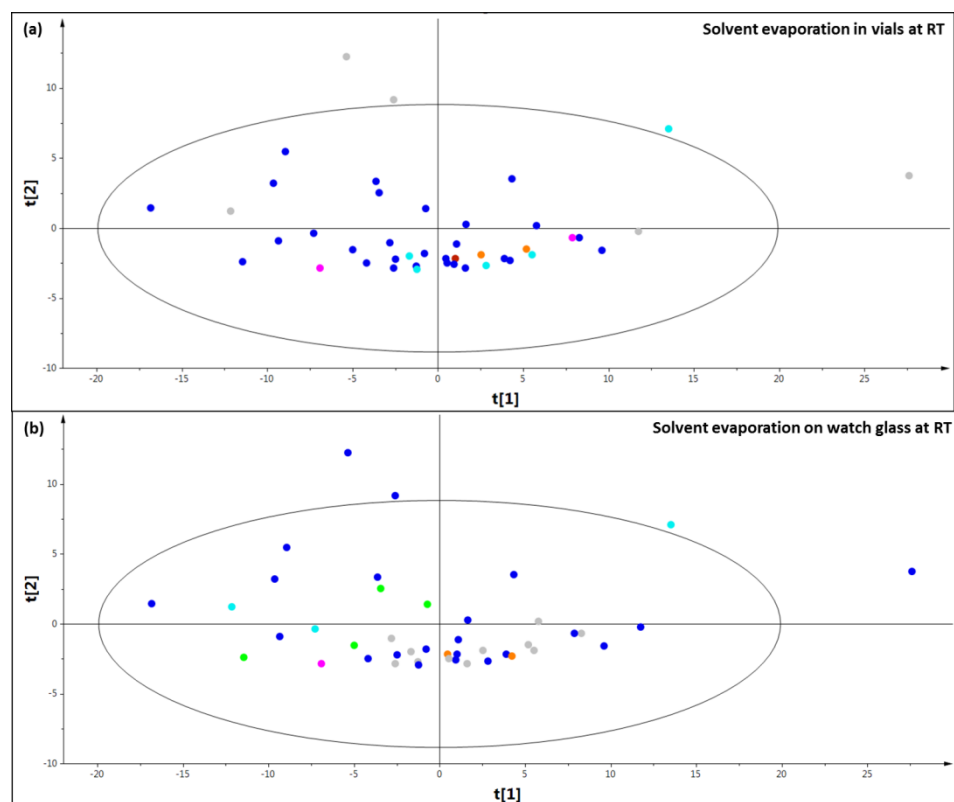


Figure 7.4. Scatter plot of solvents with positions according to the value of  $t[1]$  and  $t[2]$  from Figure 7.2 and coloured according to crystallisation outcome: clozapine anhydrous form (blue), clozapine monohydrate (magenta), mixture of clozapine anhydrous form and clozapine monohydrate (green), amorphous (orange), no sample obtained (cyan) and solvated form (maroon). Solvents not used in crystallisation condition are coloured grey. Plots shown for crystallisation conditions listed in Table 7.1: (a) solvent evaporation in vials at RT and (b) solvent evaporation on watch glass at RT.

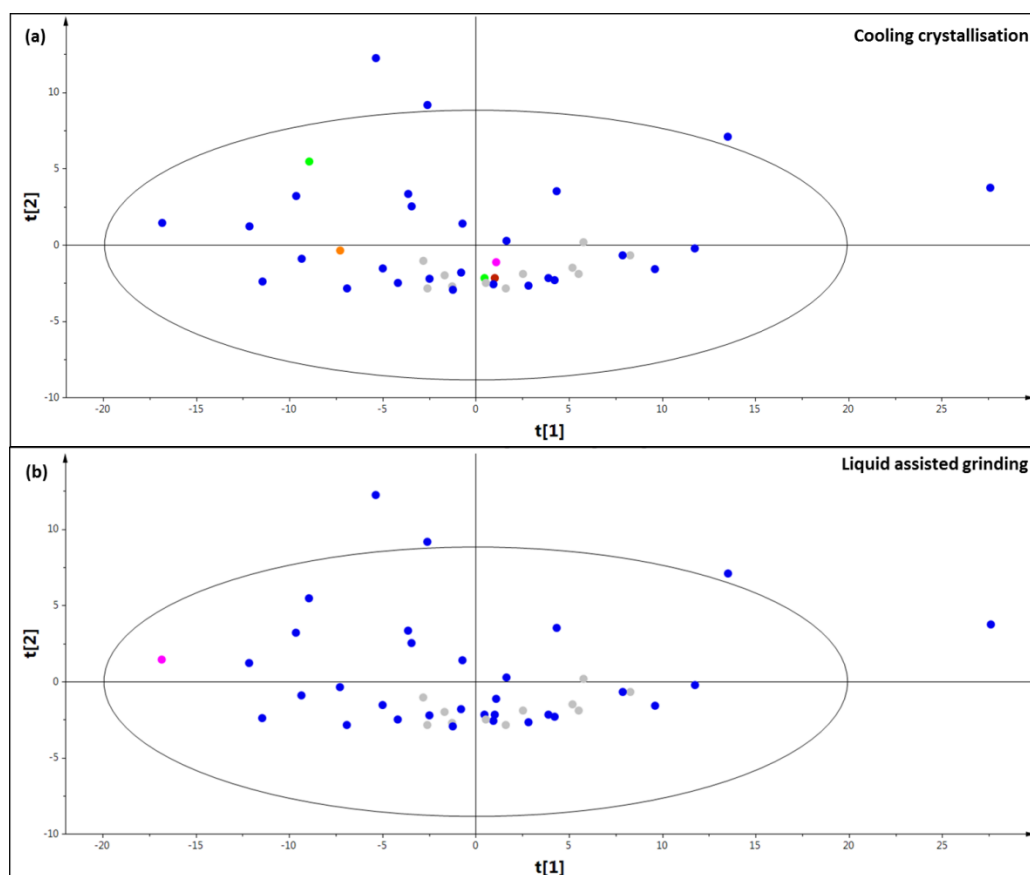


Figure 7.5. Scatter plot of solvents with positions according to the value of  $t[1]$  and  $t[2]$  from Figure 7.2 and coloured according to crystallisation outcome: clozapine anhydrous form (blue), clozapine monohydrate (magenta), mixture of clozapine anhydrous form and clozapine monohydrate (green), amorphous phase (orange), no sample obtained (cyan) and solvated form (maroon). Solvents not used during crystallisation condition are coloured grey. Plots shown for crystallisation conditions listed in Table 7.1: (a) cooling crystallisation and (b) liquid-assisted grinding.

The novel solvated forms obtained comprise a monohydrate, DMA and EG solvates of CZPN. Crystal structures of all the four physical forms were determined by SXD and the lattice parameters are tabulated in Table 7.4. Crystal data, data collection and refinement details are provided in Table Apx 1.15 of the appendix.

Table 7.4. Lattice parameters for clozapine crystal structures obtained in the experimental search (standard uncertainties in parentheses). All parameters were determined from single-crystal diffraction data collected at 123(2) K.

Structure ID	Form (CZPN: solvate)	Space Group	Stoichiometry CZPN:solvent	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\beta$ (°)	Cell Volume (Å <sup>3</sup> )
1	CZPN	<i>P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></i>	NA	9.328(7)	9.632(7)	17.836 (13)	90.00	1603.0(2)
2	CZPN_monohydrate	<i>P2<sub>1</sub>/c</i>	1:1	9.915(3)	16.365(4)	10.920(3)	102.23(1)	1731.6(1)
3	CZPN_N,N-dimethylacetamide (DMA)	<i>Pbca</i>	1:1	8.866(4)	17.542(7)	28.124(1)	90.00	4374.0(3)
4	CZPN_ethylene glycol (EG)	<i>P2<sub>1</sub>/c</i>	1:1	20.628(1)	10.849(1)	17.502 (1)	93.13(2)	3911.0(3)



In contrast to OZPN (Chapter 6), no evidence of polymorphism was obtained for CZPN and it has a significantly lower propensity for solvate formation. The powder diffraction patterns of all novel forms are shown in Figure 7.6. The XRPD patterns of these forms are unique and do not show signs of isostructurality as observed in OZPN solvates.

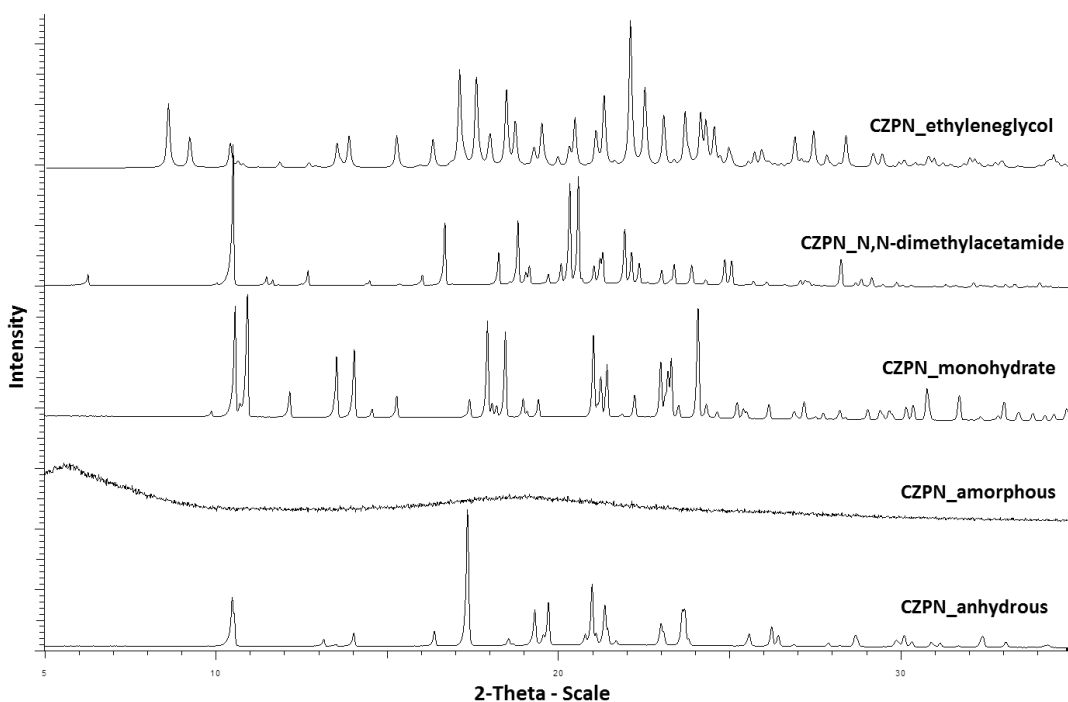


Figure 7.6. Stack plot of XRPD patterns of physical forms of clozapine observed during experimental screening in the range of 5-35° 2 $\theta$ .

STA shows that CZPN melts at 457-458 K with no other thermal transition between RT and the melt (Figure 7.7). The CZPN monohydrate and DMA solvate lose water and DMA with onset temperatures of ~340 K and ~372 K, respectively, before melting at 458 K (Figure 7.7). Due to insufficient quantity of the ethylene glycol solvate being available, STA could not be carried out.

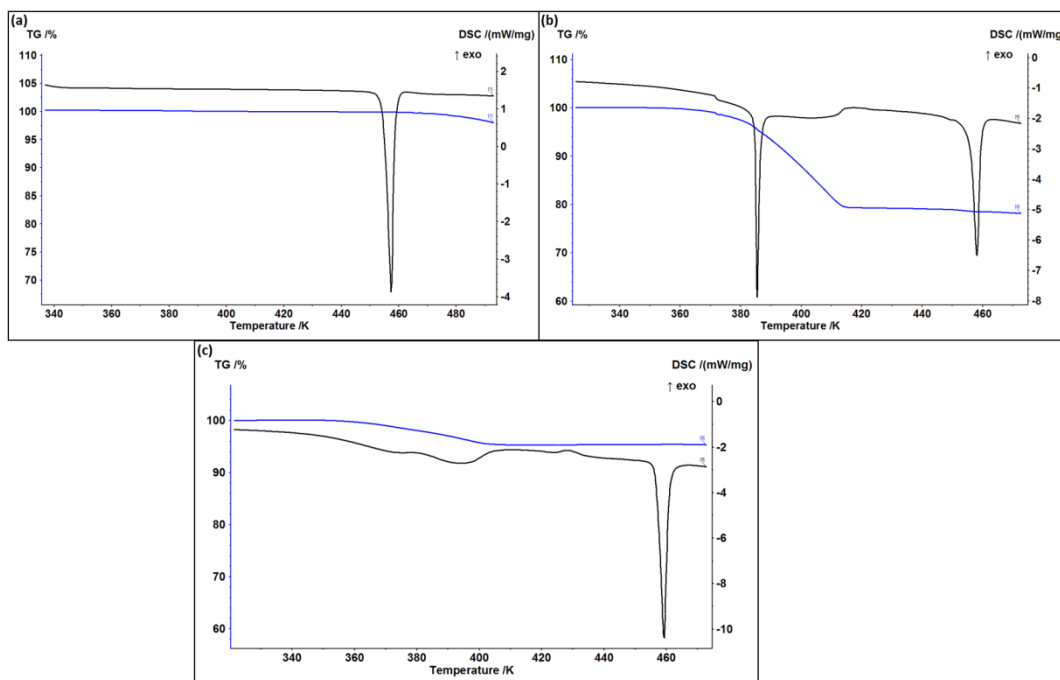


Figure 7.7. STA thermograms of (a) clozapine anhydrous form, (b) clozapine N,N-dimethylacetamide and (c) clozapine monohydrate. DSC and TG curves are shown by black and blue colour, respectively.

There is no evidence that solvate desolvation leads to the formation of further polymorphs. Small exotherms with an onset at  $\sim 411$  K and  $\sim 423$  K in the case of DMA solvate and monohydrate could be due to recrystallisation of partial amorphous form created after desolvation. Further *in-situ* VT-XRPD experiments are required to confirm this.

### 7.3.1.1 Molecular Packing Analysis of Physical Forms of Clozapine

Molecular packing analyses of all the experimental and a few predicted crystal structures of OZPN have been described in Section 6.5 of Chapter 6. This section describes the molecular packing analysis using XPac and Mercury (see Section 3.2.6) carried out on the experimental crystal structures of CZPN to investigate

the effect of the substituents and chemical modifications on the molecular conformation and packing motifs.

Both OZPN and CZPN exhibit similar molecular conformation in all the experimental crystal structures. In this conformation the diazepine ring exists in a puckered conformation between the planes of the two phenyl rings, and the piperazine ring adopts a chair conformation with the methyl group in an equatorial position.

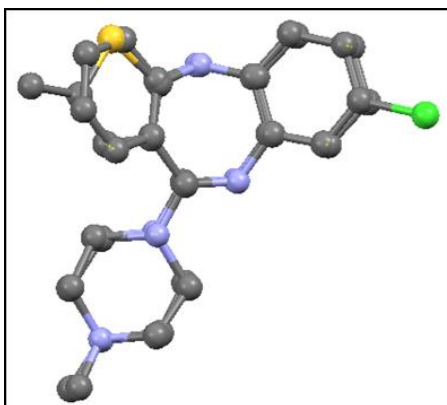


Figure 7.8. Overlay of the molecular conformations of olanzapine and clozapine which are present in the experimental crystal structures (RMSD of non-hydrogen atoms = 0.21 Å).

#### ***7.3.1.1.1 Molecular Packing in Anhydrous Form of Clozapine***

In the anhydrous form of CZPN, only one enantiomer is present in the crystal lattice as compared to all the known OZPN crystal structures where both mirror related enantiomers are present in the crystal lattice. The CZPN molecules form side-ways dimers in which they are arranged in an edge to face arrangement [Figure 7.9(a-b)] as compared to the SC<sub>0</sub> (face to face arrangement of OZPN

molecules) observed in all OZPN known crystal structures (see Chapter 6) and in other physical forms of CZPN explained later in this section (e.g. monohydrate and EG solvate).

These side-ways dimers stack to form columns along the *a*-axis. These columns arrange themselves in a corrugated manner to make the 3-D structure. The CZPN dimers between the columns are connected via N1-H...N2 H-bond (Figure 7.9c). Each CZPN molecule in the dimer is involved in two H-bonds with two other CZPN molecules from the adjacent dimers. A detailed PIXEL analysis of intermolecular interactions in the crystal lattice is presented in Section 7.3.1.3.

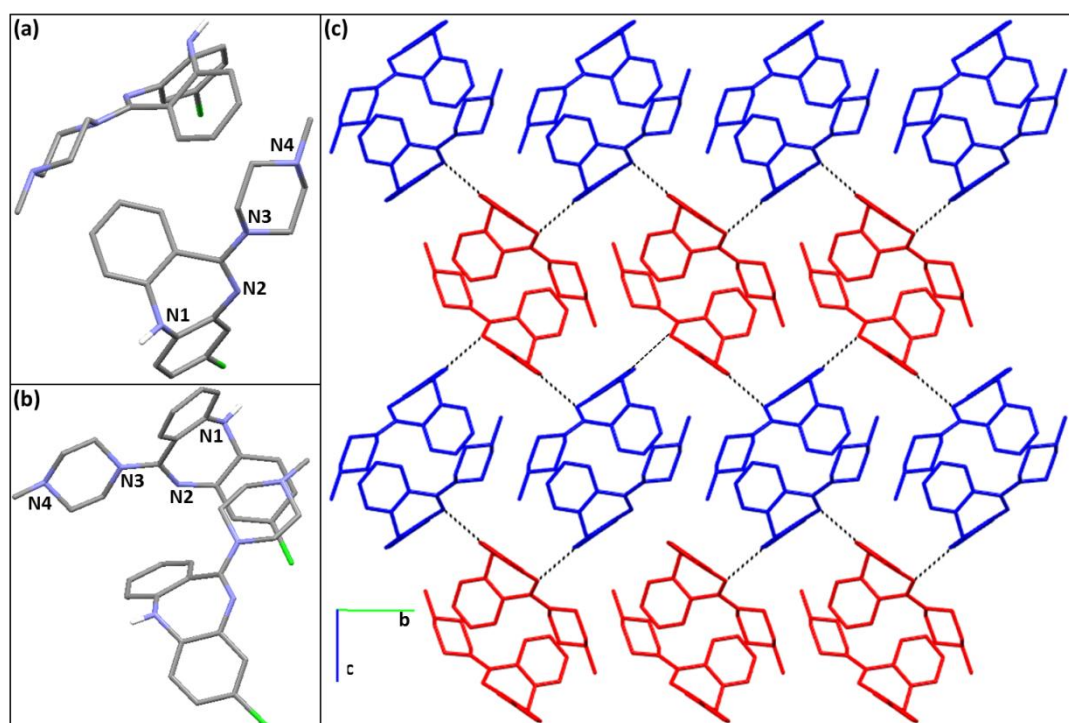


Figure 7.9. (a-b) Two views of the dimer formed by clozapine molecules, (c) 3-D arrangement of columns of clozapine dimers. Active H-bonds between dimers are shown by dotted black lines. Other hydrogen atoms are omitted for clarity.

### 7.3.1.1.2 Molecular Packing Analysis of Clozapine Solvates

In the DMA solvate, each CZPN molecule forms a hydrogen bond with the DMA molecule (N1-H...O). The CZPN molecules form two types of chains (each consists of one enantiomer) along the *c*-direction (Figure 7.10a). Two types of chains stack antiparallel to each other to form 2-D sheets which further stack in a staggered manner (Figure 7.10b).

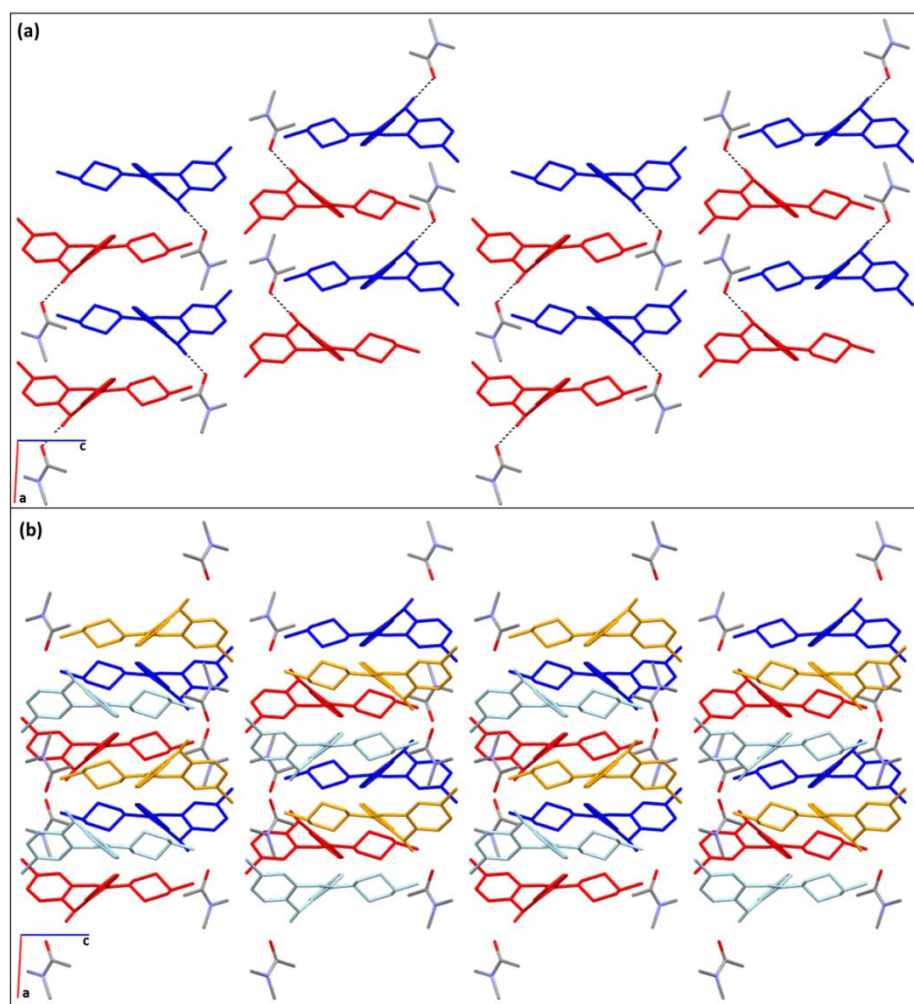


Figure 7.10. (a) 2-D sheet structure formed by the two antiparallel chains of clozapine molecules (shown by red and blue colour) and N,N-dimethylacetamide molecules. H-bonds are shown by black dotted lines. (b) 3-D arrangement of sheets. All the hydrogen atoms are omitted for clarity.

In both CZPN monohydrate and CZPN EG solvate, CZPN molecules form a centrosymmetric dimer which is similar to the SCo present in all crystal structures of OZPN (see Chapter 6). In the CZPN monohydrate structure, centrosymmetric dimers form columns along the *a*-axis. These columns stack in a zig-zag manner to create the 3-D structure (Figure 7.11a). Water molecules are present between the columns and connect the centrosymmetric dimers of parallel columns. Each CZPN molecule is involved in three H-bonds with three water molecules (Figure 7.11a).

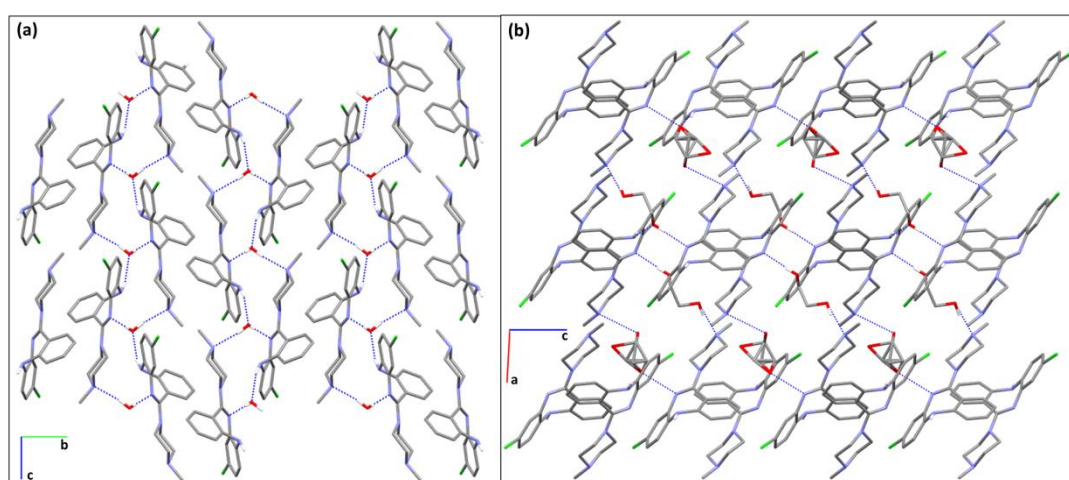


Figure 7.11. 3-D arrangement formed by the columns of the centrosymmetric dimers of clozapine in (a) clozapine monohydrate and (b) clozapine ethylene glycol solvate. H-bonds between clozapine and solvent molecules are shown by the blue dotted line. Other hydrogen atoms are omitted for clarity.

The asymmetric unit of CZPN EG solvate consists of two CZPN and two EG molecules in which one of the EG molecules is disordered. The centrosymmetric dimers form columns along the *b*-axis (Figure 7.11b). Each CZPN molecule is involved in three H-bonds with three EG molecules similar to the solvent interactions in CZPN monohydrate.

The CZPN monohydrate and CZPN EG solvate share 1-D similarity as revealed by the XPac analysis. The 2-D similarity between CPZN monohydrate and EG solvate is lost because of the larger size of EG molecules (molecular volume – 92 Å<sup>3</sup>) present between the columns compared to the smaller water molecules (molecular volume - 24 Å<sup>3</sup>). No other similarity at any level was observed amongst the four CZPN crystal structures and is therefore, not discussed further.

### 7.3.1.1.3 Comparison of Crystal Structures of Clozapine and Olanzapine

Two of the CZPN crystal structures (monohydrate and EG solvate) share only 0-D similarity (i.e. centrosymmetric dimer) with OZPN crystal structures but no packing similarity was observed amongst the anhydrous structures.

Packing efficiencies of form I of OZPN (packing coefficient = 69.7%) and CZPN (packing coefficient = 69.3%) are similar whereas form II of OZPN has a lower packing efficiency (packing coefficient = 67.9%) (Figure 7.12).

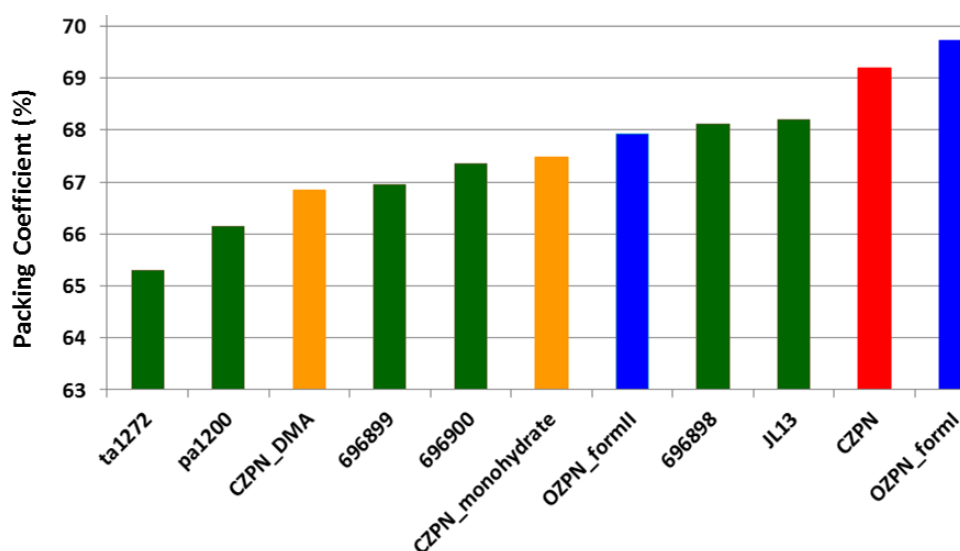


Figure 7.12. Packing coefficient (%) for the anhydrous form (red), solvates (orange), analogues (green) of clozapine and anhydrous polymorphs (blue) of olanzapine.

Inspection of the calculated packing efficiency for the known structures of CZPN shows that the stabilisation of the CZPN solvates by solvent molecules is not just a density packing effect similar to OZPN (Section 6.11), as the packing coefficient for the two CZPN solvates (monohydrate and DMA solvate) is not higher than the anhydrous form (Figure 7.12). Similarly packing efficiency of the other reported CZPN analogues with or without large substituents at position 4 of piperazine ring (Figure 7.13) (Dupont et al., 1996; Capuano et al., 2000; Dupont and Liegeois, 2003; Capuano et al., 2010), have relatively lower packing efficiency than CZPN (Figure 7.12). It is expected that large bulky substituents at N4 of piperazine ring would reduce the packing efficiency of the resulting crystal structure but the methyl substituent at N1 (pa1200) also reduces the packing efficiency considerably.

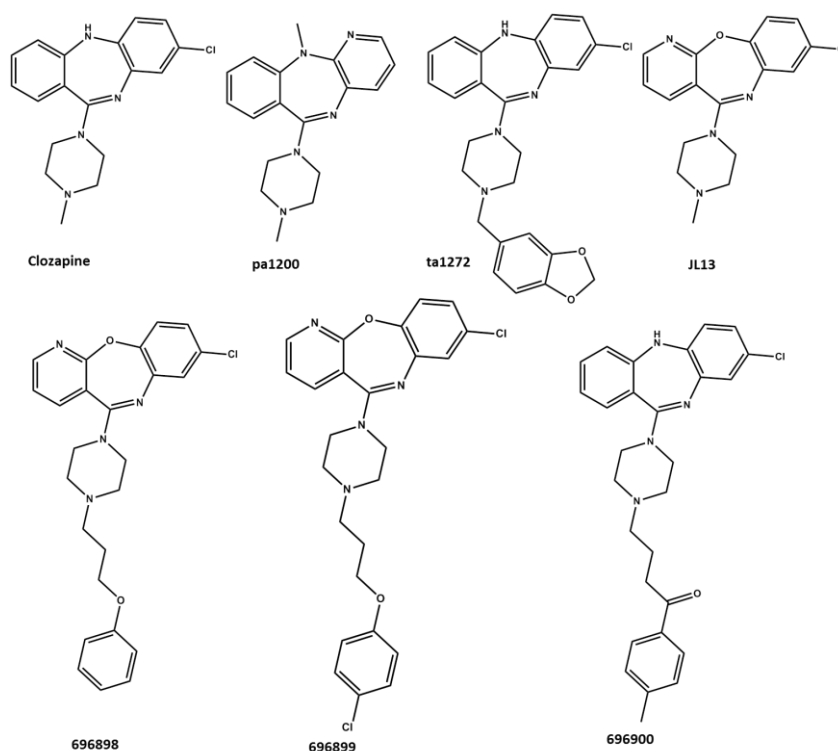


Figure 7.13. Chemical Structures of clozapine and its analogues.



### 7.3.1.2 Preliminary Crystal Structure Prediction of Clozapine

Encouragingly the experimental structure was obtained as the global minimum in lattice energy in the search (Figure 7.14). The next two lowest energy structures are very close to each other and are almost  $4.6 \text{ kJ mol}^{-1}$  and  $5.1 \text{ kJ mol}^{-1}$  higher in energy from global minimum. Another close cluster of crystal structures is observed in the region of  $-144.4$  to  $-141.1 \text{ kJ mol}^{-1}$  of lattice energy (Figure 7.14).

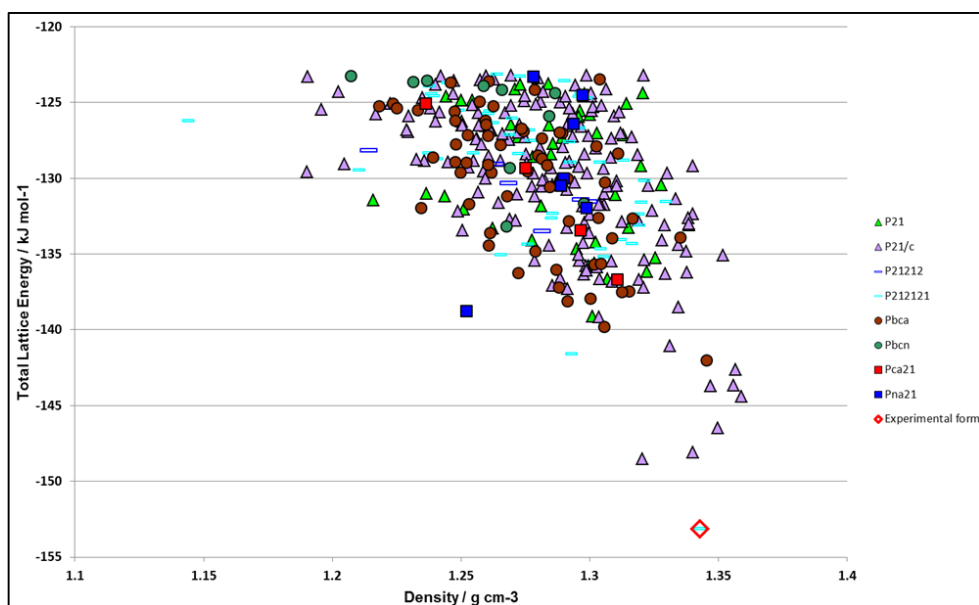


Figure 7.14. Summary of the preliminary crystal structure search of clozapine. Each point represents a hypothetical crystal structure/lattice energy minimum. The cyan rectangle with open red diamond represents the global minimum in the lattice energy, which corresponds to the known experimental structure.

In spite of being only a preliminary search, these results provide a valuable approximation of the crystal energy landscape. In contrast to CSP results of OZPN, these findings suggest that other polymorphic structures of CZPN are actually unlikely during standard crystallisation investigation due to the large

lattice energy gap between the global minimum and other predicted crystal structures.

Crystal structures with the centrosymmetric dimer were also commonly observed on the crystal energy landscape. The 2<sup>nd</sup> lowest energy structure contains the centrosymmetric dimer but is less densely packed than the global minimum. This suggests that in the two experimentally obtained CZPN solvates the solvent is stabilising the solvate structures by hydrogen bonding and other favourable interactions with the CZPN centrosymmetric dimers.

PIXEL calculations were further employed to develop a detailed picture of the interplay between molecular and crystal structures of OZPN and CZPN.

### ***7.3.1.3 PIXEL Calculations on Clozapine***

PIXEL calculations (see Section 3.2.5) were carried out on the CZPN anhydrous form, monohydrate and DMA solvate. Due to the disorder in the EG solvate crystal structure, PIXEL calculations could not be carried out. These structures (anhydrous form, monohydrate and DMA solvate) form an ideal set to study the relative contribution of the various packing motifs (centrosymmetric dimer vs. edge to face dimer vs. individual molecule) to the total lattice energy using PIXEL. The results of the crystal lattice calculations on these three structures are summarised in Table 7.5.

Table 7.5. Partitioned intermolecular interaction energies (kJ mol<sup>-1</sup>) from the PIXEL calculations on clozapine anhydrous form, monohydrate and DMA solvate.

Structure	E <sub>c</sub> <sup>a</sup>	E <sub>p</sub> <sup>b</sup>	E <sub>d</sub> <sup>c</sup>	E <sub>attrac</sub> <sup>d</sup>	E <sub>r</sub> <sup>e</sup>	E <sub>t</sub> <sup>f</sup>
CZPN	-66.9	-34.8	-201.2	-302.9	143.2	-159.8
% of E <sub>attrac</sub>	22.1	11.5	66.4	---	---	---
CZPN monohydrate	-91.3	-41.5	-113.3	-246.1	132.4	-113.8
% of E <sub>attrac</sub>	37.1	16.9	46.0	---	---	---
CZPN DMA	-50.3	-22.8	-121.4	-194.5	85.8	-108.8
% of E <sub>attrac</sub>	25.9	11.7	62.4	---	---	---

<sup>a</sup> Electrostatic (Coulombic) energy, <sup>b</sup> Polarisation term, <sup>c</sup> Dispersion term, <sup>d</sup> E<sub>attrac</sub> = E<sub>c</sub>+E<sub>p</sub>+E<sub>d</sub>, <sup>e</sup> Repulsion term, <sup>f</sup> Total intermolecular interaction energy, E<sub>t</sub> = E<sub>c</sub> + E<sub>p</sub> + E<sub>d</sub> + E<sub>r</sub>

In all three CZPN structures, the attractive forces are dominated by the E<sub>d</sub>. This is followed by the contribution from the E<sub>c</sub>, and then the E<sub>p</sub>. In the case of the monohydrate, the contribution of E<sub>c</sub> towards the total attractive energy (E<sub>attrac</sub>) is highest amongst the three structures. This might be due to the involvement of each CZPN molecule in three H-bonds with water molecules compared to two hydrogen bonds in the anhydrous form and one hydrogen bond in the DMA solvate.

### 7.3.1.3.1 *Pair-wise Molecular Interactions in Crystal Structure of Anhydrous*

#### *Clozapine*

The PIXEL calculations also identified several molecular pairs within the 2.5 kJ mol<sup>-1</sup> threshold of intermolecular interaction energy in all the three structures. Analysis of the crystal packing in terms of the individual pairwise molecular interactions for anhydrous form shows that the overall

intermolecular energy in each molecular pair is dominated by the contribution from the  $E_d$  (Table 7.6). Non-localised interactions between the diffuse electron clouds of CZPN molecules in dimers are most likely responsible for the higher dispersion energy contribution (Dunitz and Gavezzotti, 2005b).

Table 7.6. Intermolecular interaction energies ( $\text{kJ mol}^{-1}$ ) for the top three ranked molecular pairs in the anhydrous form of clozapine. These pairs are shown in Figure 7.15.

Rank <sup>a</sup>	$E_c$	$E_p$	$E_d$	$E_{\text{attrac}}$	$E_r$	$E_t$	COM ( $\text{\AA}$ ) <sup>b</sup>
1 <sup>st</sup>	-27.5	-12.3	-46.1	-85.9	40.4	-45.5	6.997
% of $E_{\text{attrac}}$	32.0	14.3	53.7	---	---	---	---
2 <sup>nd</sup>	-14.3	-8.1	-63.8	-86.2	45.5	-40.7	6.178
% of $E_{\text{attrac}}$	16.6	9.4	74.0	---	---	---	---
3 <sup>rd</sup>	-8.4	-3.8	-20.6	-32.8	17.4	-15.4	9.609
% of $E_{\text{attrac}}$	25.6	11.6	62.8	---	---	---	---

<sup>a</sup> Rank contribution of a pair of molecules to the overall intermolecular interaction energy in the structure studied, <sup>b</sup> Distance between the centres of mass of the two molecules ( $\text{\AA}$ ).

The topmost molecular pair-wise interaction involves CZPN molecules connected via NH...N hydrogen bond (Table 7.6 and Figure 7.15). The 2<sup>nd</sup> ranked molecular pairwise interaction includes contributions from  $\text{CH}_2\cdots\pi$  interactions between the CZPN molecules which form a side-ways dimer (Table 7.6 and Figure 7.15) and is over  $4.8 \text{ kJ mol}^{-1}$  less favourable than the top ranked pair-wise molecular interaction.

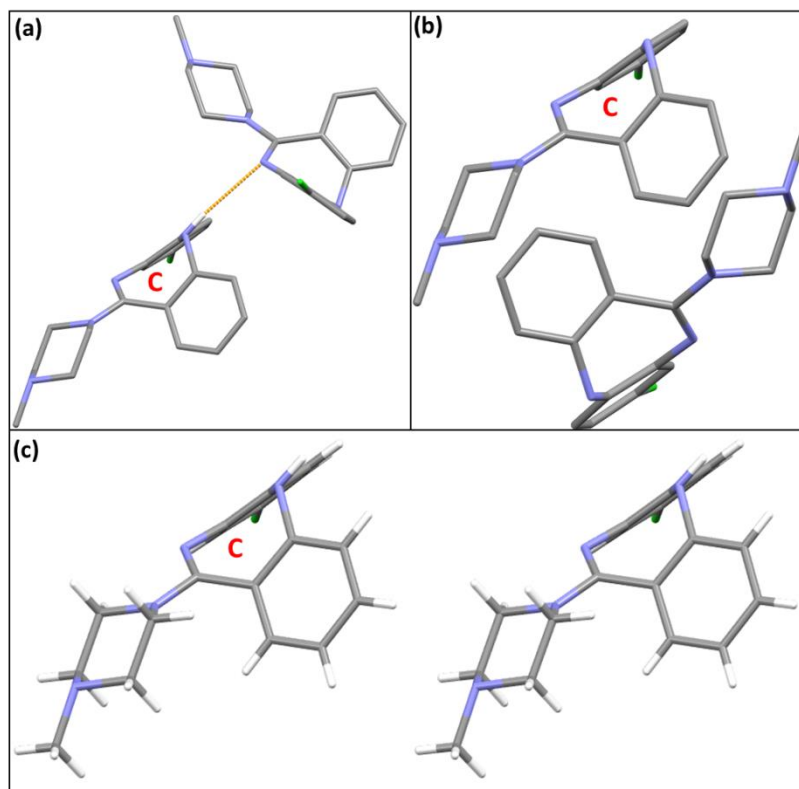


Figure 7.15. The top three molecular pairs present in the crystal lattice of clozapine identified in Table 7.6. (a) dimer involving hydrogen-bond between clozapine molecules (shown by orange line), (b) side-by-side dimer, and (c) dimer involving weaker molecule-molecule interactions. The central molecule is labelled as 'C'.

The top two molecular pairs are significantly more favourable ( $>25 \text{ kJ mol}^{-1}$ ) than the others within the coordination shell. The 3<sup>rd</sup> ranked molecular pair includes van der Waals interactions between the CZPN molecules (Figure 7.15) from adjacent dimers along the *b*-direction.

### 7.3.1.3.2 *Pair-wise Molecular Interactions in Crystal Structures of Clozapine Monohydrate and DMA Solvate*

Unlike the CZPN anhydrous form, the top most molecular pair in the DMA solvate and monohydrate involves non-hydrogen bonding CZPN molecules

(Table 7.7, Figure 7.16) and  $E_d$  is not the key contributor towards  $E_t$  in all the molecular pairs.

Table 7.7. Intermolecular interaction energies ( $\text{kJ mol}^{-1}$ ) for top three ranked dimer interactions in clozapine monohydrate and N,N-dimethylacetamide solvate. These pairs are shown in Figure 7.16.

	<b>Rank</b>	<b><math>E_c</math></b>	<b><math>E_p</math></b>	<b><math>E_d</math></b>	<b><math>E_{\text{attrac}}</math></b>	<b><math>E_r</math></b>	<b><math>E_t</math></b>	<b>COM (Å)</b>
DMA solvate	1 <sup>st</sup>	-18.0	-8.7	-68.4	-95.1	47.1	-48.0	5.861
	% of $E_{\text{attrac}}$	18.9	9.1	71.9	---	---	---	---
	2 <sup>nd</sup>	-40.9	-15.5	-18.5	-74.9	34	-41.0	6.651
	% of $E_{\text{attrac}}$	54.6	20.7	24.7	---	---	---	---
	3 <sup>rd</sup>	-7.2	-3.2	-25.2	-35.6	14.9	-20.8	5.461
	% of $E_{\text{attrac}}$	20.2	9.0	70.8	---	---	---	---
monohydrate	1 <sup>st</sup>	-28.8	-10.8	-94.9	-134.5	70.9	-63.6	4.446
	% of $E_{\text{attrac}}$	21.4	8.0	70.6	---	---	---	---
	2 <sup>nd</sup>	-46.3	-20.6	-18.8	-85.7	51.2	-34.6	4.003
	% of $E_{\text{attrac}}$	54.0	24.0	22.0	---	---	---	---
	3 <sup>rd</sup>	-49.7	-23.0	-15.5	-88.2	58.5	-29.7	6.733
	% of $E_{\text{attrac}}$	56.3	26.1	17.6	---	---	---	---

The most significant molecular pair in DMA solvate is a dispersion dominated dimer consisting of CZPN molecules from the adjacent chains and involves contribution from van der Waals interactions between CZPN molecules (Figure 7.16a). The 2<sup>nd</sup> ranked molecular pair shows an association mode based on intermolecular hydrogen bonds, namely N1-H...O1 (Figure 7.16) and  $E_{\text{attrac}}$  is dominated by the  $E_c$ . Similar to the anhydrous CZPN, the top two molecular pairs in the DMA solvate are significantly more favourable ( $>20 \text{ kJ mol}^{-1}$ ) than

the others within the coordination shell. The 3<sup>rd</sup> ranked molecular pair includes contributions from van der Waals interactions between CZPN and the DMA molecule and the  $E_d$  is the main contributor towards  $E_{\text{attrac}}$  (Figure 7.16 and Table 7.7).

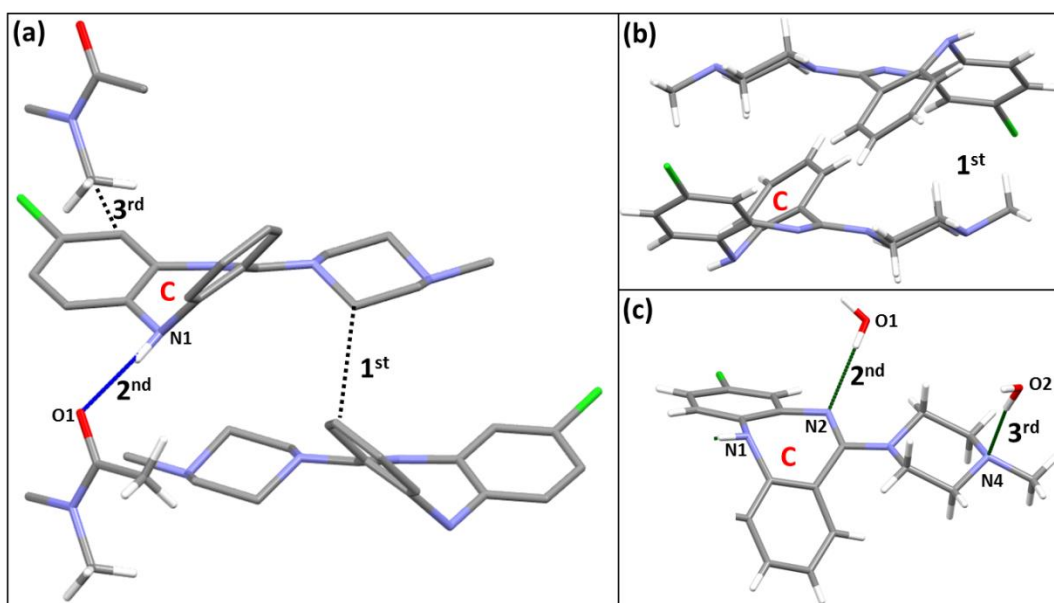


Figure 7.16. Top three pair-wise intermolecular interactions in (a) clozapine N,N-dimethylacetamide solvate and (b-c) clozapine monohydrate, quantified and labelled as in Table 7.7. The central molecule is labelled as 'C'. Other hydrogen atoms are omitted for clarity.

The topmost molecular pair in terms of interaction energy in CZPN monohydrate is a centrosymmetric dimer formed by CZPN molecules (Figure 7.16b) which is significantly more favourable than other molecular pairs within the coordination shell (Table 7.7). The 2<sup>nd</sup> and 3<sup>rd</sup> ranked molecular pairs involve contribution from H-bonds (O1-H...N2 and O2-H...N4 respectively) between CZPN and water molecules (Table 7.7 and Figure 7.16c). As expected from the H-bonded motif, the attractive forces are dominated by the  $E_c$ , which

represents over half of the total attractive forces in these two molecular pairs (Table 7.7 and Figure 7.16c). Dispersion forces in these two H-bonded pair-wise molecular interactions are not as significant as those of the anhydrous form of CZPN (~22% vs. ~54%) due to the fact that this dimer involves one CZPN with a water molecule as compared to the dimer of two CZPN molecules in the anhydrous form.

To summarise, PIXEL calculations have revealed that, independent of the packing motif,  $E_d$  is a key contributor towards stabilisation of the crystal lattices of anhydrous form and solvates of CZPN. In terms of pair-wise interactions, the relative contribution from the top most interaction is significantly less in non-centrosymmetric dimer based structures (CZPN and its DMA solvate) than in the centrosymmetric dimer based structure (CZPN\_monohydrate). Particularly, in CZPN solvates the contribution of  $E_c$  towards total energy is higher in H-bonded molecular pair as compared to anhydrous CZPN where H-bonded molecular pair is dominated by  $E_d$ .

On comparing both the anhydrous crystal structures of OZPN and CZPN, the total lattice energy is dominated by  $E_d$ . In contrast to OZPN, where the  $SC_0$  is the main contributor towards total energy, the lattice energy is more evenly distributed amongst the top two molecular pairs (H-bonding dimer and side-ways dimer) in anhydrous CZPN.

The second group of molecules (AXPN and LXPN) studied are discussed in the next section.



### 7.3.2 Physical Form Screening Results of Amoxapine

Across all the crystallisation experiments employed, the same physical form as the starting material, acetic acid salt and an amorphous phase of AXPN was obtained.

No new polymorphs, solvates or hydrates were observed from solution crystallisations from 47 solvents under three conditions (see Section 7.2.3) except acetic acid which yielded an acetate salt of AXPN. Clearly the solution experiments did not lead to alternate packing arrangements being produced. Neat and liquid-assisted grinding, vapour phase crystallisation and recrystallisation from amorphous phase returned the same non-solvated form as the starting material. Crystal structure of AXPN was redetermined at 123 K to allow comparison with CSP.

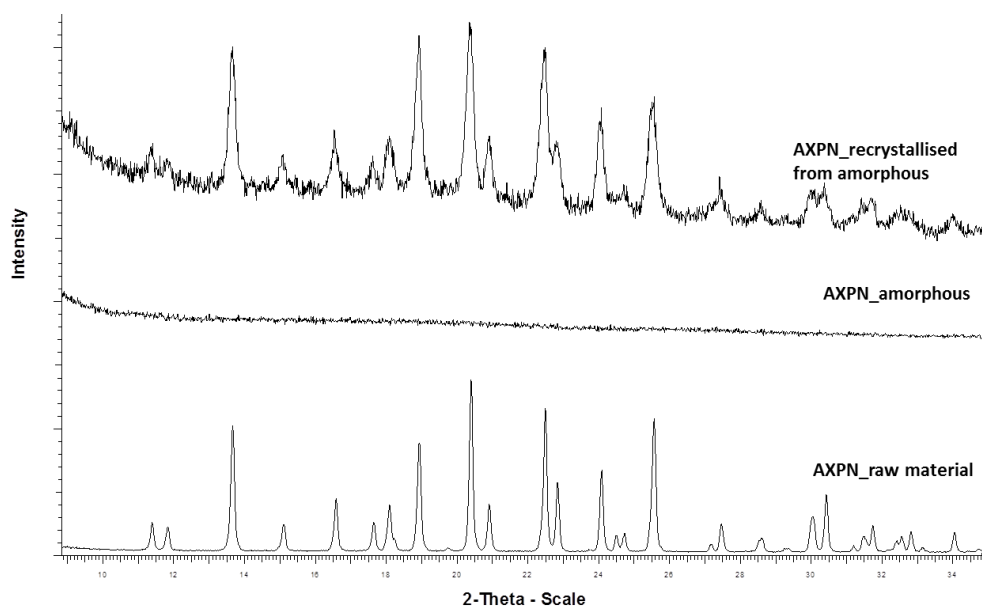


Figure 7.17. Stack plot of XRPD patterns of amoxapine raw material, melt quenched amorphous phase and partially recrystallised sample of amorphous amoxapine in the range of 10-35° 2 $\theta$ .

In DSC, AXPN melts at 453-454 K. *In-situ* recrystallisation of the amorphous sample and a thermal heat-cool-heat cycle (see Section 3.2.3) of the starting material of AXPN yielded the same starting material (endotherm at 454 K, Figure 7.18). Under these conditions, again no polymorphs other than known anhydrous form were observed.

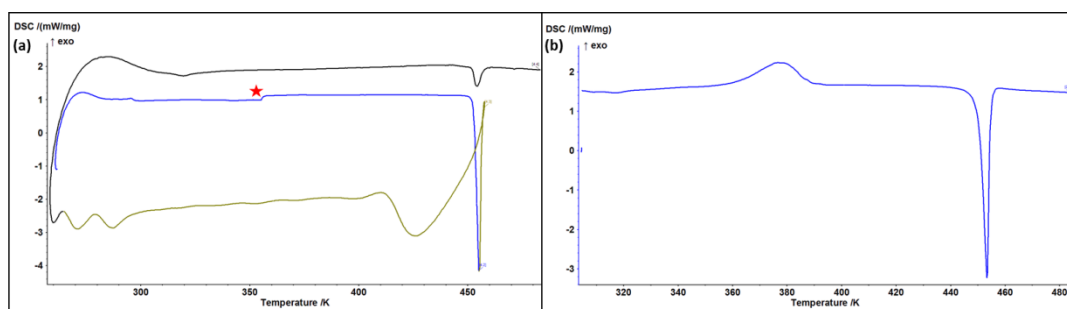


Figure 7.18. DSC curves of amoxapine (a) heat-cool-heat cycle of amoxapine. The heat-cool-heat segments are shown by blue, golden and black colour respectively. In heating curve, star denotes an instrument artefact. (b) amorphous amoxapine.

### 7.3.2.1 Molecular Packing Analysis of Crystal Structure of Amoxapine

AXPN molecules form two types of chains along *b*-axis and each chain consists of one enantiomer. These chains stack parallel to each other in a staggered manner to give rise to a layered structure (Figure 7.19a). The 2-D layers stack over each other along *a*-direction (Figure 7.19b) and are connected by C-H...O hydrogen bonds.

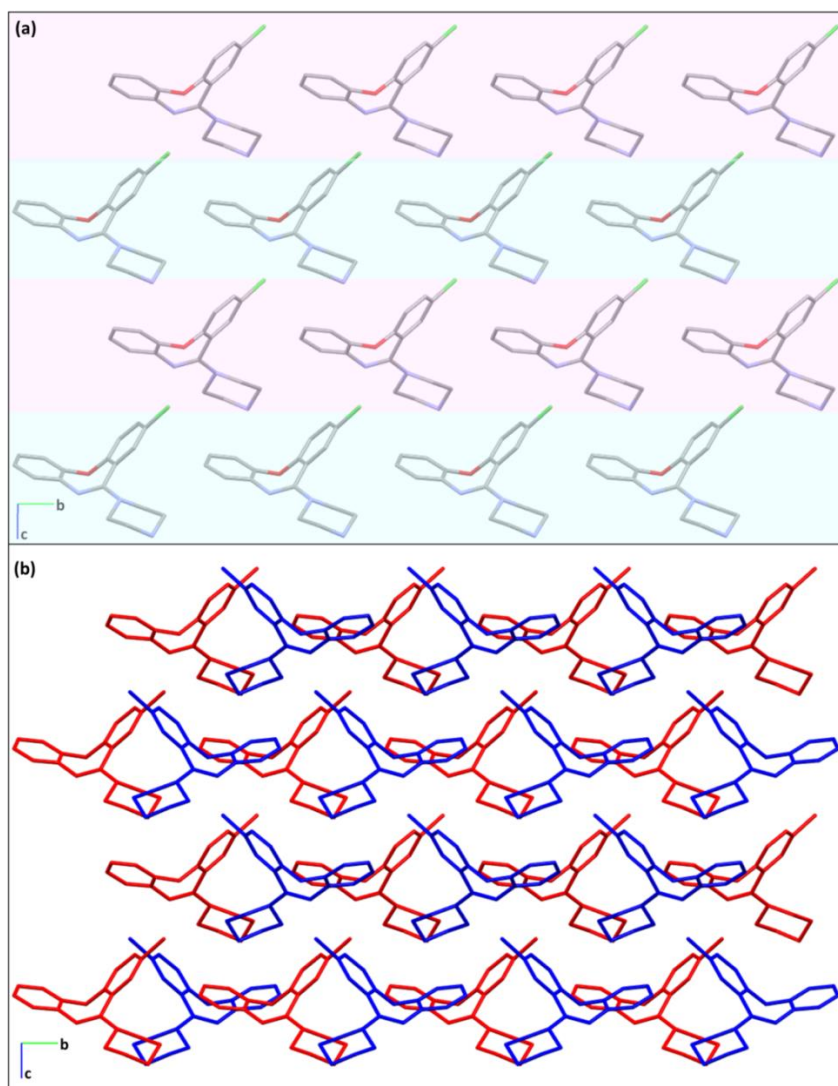


Figure 7.19. a) Layer structure formed by stacking of two types of chains (highlighted in pink and green colour) of amoxapine molecules. (b) The stacking of layers (shown by red and blue colours). All the hydrogen atoms are omitted for clarity.

Although AXPN molecule has one donor and three acceptors, it does not form any conventional intermolecular hydrogen bonds in the crystal lattice (Figure 7.19). AXPN crystal packing seems to be driven by space filling van der Waals interactions, stabilising the structure with reasonably high packing efficiency (packing coefficient = 70.5%).

### 7.3.2.2 Preliminary Crystal Structure Prediction of Amoxapine

Encouragingly, the experimental structure was found in the search as the global minimum in lattice energy (Figure 7.20). The next lowest energy structure is almost 5 kJ mol<sup>-1</sup> higher in energy, and the third lowest energy structure in the search has a gap of 6 kJ mol<sup>-1</sup> to the global minimum.

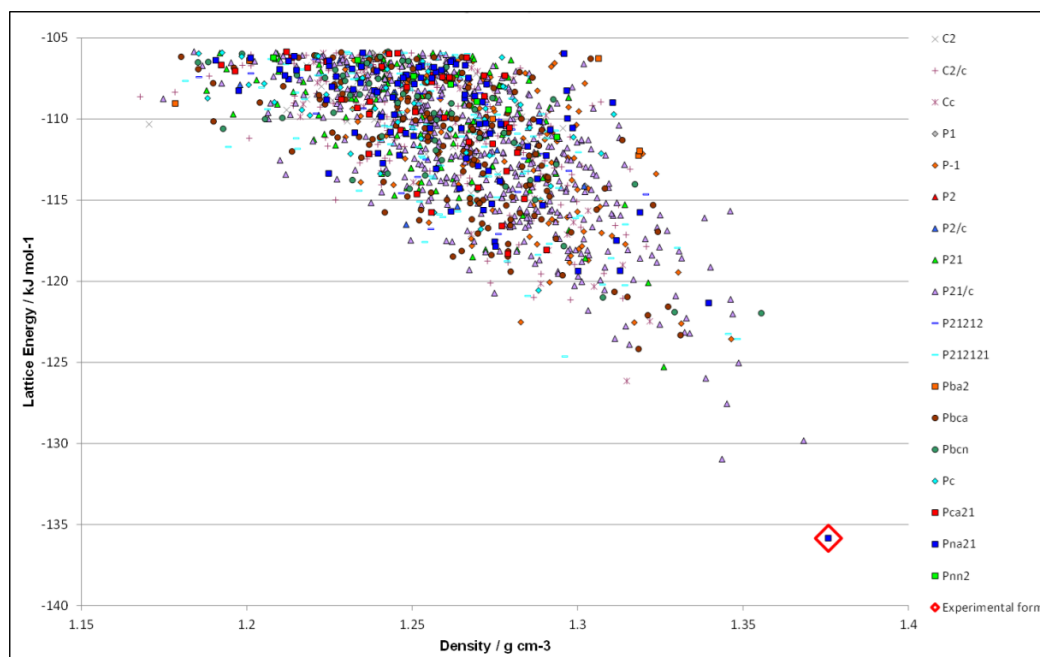


Figure 7.20. Summary of preliminary crystal structure search of amoxapine. Each point represents a hypothetical crystal structure/lattice energy minimum. The blue square with open red diamond represents the global minimum in lattice energy, which corresponds to the known experimental structure.

The approximate crystal energy landscape highlights that other polymorphic structures are unlikely to be obtained during routine crystallisation investigation because of the large gap between global minimum and other predicted crystal structures. In other words the known form is the most favourable packing arrangement by far and other structures are not thermodynamically competitive.

This type of landscape provides an answer to the experimentalist question “When should we stop the experimental search?” (Braun et al., 2010). It is notable, however, that this is the first time that this type of landscape has been observed for pharmaceutical molecules in this group’s experience.

### 7.3.2.3 *PIXEL Calculations on Amoxapine*

PIXEL calculations show that the majority of the contribution towards crystal stabilisation comes from  $E_d$  followed by  $E_c$  and  $E_p$  terms (Table 7.8).

Table 7.8. Partitioned intermolecular interaction energies ( $\text{kJ mol}^{-1}$ ) from the PIXEL crystal lattice calculations of amoxapine.

<b>Structure</b>	<b><math>E_c</math></b>	<b><math>E_p</math></b>	<b><math>E_d</math></b>	<b><math>E_{\text{attrac}}</math></b>	<b><math>E_r</math></b>	<b><math>E_t</math></b>
form I	-50.2	-23.9	-192.1	-274.0	123.3	-142.9
% of $E_{\text{attrac}}$	18.8%	9.0%	72.2%	---	---	---

Eight pair-wise interactions within a threshold of  $2.5 \text{ kJ mol}^{-1}$  were identified from the crystal lattice calculations. Analysis of the individual pairwise molecular interactions (Table 7.9 and Figure 7.21) revealed that the overall intermolecular energy in each molecular pair is dominated by the contribution from the  $E_d$  term.

The first three top ranked molecular pairs have been found to have similar intermolecular interaction energies and over 69% of the contribution towards attractive forces comes from the  $E_d$  followed by  $E_c$  and  $E_p$  (Table 7.9).

Table 7.9. Intermolecular interaction energies ( $\text{kJ mol}^{-1}$ ) for the top three ranked molecular pairs in amoxapine. These pairs are shown in Figure 7.21.

Rank	$E_c$	$E_p$	$E_d$	$E_{\text{attrac}}$	$E_r$	$E_t$	COM ( $\text{\AA}$ )
1 <sup>st</sup>	-9.7	-4.6	-33.3	-47.6	21.1	-26.5	7.300
% of $E_{\text{attrac}}$	20.4	9.7	69.9	---	---	---	---
2 <sup>nd</sup>	-7.6	-3.3	-31.8	-42.7	16.6	-26.0	7.903
% of $E_{\text{attrac}}$	17.8	7.7	74.5	---	---	---	---
3 <sup>rd</sup>	-6.2	-4.6	-37.2	-48.0	23.1	-24.9	7.410
% of $E_{\text{attrac}}$	12.9	9.6	77.5	---	---	---	---

1<sup>st</sup>-3<sup>rd</sup> ranked molecular pairs involve AXPN molecules from adjacent layers. The top ranked molecular pair involves contributions from C-H...Cl and other van der Waals interactions between piperazine rings (Figure 7.21a). The 2<sup>nd</sup> ranked pair involves contributions from the C-H...O bond and van der Waals interactions between piperazine and phenyl rings (Figure 7.21b) with CH... $\pi$  interactions between the piperazine and chlorophenyl rings contributing to the 3<sup>rd</sup> ranked pair (Figure 7.21c).

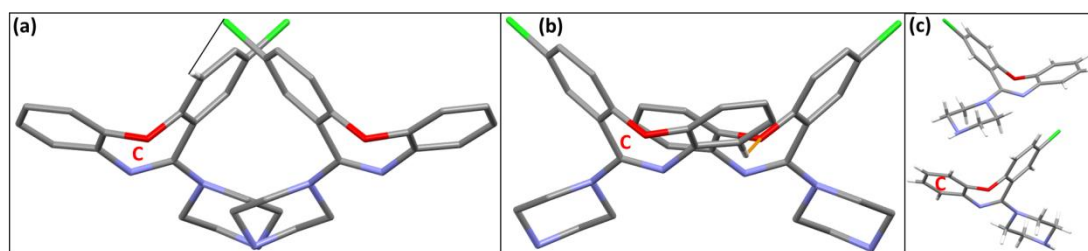


Figure 7.21. Top five pairwise intermolecular interactions in amoxapine, quantified in Table 7.9, (a-b) molecular pairs involve contribution from C-H...Cl interaction (black line) and C-H...O interaction (orange line) and (c) a molecule-molecule interaction. The central molecule is labelled as 'C'.

PIXEL calculations indicated that in the absence of strong intermolecular hydrogen bonds, the van der Waals interactions play a significant role in stabilising the crystal lattice of AXPN. These calculations also revealed that a large part of the cohesive energy in molecular pairs containing weak hydrogen bonds (C-H...O present in 2<sup>nd</sup> ranked molecular pair and C-H...N present in 4<sup>th</sup> ranked molecular pair and is not shown here) also arises from the dispersion term which is difficult to characterise and particularly challenging for computational CSP methods (Dunitz and Gavezzotti, 2005b; Bhardwaj et al., 2013).

The calculated values for intermolecular interactions cannot however explain why the observed associated patterns dominate over other potential molecular pairs that could be formed involving strong H-bonds between N-H groups and O atoms of AXPN molecules during molecular self-assembly. The absence of strong H-bond could be due to the steric effect associated with the benzene rings next to ether moiety which may hinder the access of available oxygen atom for the formation of H-bonds in the crystal lattice. CSD (Frank, 2002) (version 5.34) analysis revealed that only 3% of the structures containing ether motif were found to be involved in conventional H-bonding. This also suggests that oxygen atom in this kind of ether motif is not easily accessible for H-bonding.

### 7.3.3 Physical Form Screening Results of Loxapine

#### 7.3.3.1 Purity of Supplied Material of Loxapine

LXPN was purchased from Molekula, UK. The XRPD data of LXPN did not match the unit cell of either of the two reported polymorphs (Petcher and Weber, 1976; Cosulich and Lovell, 1977). STA analysis revealed a 1% mass loss on heating at ~373 K and the melting point of the supplied material (428 K) was much higher than the two known LXPN polymorphs (386 K and 388 K). Recrystallisation of the supplied material from 2-butanone yielded a single crystal suitable for structure determination from SXD. The crystallography data confirmed the identity of the sample as monohydrate of the succinate salt of LXPN. XRPD and STA analysis of a recrystallised sample showed that the supplied material was a mixture of hydrated and anhydrous LXPN succinate salts. The LXPN hydrogen succinate monohydrate crystal structure is known although due to disorder the H-atoms positions were not reported (Fillers and Hawkinson, 1982a). Hence, redetermination of the crystal structure in this investigation has enabled a complete description of the structure including disorder with all the H-atoms positions to enable a better understanding of the intermolecular packing.



### ***7.3.3.2 Crystal Structure Determination of Anhydrous Loxapine Succinate using XRPD***

The supplied material was oven dried for 5 hours at 373 K to dehydrate the sample. The crystal structure of the anhydrous material was subsequently determined using XRPD. Indexing using DICVOL04 (Boultif and Louer, 1991) was unsuccessful, so singular valued decomposition (SVD)-Index algorithm (Coelho, 2003a) in TOPAS was used to index the powder pattern. A monoclinic solution with unit cell parameters of  $a = 36.182 \text{ \AA}$ ,  $b = 7.083 \text{ \AA}$ ,  $c = 18.759 \text{ \AA}$ ,  $\beta = 66.171^\circ$  and volume =  $4397.54 \text{ \AA}^3$  was obtained with suggested space group  $C2/c$  indicated by consideration of volume and systematic absences. A Pawley type fit of this unit cell and space group against the powder data ( $4\text{-}37^\circ 2\theta$ ) in DASH returned a Pawley  $\chi^2 = 3.74$ .

It is generally observed that stable salt formation occurs when the difference in pKa between free acid and base counterions is more than 3 (Bastin et al., 2000). The difference in pKa between LXPN and succinic acid is 4.0 (calculated using the Pipeline Pilot Professional Client interface (Accelrys, 2010)), which indicates salt formation is likely. Therefore, this structure was assumed to be the succinate salt and structure solution using SA was attempted as a salt configuration. Geometry optimisation using CASTEP was then applied to the structure obtained from structure solution to confirm the positions of all H-atoms and to verify the salt configuration of the structure (Florence et al., 2009). SA runs were used to optimise the LXPN hydrogen succinate model against the diffraction data set (166 reflections) using the direct space method. An internal

coordinate description of protonated LXPB was derived from the single crystal of the hydrated LXPB hydrogen succinate structure (determined in this work) and for succinate anion it was derived from the crystal structure of the succinate salt available in the CSD (refcode: BOVWIT). The O–H and C–H distances were normalised to 0.90 Å to 0.95 Å respectively. The structure was solved using 500 SA runs of  $2.5 \times 10^7$  moves per run as implemented in DASH in which the position, orientation and conformation of internal coordinate description (z-matrix) of a protonated LXPB and a succinate anion were varied. Each protonated LXPB and succinate anion molecule was allowed 7 (6 external and 1 internal) and 9 (6 external and 3 internal) degrees of freedom respectively. Structures returned from SA with ratios of Profile  $\chi^2$ /Pawley  $\chi^2$  less than 10 are generally considered to be solved (Florence et al., 2005). 57% of the runs successfully located the global minimum (Profile  $\chi^2$  to Pawley  $\chi^2$  ratio threshold <10).

CASTEP was then used to confirm that the diffraction based minimum structure also matches with an energy minimum structure, thus verifying the accuracy of salt configuration of the structure. The SA solution with lowest ratio of Profile  $\chi^2$  to Pawley  $\chi^2$  was taken for geometry optimisation in CASTEP. Although the pKa difference between LXPB and succinic acid suggests that the resulting structure is most likely a salt, to verify this, the salt configuration was changed to a co-crystal (by manually moving the proton from protonated LXPB to the succinate anion) and a geometry optimisation carried out using CASTEP to confirm its position. After geometry optimisation of the LXPB hydrogen succinate salt structural model, no movement in the proton position (H1) was observed.

However, in the structural model corresponding to a co-crystal, the proton from the carboxylic acid of succinic acid (H1 (SA) in Figure 7.22) was transferred to the nitrogen atom of LXPN molecule, (H1 (DFT) in Figure 7.22) confirming the salt configuration as accurate. No significant movement was observed for any of the heavier atoms e.g. C, N, O and Cl (Figure 7.22). These results confirm that the structure obtained from SA is accurate and is a salt.

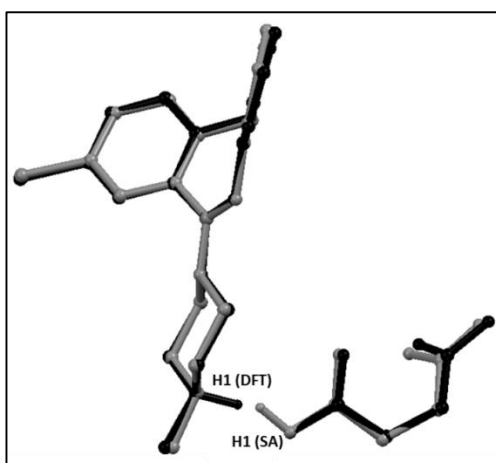


Figure 7.22. Overlay of the best simulated annealing solution with proton manually placed at succinic acid (co-crystal, grey) with the CASTEP geometry optimised structure (black) indicating proton transfer from succinic acid to loxapine. H-atom position before and the one obtained after geometry optimisation is denoted by H1 (SA) and H1 (DFT) respectively. Other hydrogen atoms are omitted for clarity. RMSD non-hydrogen atoms = 0.121Å.

After SA, the global minimum structure with a Profile  $\chi^2$ /Pawley  $\chi^2$  ratio (8.73/3.74) was used for subsequent restrained Rietveld refinement in TOPAS Academic (version 4.1). All atomic positions (including H atoms) for this structure were refined, subject to a series of restraints on bond lengths, bond angles and planarity. The final refinement included a total of 188 parameters (17 profile, 4 cell, 1 scale, 1 isotropic temperature factor and 165 positions) yielding a final  $R_{wp}$  of 3.164 and goodness of fit of 2.119 (Figure 7.23).

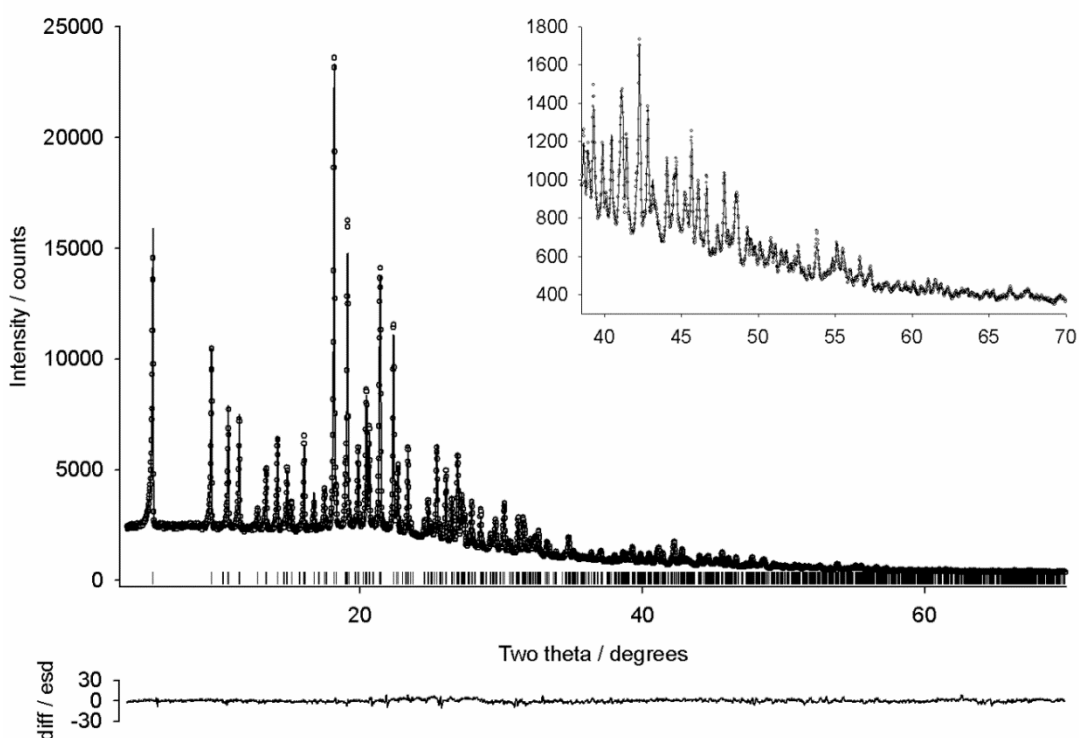


Figure 7.23. Final fit showing the observed  $y_{\text{obs}}$  (lines), calculated  $y_{\text{calc}}$  (points) and difference  $[(y_{\text{obs}} - y_{\text{calc}})/\sigma(y_{\text{obs}})]$  profiles for the restrained Rietveld refinement of the best simulated annealing solution of loxapine hydrogen succinate.

The resulting structure from Rietveld refinement was further scrutinised by allowing all fractional coordinates to refine freely (181 parameters,  $R_{\text{wp}} = 2.66$ ). As expected, the improved  $R_{\text{wp}}$  came at the expense of some chemical sense (e.g. slight distortion in planarity of benzene ring, movement of H-atoms to nonsensical positions), but otherwise, the geometry of the independent molecules was well preserved, confirming the correctness of the restrained refined crystal structure (Florence et al., 2008).

No further work was carried with LXPN succinate salt as it was beyond the scope of this thesis.

### **7.3.3.3 Physical Form Screening Results on Loxapine Free Base**

The supplied material was form I of LXPB as indicated by Pawley type refinement of XRPD data. Across all the crystallisation conditions, 2 known polymorphs and 4 novel physical forms (solvates and/or salts) of LXPB were identified. In comparison to monomorphic AXPB, LXPB exhibit polymorphism and has a tendency to form solvates.

Of the 96 solution crystallisations for LXPB (Table 7.3), 10 (10.4%) yielded no sample or insufficient solid for XRPD analysis and 3 (3.1%) provided powder patterns with no diffraction peaks. Of the 83 samples which yielded measurable diffraction, form I and II were identified in 49 (51.0%) and 22 (22.9%) samples respectively. 5 samples were identified as a mixture of forms I and II. The remaining 7 samples were recognised as novel forms (solvates and/or salts) of LXPB.

The outcome of each crystallisation technique is represented by a series of scatter plots (Figure 7.24 and Figure 7.25). Each point in the plot is positioned according to the value of  $t[1]$  and  $t[2]$  from the solvent property PCA (shown in Figure 7.2) and coloured according to the crystallisation outcome. Form I was found in majority of the cases of solvent evaporation from vials at RT and cooling crystallisation whilst in the case of faster solvent evaporation (solvent evaporation on a watch glass at RT), form II was observed in the majority of samples. A detailed listing of the form(s) identified from each recrystallised sample is provided in the Appendix (Table Apx 1.16- Table Apx 1.18).

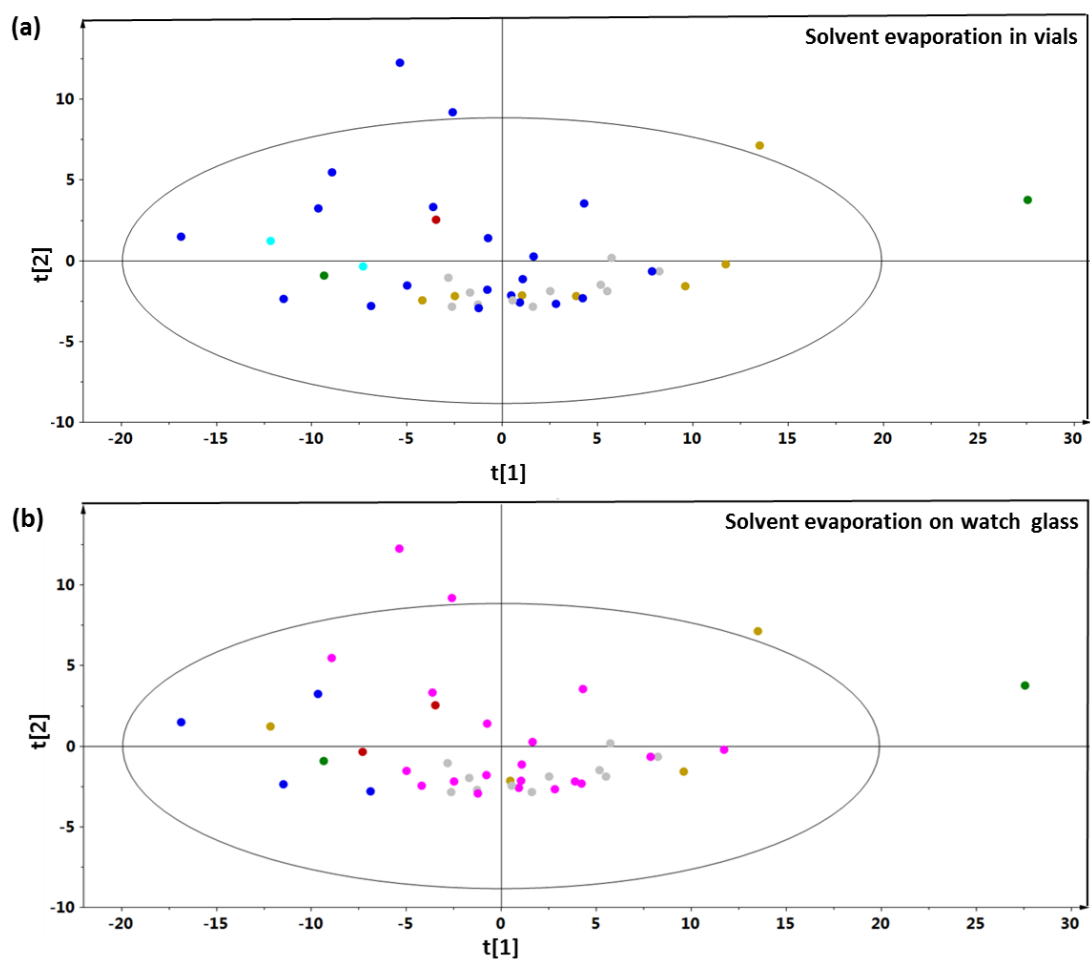


Figure 7.24. Scatter plot of solvents with positions according to the value of  $t[1]$  and  $t[2]$  from Figure 7.2. Solvents are coloured according to the crystallisation outcome: form I (blue), form II (magenta), mixture of forms I and II (green), novel physical forms, solvates and/or salts (red), no sample obtained (golden) and no spectral information (cyan). Solvents which were not used in crystallisation are coloured grey. Plots are shown for crystallisation conditions listed in Table 7.3 (a) solvent evaporation in vials and (b) solvent evaporation on watch glass.

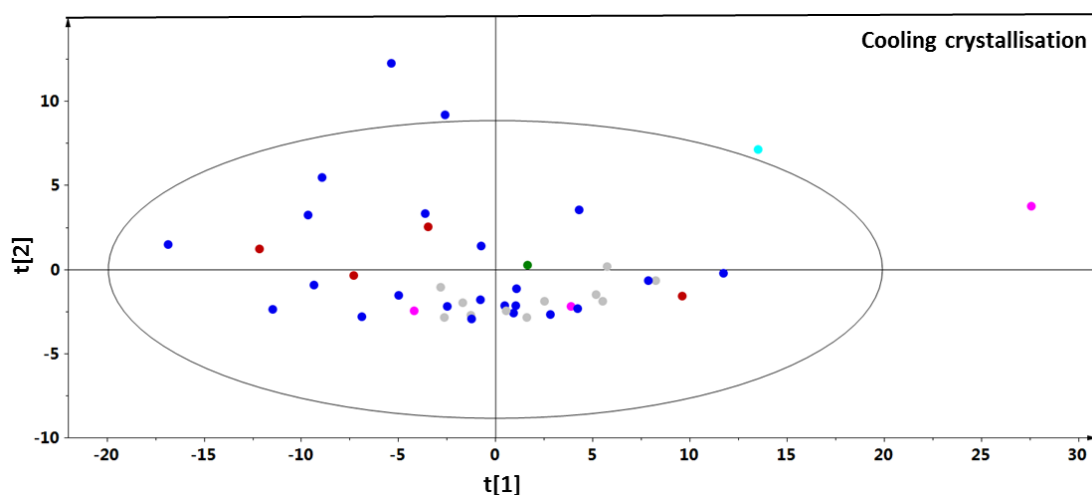


Figure 7.25. Scatter plot of solvents with positions according to the value of  $t[1]$  and  $t[2]$  from Figure 7.2. Solvents are coloured according to the crystallisation outcome: form I (blue), form II (magenta), mixture of forms I and II (green), novel forms, solvates and/or salts (red) and no spectral information (cyan). Solvents which were not used in crystallisation are coloured grey. Plot is shown for cooling crystallisations, listed in Table 7.3.

Based on XRPD data and thermal analysis, the physical forms comprise of two known polymorphs (form I and II), 2 solvates (2-butoxyethanol, 2,2,2-trifluoroethanol) and 2 solvates/salts with acetic acid and formic acid. The XRPD patterns of physical forms observed during screening are shown in Figure 7.26. The lattice parameters of form I and II (redeterminations) are given in Table 7.10. No further characterisation of the novel forms was carried out. Detailed structural analysis of only form I and II is described here.

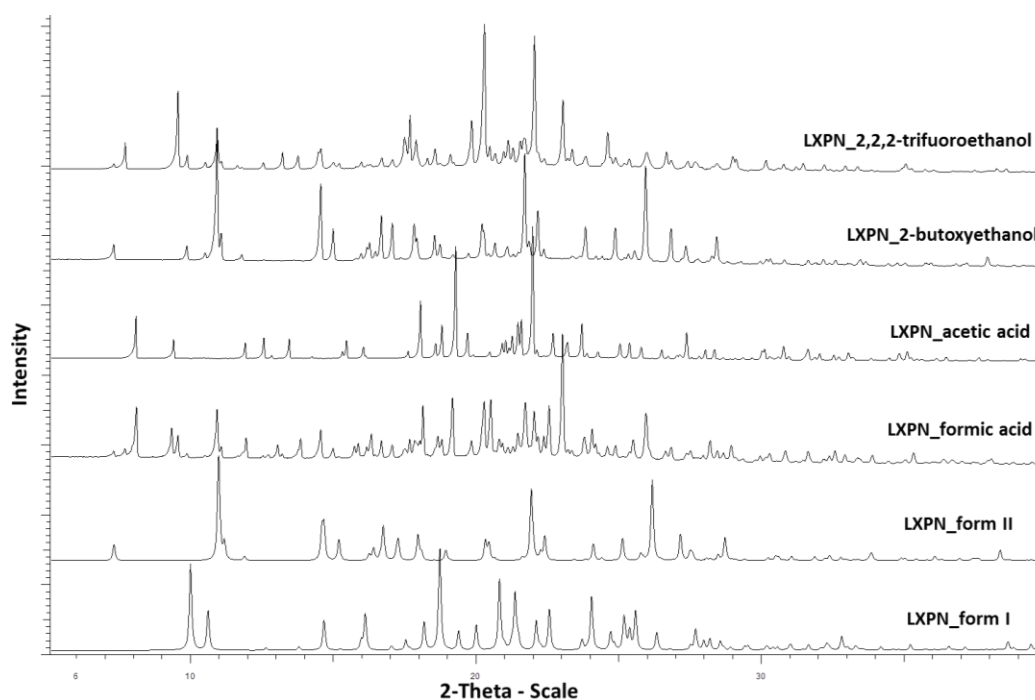


Figure 7.26. Stack plot of XRPD patterns of physical forms of loxapine identified during experimental screen.

Table 7.10. Lattice parameters for loxapine polymorph crystal structures (redeterminations) found in the experimental search (standard uncertainties in parentheses). All parameters were determined from single-crystal diffraction collected at 123(2) K.

Structure	Space Group	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\alpha$ (°)	$\beta$ (°)	$\gamma$ (°)	Vol (Å <sup>3</sup> )
Form I	<i>Pbca</i>	14.0245(5)	12.8546(5)	17.7500(7)	90.00	90.00	90.00	3200.0(2)
Form II	<i>P2<sub>1</sub>/c</i>	12.8867(7)	10.8108(6)	12.3806(6)	90.00	109.484(2)	90.00	1626.0(1)

STA analysis revealed that LXP\_N form I melts at 384-385 K and form II melts at 381-382 K (Figure 7.27). During these analyses no signs of any other polymorphs were observed.



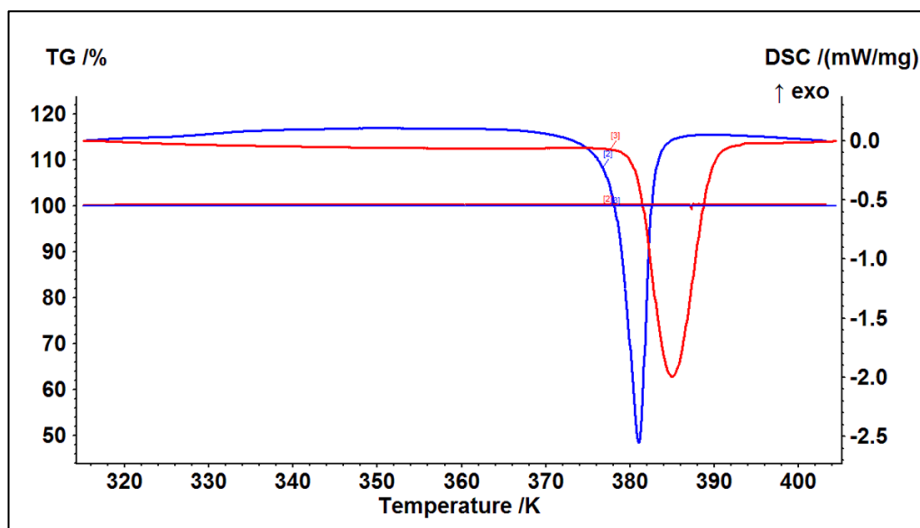


Figure 7.27. STA thermograms of loxapine form I (red) and form II (blue).

Molecular packing analysis was further carried out for form I and II for LXPN to compare it with AXPN.

#### 7.3.3.4 Molecular Packing Analysis of Form I and II of Loxapine Free Base

Both LXPN (form I and II) and AXPN exhibit similar conformation in the crystal structures (Figure 7.28). It highlights that the methyl group in the LXPN is not affecting its molecular conformation in the crystal lattices.

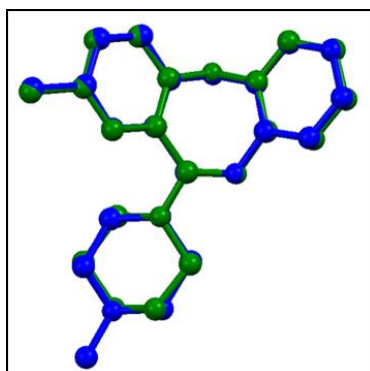


Figure 7.28. Overlay of the molecular conformations of amoxapine (green) and loxapine (blue) which are present in the experimental crystal structures (RMSD of non-hydrogen atoms = 0.08Å).

In form I, LXPN molecules form non-H bonded centrosymmetric dimers. These dimers arrange themselves in a corrugated manner giving rise to a 2-D sheet structure (Figure 7.29a). In the sheet structure, LXPN dimers are connected via weak C-H...O H-bonds. Alternate sheets stack over each other in a staggered arrangement, giving rise to the 3-D structure (Figure 7.29b).

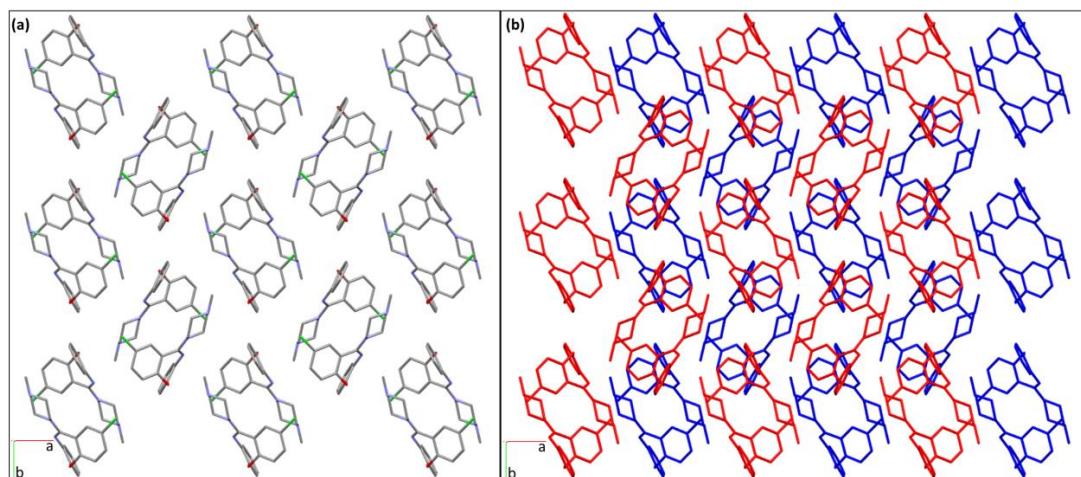


Figure 7.29. (a) 2-D packing of loxapine. (b) The staggered arrangement of sheets. Alternate sheets are shown by red and blue colour respectively. All the hydrogen atoms are omitted for clarity.

In form II, LXPN molecules form non-H-bonded antiparallel chains (each comprises one enantiomer) along *b*-direction (Figure 7.30a-b). Alternate chains of enantiomers stack along the *c*-direction to form sheets that stack along *a*-direction (Figure 7.30c).

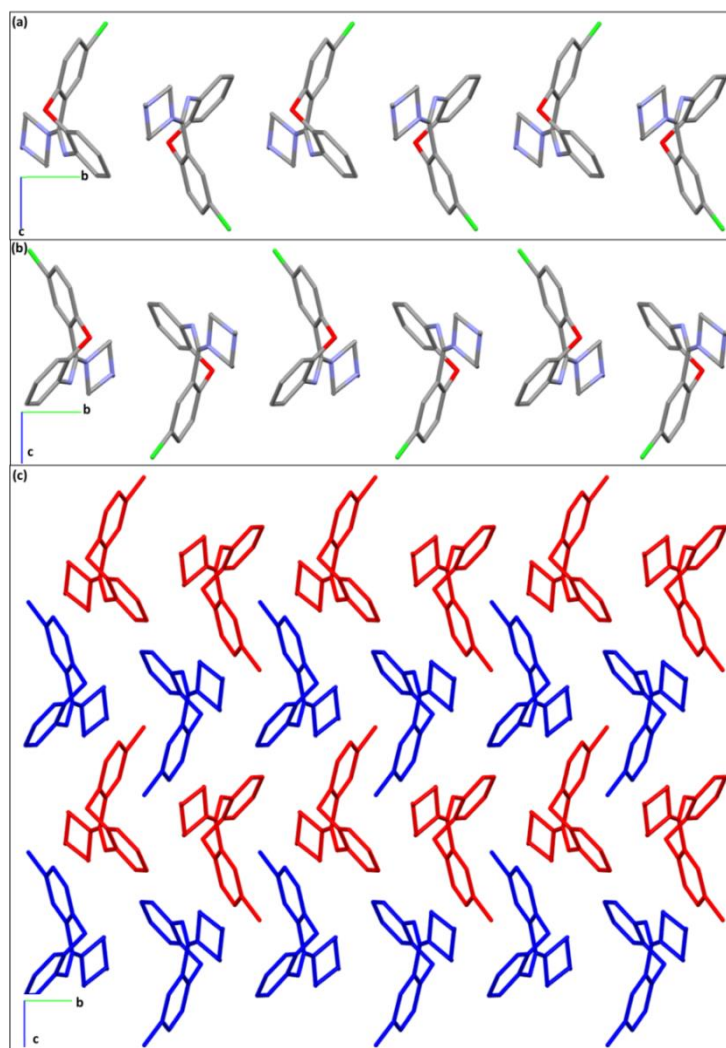


Figure 7.30. (a–b) 1-D chains of loxapine molecules along *b*-axis. (c) The stacking of chains of loxapine molecules along *c*-direction. All the hydrogen atoms are omitted for clarity.

#### 7.3.3.4.1 Comparison of Crystal Structures of AXPN and LXPN

In spite of having similar molecular conformation in the crystal lattice, the crystal structures of both the compounds do not have a common packing motif. Similar to AXPN, no conventional intermolecular H-bonds were observed in the crystal lattice of either polymorph of LXPN.

AXPN has comparatively higher packing coefficient (70.5%) than both polymorphs of LXPN (packing coefficient for form I and II are 68.7% and 67.4% respectively). This suggests that the methyl substituent leads to less efficient packing of crystal structures of LXPN as compared to AXP. Efficient packing in crystal structure of AXP might be one of the potential reasons behind lack of alternate packing arrangements. Whilst in case of LXP, the alternate packing motifs yield crystal structures with similar packing efficiencies.

#### ***7.3.3.5 Preliminary Crystal Structure Prediction of Loxapine Free Base***

Two different structures were observed as the lowest energy structures in the crystal energy landscape (Figure 7.31). One of these matches with the experimentally observed form II. The second, equi-energetic structure bears no resemblance with the experimental observed forms. The other experimental structure form I was not observed during the search.

In contrast to the crystal energy landscape for AXP, Figure 7.31 suggests that there are various available options for packing of LXP molecule that can lead to many thermodynamically competitive crystal structures. The contrasting features of the crystal energy landscapes of LXP and AXP therefore explain the differences in their observed solid-state diversity.

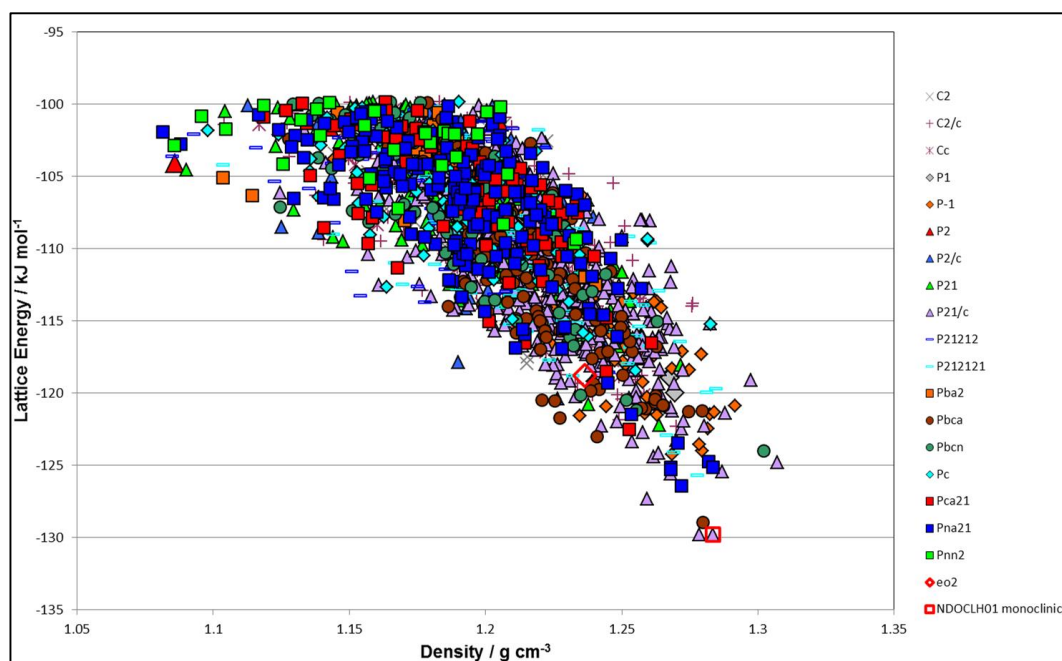


Figure 7.31. Summary of preliminary crystal structure search of loxapine. Each point represents a hypothetical crystal structure. The violet triangle with open red square represents one of the lowest energy structures, which corresponds to the known experimental structure.

PIXEL calculations were undertaken to offer a closer examination of the packing of LXPN in the two experimental crystal structures to examine any comparison with AXPN.

### 7.3.3.6 PIXEL Calculations on Form I and II of Loxapine Free Base

As expected for non/weak H-bonded systems, the attractive forces are dominated by  $E_d$  in both form I and II of LXPN (Table 7.11). The next largest contributions are from  $E_c$  and  $E_p$ , respectively. This trend of relative contributions of the different energy terms ( $E_d$ ,  $E_c$  and  $E_p$ ) towards the total lattice energy is the same as observed with AXPN.

Table 7.11. Partitioned intermolecular interaction energies (kJ mol<sup>-1</sup>) from the PIXEL calculations of loxapine polymorphs.

Structure	E <sub>c</sub>	E <sub>p</sub>	E <sub>d</sub>	E <sub>attrac</sub>	E <sub>r</sub>	E <sub>t</sub>	Density g cc <sup>-1</sup>
Form I	-46.5	-23.1	-188.9	-258.5	121.5	-136.8	1.361
% E <sub>attrac</sub>	18.0%	8.9%	73.1%	---	---	---	
Form II	-36.9	-19.8	-178.6	-235.3	110.7	-124.6	1.339
% E <sub>attrac</sub>	15.7%	8.4%	75.9%	---	---	---	

Similar to AXPN, analysis of the crystal packing in terms of the individual pairwise molecular interactions shows that the individual contribution from different energy terms in each pairwise molecular interaction follows the same trend as observed with the total intermolecular energy (E<sub>d</sub>>E<sub>c</sub>>E<sub>p</sub>) (Table 7.12).

Table 7.12. Intermolecular interaction energies (kJ mol<sup>-1</sup>) for the top three ranked dimer interactions in loxapine polymorphs. These pairs are shown in Figure 7.32 and Figure 7.33.

Structure	Rank <sup>a</sup>	E <sub>c</sub>	E <sub>p</sub>	E <sub>d</sub>	E <sub>attrac</sub>	E <sub>r</sub>	E <sub>t</sub>	COM (Å) <sup>b</sup>
Form I	1 <sup>st</sup>	-31.5	-14.8	-106.1	-152.4	82.8	-69.7	4.01
	% E <sub>attrac</sub>	20.7%	9.7%	69.6%	---	---	---	---
	2 <sup>nd</sup>	-4.3	-3.9	-32.6	-40.8	15.4	-25.4	7.688
	% E <sub>attrac</sub>	10.5%	9.6%	79.9%	---	---	---	---
	3 <sup>rd</sup>	-10.4	-2.6	-16.5	-29.5	8.2	-21.2	9.356
	% E <sub>attrac</sub>	35.3%	8.8%	55.9%	---	---	---	---
Form II	1 <sup>st</sup>	-11.8	-7.3	-59.0	-78.1	41.9	-36.1	6.476
	% E <sub>attrac</sub>	15.1%	9.3%	75.6%	---	---	---	---
	2 <sup>nd</sup>	-10.8	-5.5	-45.9	-62.2	30.3	-31.8	6.849
	% E <sub>attrac</sub>	17.4%	8.8%	73.8%	---	---	---	---
	3 <sup>rd</sup>	-11.5	-5.1	-29.1	-45.7	20.5	-25.2	8.963
	% E <sub>attrac</sub>	25.2%	11.1%	63.7%	---	---	---	---

The most favourable interaction in form I is between the molecules forming the centrosymmetric dimer (Table 7.12, Figure 7.32), which is significantly higher than the other interactions within the coordination sphere. The 2<sup>nd</sup> and 3<sup>rd</sup> ranked molecular pairs involve contributions from van der Waals contacts between aromatic rings of LXPN molecules from adjacent centrosymmetric dimers and C-H...O interactions, respectively. The molecule-molecule interaction containing centrosymmetric dimer is significantly more favourable than the other interactions within the coordination shell.

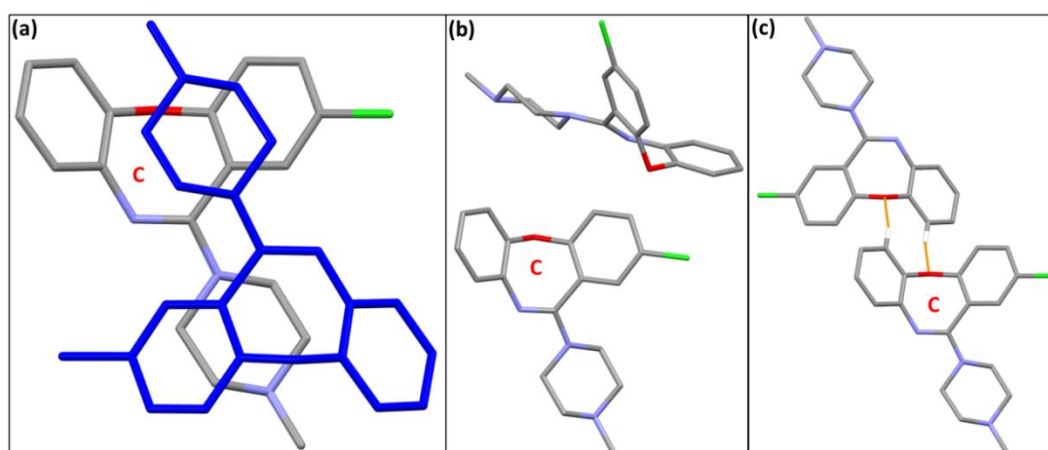


Figure 7.32. Top three molecular pairs in form I of loxapine, quantified in Table 7.12, (a) centrosymmetric dimer, (b) molecule-molecule interaction and (c) molecular pair involves C-H...O interaction (shown in orange line). The central molecule is labelled as 'C'. Other hydrogen atoms are omitted for clarity.

The total lattice energy is more evenly distributed between the nearest neighbouring interactions in form II as compared to the centrosymmetric dimer based form I (Table 7.12). The topmost interaction in form II is significantly less favourable (33.6 kJ mol<sup>-1</sup>) than that of the centrosymmetric dimer in form I (Table 7.12, Figure 7.33). This pair involves contributions from C-H...O and

CH... $\pi$  interactions between LXPN molecules from adjacent anti-parallel chains. The 2<sup>nd</sup> and 3<sup>rd</sup> ranked molecular pairs involve contribution from CH... $\pi$  interactions between neighbouring LXPN molecules.

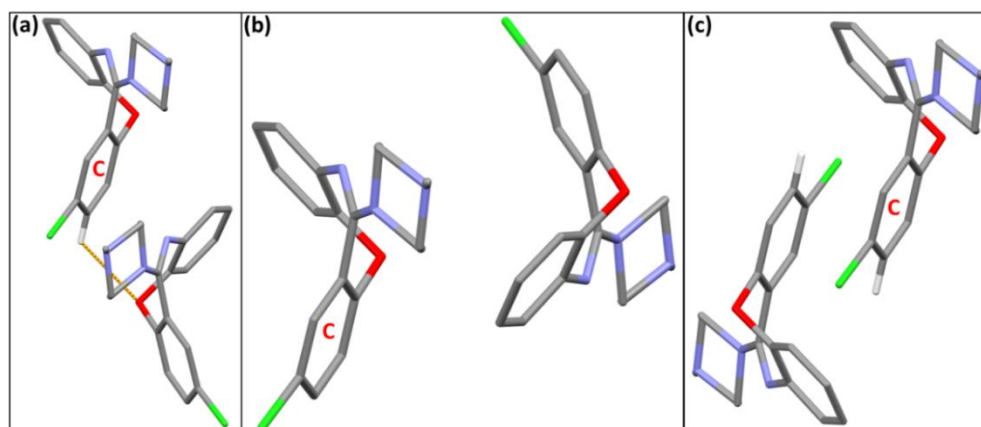


Figure 7.33. Top three pairwise intermolecular interactions in form II of loxapine quantified in Table 7.12. (a) Molecular pair involving C-H...O (shown by orange line) and CH<sub>2</sub>... $\pi$  interactions and (b-c) molecular pairs involving contribution from CH... $\pi$  interactions. The central molecule is labelled as 'C'. Other hydrogen atoms are omitted for clarity.

To summarise, PIXEL calculations showed that the dispersion is the key contributor towards crystal stabilisation for AXPN and both polymorphs of LXPN. In the case of the centrosymmetric dimer based form I of LXPN, the centrosymmetric dimer dominates the total energy whilst for the non-centrosymmetric dimer based form II of LXPN and AXPN the lattice energy is more evenly distributed amongst neighbouring interactions. In the absence of strong H-bonds being formed in the crystal structures of AXPN and LXPN, the crystal packing is stabilised by weaker, less directional C-H...O, C-H...N (Desiraju, 1991a; Desiraju, 2002) and van der Waals contacts. C-H...O and C-H...N H-bonds



are not as strong as the conventional H-bonds and provide minor contribution to the cohesive energy in the crystal lattice. However, in sufficient number these interactions can provide additional cohesive stability and clearly cannot be ignored.

## 7.4 Summary

Multidisciplinary studies involving experimental and computational approaches were applied to pharmaceutical molecules to facilitate the comparison of their solid-state structures. Specifically, this study has focused on exploring how changes in molecular structure affect the molecular conformation, packing motifs, interactions in the resultant crystal lattices and the extent of solid-state diversity of these compounds. The results highlight the value of crystal structure prediction studies and PIXEL calculations in the interpretation of the observed solid-state behaviour and quantifying the intermolecular interactions in the packed structures and identifying the key stabilising interactions.

An experimental screen yielded 4 physical forms for CZPN as compared to 60 distinct physical forms for OZPN. The experimental screening results of CZPN are consistent with its crystal energy landscape which confirms that no alternate packing arrangement is thermodynamically competitive to the experimentally obtained structure. Whilst in case of OZPN, crystal energy landscape highlights that the extensive experimental screening has probably not found all possible polymorphs of OZPN, and further solid form diversity could be targeted with a better understanding of the role of kinetics in its crystallisation. CSP studies were able to offer an explanation for the absence of the centrosymmetric dimer in anhydrous CZPN. PIXEL calculations on all the crystal structures of CZPN revealed that similar to OZPN, the intermolecular interaction energy in each structure is also dominated by the  $E_d$ .

Despite the molecular structure similarity between AXPB and LXPB (molecules in group 2), the crystal packing observed in polymorphs of LXPB differs significantly from the AXPB. A combined experimental and computational study demonstrated that the methyl group in LXPB has a significant influence in increasing the range of accessible solid forms and favouring various alternate packing arrangements.

CSP studies have again helped in explaining the observed solid-state diversity of LXPB and AXPB. PIXEL calculations showed that in absence of strong H-bonds, weak H-bonds such as C-H...O, C-H...N and dispersion interactions play a key role in stabilising the crystal lattice of both the molecules. Efficient crystal packing of AXPB seems to be contributing towards its monomorphic behaviour as compared to the comparatively less efficient packing of LXPB molecules in both polymorphs. The combination of experimental and computational approaches has provided a deeper understanding of the factors influencing the solid-state structure and diversity in these compounds.

Hirshfeld surfaces using Crystal Explorer represent another way of exploring packing modes and intermolecular interactions in molecular crystals (McKinnon et al., 2004). The influence of changes in the small substituents on shape and electron distribution can also be investigated by mapping the total electron density on the electrostatic potential for molecules in the gas phase. This allows straightforward visualisation and comparison of overall shape, electron-rich and electron-deficient regions within molecules. The shape of these molecules can be further investigated to study its influence on diverse solid-state diversity.

## **8 Conclusions and Further Work**

## 8.1 Conclusions and Further Work

The overall aim of this research was to investigate a range of experimental and computational approaches for the discovery and better understanding of the key factors underpinning solid-structure and diversity. Two groups of structurally similar compounds; olanzapine (OZPN) and clozapine (CZPN) in group 1 and amoxapine (AXPN) and loxapine (LXPN) in group 2 (Figure 8.1) were studied to investigate the influence of small changes in molecular structure and substituents on molecular conformation, the resultant intermolecular interactions, crystal packing motifs and the extent of solid form diversity. In addition, a fast validated method for physical form screening using Raman microscopy on multi-well plates was developed to explore the experimental crystallisation space. Also developing the theme of prediction, a statistical model for predicting crystallisability of small organic molecules was generated, which afforded ~70% prediction accuracy.

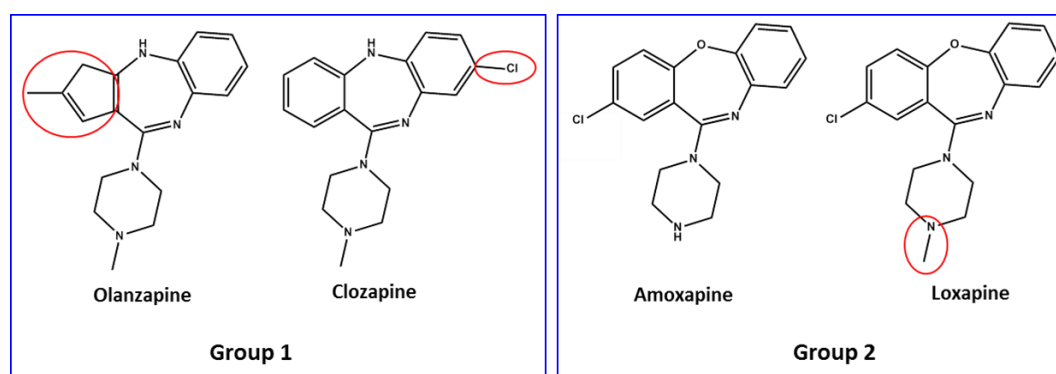


Figure 8.1. The four compounds studied in this thesis. The differences in the chemical structures within a group are highlighted in red.

The work therefore spans a broad range of experimental (Table 8.1) and computational approaches (Table 8.2). Underpinning the overall approach has been the use of a combination of analytical approaches to deliver complementary data on the systems under investigation.

Table 8.1. Recrystallisation methods used in this study to generate various solid forms

<b>Crystallisation Method</b>	<b>Degree of Freedom Used During Crystallisation</b>
Solvent evaporation	Solvent, rate of evaporation, initial concentration, temperature
Cooling crystallisation	Solvent, cooling profile, stirring/agitation, vacuum and concentration
Anti-solvent crystallisation	Solvent and anti-solvent
Crystallisation as a function of pH	Rate of change in pH and base
Crystallisation on functionalised surfaces	Surfaces
Mechanical grinding	Solvent, temperature, milling time and frequency
Recrystallisation from amorphous	Temperature and sample presentation (capillary, glass slides and vials)
Melt quenching	Temperature changes (minimum and maximum)
Vapour phase crystallisation	Temperature gradient
Desolvation of solvates	Temperature (isothermal or <i>in-situ</i> ) and vacuum
Spray drying	Solvent, concentration, pump speed and temperature programs
Freeze drying	Temperature programs.

Table 8.2. List of the computational tools used in this study

<b>Method</b>	<b>Application</b>
DASH, TOPAS and WinGX	Determination of crystal structure from XRPD data and/or SXD data
CASTEP	Verification of the crystal structure solved from XRPD data and determination of accurate positions of H-atoms.
Crystal structure prediction (CSP)	Prediction of thermodynamically feasible crystal packing arrangement from molecular structure.
PIXEL	Calculation of total lattice energy, contribution of polarisation, dispersion, Coulombic and repulsive interactions towards the total lattice energy and identification of the most significant intermolecular interaction within the crystal structures.
Gaussian	Provide molecular geometry and the electronic distribution around the molecules.
XPac	Determination of structural similarity at various levels (0-D, 1-D, 2-D and 3-D) amongst crystal structures.
Mercury	Visualisation and analysis of crystal structures e.g. H-bonding networks, exploration of crystal packing similarity and propensity of polymorphism etc.
Randomforests	Building of classification and regression models, Determination of important descriptors responsible for classification or regression models.
Principal Component Analysis	Identification of clusters, elucidation of similarities and differences within the data.
Group Analysis	Unsupervised clustering of the data.

During the course of this work the schematic workflow shown in Figure 8.2 has been developed providing a systematic approach to explore the organic solid state. Across the work presented herein, the complementary application of the experimental and computational approaches has helped in providing a more complete and accurate interpretation of the physical form diversity of the

molecules under study and in highlighting the factors impacting on the distinctly different solid-state behaviour of the 4 molecules studied.

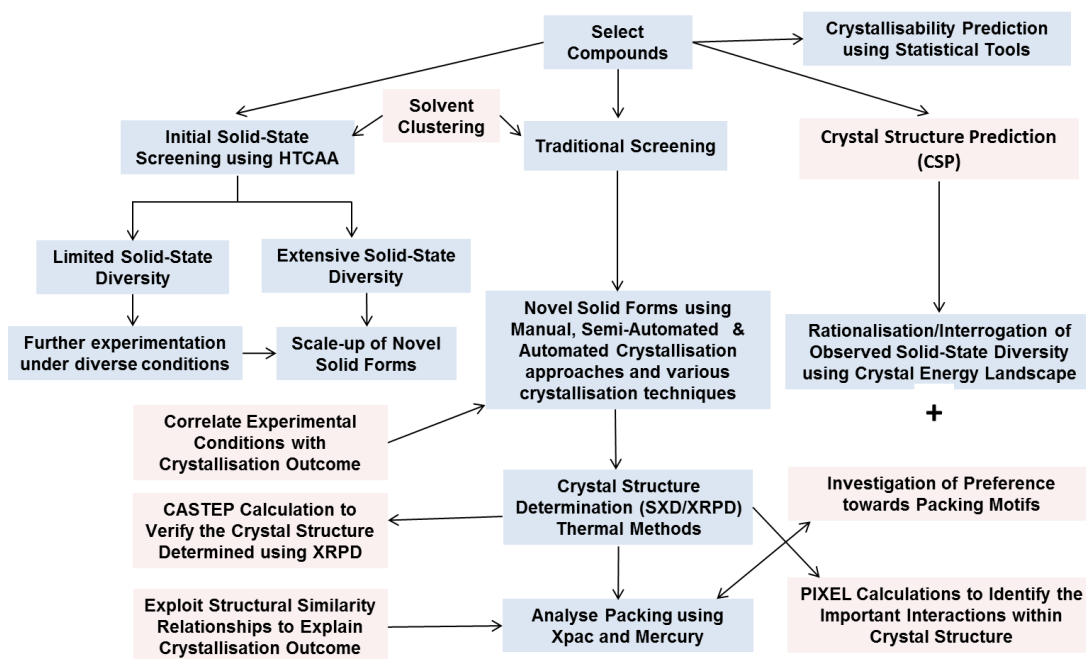


Figure 8.2. Schematic workflow developed in this work.

The conclusions of this thesis can be summarised in three sections namely, (i) development and validation of HTCAA methodology, (ii) crystallisability prediction of organic molecules using statistical modelling techniques and (iii) exploring the physical form landscape of structurally related pharmaceutical molecules in groups 1 (olanzapine and clozapine) and 2 (amoxapine and loxapine).



### **8.1.1 Development and Validation of High Throughput Crystallisation and Analysis Methodology for Physical Form Screening**

In the absence of reliable predictive models, systematic and rigorous experimental screening remains the best means to explore the full experimental crystallisation space and to provide a comprehensive understanding of solid form diversity and the conditions under which various solid forms of a molecule can be produced. Moreover in some cases, limited material is available for physical form screening which emphasises the requirement of an efficient means to maximise the information gained with minimal amount of a compound (less than ~250 mg). The physical form screening methodology developed in Chapter 4 can play an important role in achieving a detailed and accurate picture of solid-state diversity of compounds in a shorter period of time by utilising only milligram quantities of material.

However, the results obtained with OZPN suggest that this methodology might not provide a complete picture of the solid form diversity. It is advisable to use other complementary crystallisation techniques e.g. grinding and vapour phase crystallisation to maximise the coverage of experimental search space. Limitations associated with similar Raman spectra for physical forms emphasise the need of a combined use with another analytical tool such as XRPD to maximise the chances of correctly identifying different physical forms. Additionally, this methodology can be tricky for compounds exhibiting

fluorescence especially when the interference cannot be overcome by applying background corrections.

Due to its relative simplicity, parallelisation and miniaturisation, this methodology has potential in routine physical form screening and assessing the extent of solid form diversity of compounds. It has considerable scope and by simulating the typical conditions during downstream processing or secondary manufacturing, it can also be employed for wider applications such as in the study of the influence of excipients and process variables on possible phase transformation on a very small scale. This prior knowledge of the landscape of parameters affecting the physicochemical stability of the physical form can potentially be used during later stages of development by creating the design space for implementation of QbD approaches to guarantee the best quality pharmaceutical product.

### **8.1.2 Predicting Crystallisability of Organic Molecules using Statistical Modelling Techniques**

The studies in chapters 4, 6 and 7 investigate the effects of molecular structure on crystal packing, intermolecular interactions and extent of physical form diversity. Whereas the statistical model built in chapter 5 explores the impact of molecular structure on crystallisability and provides a means to highlight and understand the molecular factors that inhibit or promote crystallisation.

The RF classification model for predicting crystallisability of small organic molecules yielded a prediction accuracy of ~70%. This is the first study on

crystallisability prediction for small molecules using statistical modelling techniques and although it does not give 100% accuracy, it shows good potential for further development. This model provides a reasonable opportunity to highlight problematic compounds (e.g. those exhibiting nano/micro crystal formation, agglomeration, oiling out, slow nucleation etc.) so that resource planning can be accommodated to obtain effective crystallisation processes.

Prediction of crystallisability (here defined as the ability of a molecule to come out of solution and form a crystal suitable for SXD) is a challenge. This model essentially deals with the process of crystallisation i.e. nucleation and growth under the conditions tested but not the resultant crystal structure. Although, this model does not provide a mechanistic understanding of the crystallisation process, it still represents a rational and pragmatic approach which enables crystallisability prediction with reasonable degree of confidence.

The use of this approach also allowed identification of the most important descriptors responsible for differences in crystallisation behaviour of structural analogues and they matched well with the experimental observations (Hursthouse et al., 2009). However, the importance of descriptors cannot be quantified using this method. Still this information provides a potential means to identify crystallisation issues during initial studies and can help in designing improved crystallisation processes. These types of statistical models provide reasonably accurate predictions for molecules similar to the molecules used for

training the model, but might not be as useful for compounds which are not structurally related to the molecules employed in the training dataset.

### **8.1.3 Exploring the Physical Form Landscape of Structurally Related Pharmaceutical Molecules in Group 1 (Olanzapine and Clozapine) and 2 (Amoxapine and Loxapine)**

The work in chapter 6 and 7 was focussed on studying the effect of minor changes in the substituents on the solid-state structure and diversity and to develop a better understanding of the factors underpinning those changes.

Extensive solid form screening of OZPN yielded 60 solid forms including three polymorphs, 56 solvates and an amorphous form. For OZPN, the 124 experiments specifically targeting non-solvated forms (grinding, desolvation, sublimation, melt-quenching, crystallisation from amorphous form, spray and freeze-drying) only produced 3 polymorphs. All these efforts did not find any conditions that could produce phase pure samples of the metastable polymorph III, though a single crystal of form II could be extracted and was used to solve the crystal structure. From a mixed phase sample, no such crystals for form III were obtained. It is plausible that form III contains the same  $SC_{24}$  layers as form II, but stacked in the alternative manner similar to the hypothetical structure, A162, which is only slightly less stable than form II (see Section 6.8). Although A162 gave the reasonable fit to the form III powder data in a two-phase refinement, the residual misfit confirms that the space group and/or lattice does not match to that of form III. (see Section 6.8). Nonetheless, the close

relationship suggested between the known and unknown structures by available experimental and hypothetical structures provides a potential explanation for the observed difficulty of preparing phase pure samples.

In OZPN, there is a favoured layer structure (SC<sub>21</sub>), which stacks badly with itself, but when water and other solvent molecules are introduced between the layers, it forms three sets of iso-structural solvates. The RF classification model has illustrated that the solvate formation in these three sets of iso-structural solvates is directed by the physicochemical properties of the solvents (see Section 6.5 and 6.7). The extensive range of almost three hundred solution crystallisations (evaporations, cooling, anti-solvent, pH change and on functionalised surfaces) of OZPN and the resulting structures of the 56 solvates suggest that other solvates, particularly solvate hydrates, are likely. In contrast only 3 solvates for CZPN have been obtained so far and number of solvates would very likely to be limited.

A key finding from the experimental and CSP studies on CZPN is that the appearance of its alternative polymorph using standard crystallisations techniques is unlikely. The crystal energy landscape highlights limited potential of CZPN in forming crystal structures with a centrosymmetric dimer. However the presence of the centrosymmetric dimer in solvates might be due to the stabilisation of their crystal lattices by the inclusion of solvents. The crystal energy landscape for OZPN shows that there are alternative, probably metastable, structures for anhydrous OZPN. Some are not based on the SC<sub>0</sub> dimer found in all the experimental structures. PIXEL calculations on OZPN

show that the strongly dispersion-bound dimer unit SC<sub>0</sub> dominates the lattice energy of form I and II, whereas the lattice energy can be more evenly distributed between the nearest neighbour interactions in some hypothetical structures. This raises the question as to whether the inability to crystallise any novel polymorphs or solvates of OZPN without the SC<sub>0</sub> dimer is a kinetic effect due to the SC<sub>0</sub> dimer being the key growth unit in all the crystallisation experiments. Whether the computed structures without the SC<sub>0</sub> dimer are genuinely credible thermodynamic polymorphs, whose kinetic disadvantage or ease of transformation to form I might be overcome in a specific experiment, or theoretical anomalies, remains an important question.

Another pair of molecules studied in this work was of LXP<sub>N</sub> and AXP<sub>N</sub>. An important outcome of the multidisciplinary study on AXP<sub>N</sub> is that it neither exhibits polymorphism nor has the tendency of solvate formation. This study also illustrated that the presence of methyl substituents at the 4 position of the piperazine ring in LXP<sub>N</sub> has a significant influence on the physical form diversity. This study demonstrates that the minor substituents can have a disruptive effect on the intermolecular interactions, packing motifs, resultant crystal structures and the extent of solid-state diversity. It also illustrates the challenges associated with the prediction of small substituent's effect and its magnitude and importance of computational tools in providing reasonable explanations of the experimental results.

Therefore, further studies of this type have an important role to play in better understanding of substituent effects on solid-state behaviour. Furthermore,

these studies might be able to provide information about structural traits which are responsible for polymorphism and whether these structural traits can be used to predict solid-state behaviour of new compounds. The presence of similar SCs in 0-D (dimer) or 1-D (chains) or 2-D (sheets) in two solid forms of structurally similar molecules may indicate the particular preferences in both nucleation and crystal growth processes independent of the bonding environment. The knowledge of common SCs may help in controlling the crystallisation behaviour of a given compound (Gelbrich and Hursthouse, 2005). Knowledge of SCs present across different crystal environments may also be used to predict the crystal structures (Sarma and Desiraju, 2002).

As has been demonstrated in this study, a realistic crystal energy landscape is an important tool for interpreting the range of experimental solid forms as well as in proposing possible structures that have yet to be crystallised. However, information about the experimental conditions and control to obtain the predicted polymorphs cannot be as yet deduced from the crystal energy landscape. The development of improved theoretical approaches for accurate calculated energy landscapes alongside establishing experimental approaches that enable specific forms to be targeted, rather than found through necessarily limited screening or serendipity is an important and still elusive goal for the control and prediction of pharmaceutical solids.

## **8.2 Further Work**

### **8.2.1 Development and Validation of High Throughput Crystallisation and Analysis Methodology for Physical Form Screening**

There is a considerable scope for the further improvement and refinement of this methodology to have better control over crystallisation conditions and reproducible results. For example, implementation of stirring element for each well such as magnetic stir disc or overhead stirrers (VP 721F-1 or VP 177-STIR system, available from VP-Scientific, US) would enable better control and dissolution as compared to currently adopted orbital shaker. Adoption of automated solvent and solute dispensing unit would enable better control over initial concentration of solution in all wells of MWP, reduce errors due to manual dispensing and provide reproducible results. Implementation of a filtration unit (96/48-well filter plate) could potentially remove the seeds or the undissolved solute left due to its less solubility in the respective solvent which can affect the crystallisation outcome. Implementation of mica insulated temperature units specifically designed for keeping MWPs at constant temperature with/without computer control (e.g. VP 741AW-RCE-MICA-B system available from VP-Scientific, US) would allow uniform and better control over temperature during crystallisation at various temperatures.

At present, this methodology uses Raman microscopy as the analytical technique. Different solid forms of a compound may provide similar Raman



spectra and novel solid forms with similar Raman spectra can be missed out (Stahly, 2007). Therefore, it would be advantageous to include XRPD also to the set up to confirm the sample assignments. For initial high-throughput screening, XRPD is widely used as an analytical tool for primary identification of recrystallised samples and/or for confirmation of the sample identified by an initial Raman analysis (Morissette et al., 2003; Remenar et al., 2003). An ideal implementation would be the Bruker-AXS General Area Diffraction Detector System (GADDS™) equipped with automated *x-y-z* sample stage with video and laser-guided sample alignment, X-ray source with small focal point and a large 2-D area detector (Storey et al., 2004). Chemometric analysis of the data obtained from both techniques would reduce chances of false positives/negatives compared to Raman data alone.

This methodology has been applied for salt, polymorph and solvate screening using solvent evaporation technique. Further investigation of this methodology for various crystallisation techniques (e.g. anti-solvent crystallisation, slurry, and co-crystal screening) and studies involving effect of water/relative humidity on API-excipient interaction, role of excipient in salt disproportionation in formulation, possible API-polymer interactions and the effect of process-induced stresses like heat, solvent, shear and pressure on the formulation will provide more confidence in its general applicability.

A large amount of data is generated using HTCAA. The application of statistical modelling techniques to this data offers a potential tool for modelling the property space which involves fitting of the crystallisation outcome (solid form,

morphology etc.) with the physicochemical descriptors and experimental conditions. These models can reveal various trends between the molecular properties, experimental conditions and crystallisation outcome thus identifying the key factors responsible for observed outcome (Johnston et al., 2008; McCabe, 2010). The significant statistical models can be stored and retrieved later to direct future experiments.

### **8.2.2 Predicting Crystallisability of Organic Molecules using Statistical Modelling Techniques**

The next step towards achieving this kind of statistical modelling approach would be to incorporate information about crystallisation conditions such as solvent, rate of solvent evaporation, RH, temperature etc. which would certainly improve the value of the training data set and hence, predictive capability. These models need to be updated with the increasing knowledge to provide valuable information about crystallisation behaviour of a wider range of molecules. The extension and application of this kind of statistical model to salts and co-crystal may also provide additional insights into the ease to crystallisation of multi-component systems.

Various predictive models for peptides and proteins are available (see Section 5.1.5.6); however there are no reported models for prediction of crystallisability of small molecules. It is worth considering the reasons for this. One key reason may be the absence of large, suitable and systematically studied datasets and lack of associated data required to make these models. To make robust models,

systematic studies are required to obtain sufficiently comprehensive datasets which are not commonly done. However with advancements in instrumentation and automation, it is now possible to generate huge datasets of crystallisation properties (Hursthouse, 2004; Storey et al., 2004). In addition, it is also important to store all the relevant data in accessible electronic database formats. These databases with suitable statistical modelling techniques would open the avenues for researchers to study relationships between solute, solvent, physical form and crystallisation conditions (Johnston et al., 2008).

### **8.2.3 Exploring the Physical Form Landscape of Structurally Related Pharmaceutical Molecules in Group 1 (Olanzapine and Clozapine) and 2 (Amoxapine and Loxapine)**

In OZPN, the large  $SC_0$  dimer binding energy dominates the lattice energy. This merits further investigation of the degree of self-association in the prenucleation solution state (Parveen et al., 2005) and amorphous phase (Billinge et al., 2010).

Evidently, during OZPN nucleation in the solution recrystallisations, solid-state desolvation, grinding experiments and vapour phase crystallisations tested, OZPN molecules preferentially associate as the centrosymmetric dimer. The dominant  $SC_0$  interactions may therefore swamp initial short length-scale aggregation such that possible structures based on alternative growth units are starved of single molecule building blocks. The possibility of finding other non-solvated OZPN polymorphs may well be enhanced by less typical crystallisation

conditions (Llinàs and Goodman, 2008) designed to disrupt the centrosymmetric dimer or favour the alternative molecular conformation. This might be achieved by laser induced nucleation (Sun et al., 2008), ultra-sound assisted crystallisation (Gracin et al., 2005), systematic attempts at templating, for example by crystals of related molecules (Arlin et al., 2011), polymers (Grzesiak et al., 2006; McKellar et al., 2012) or other surfaces including metals or self-assembled monolayers (Hiremath et al., 2005). The chances of success are likely to be increased by the design of a specific template to target the crystallisation of a particular alternative structure. This kind of experimental interventions might enable the nucleation process to follow an alternative pathway to one based on SC<sub>0</sub>.

Experimental work on LXP<sub>N</sub> was based only on solution crystallisations. Further investigation of the experimental structure landscape by utilising multiple crystallisation techniques would potentially reveal the full extent of its solid-state diversity. Comparisons of molecular shape of all these molecules using electrostatic potential surfaces calculations and Hirshfeld surface analysis (McKinnon et al., 2004) should be explored to study the influence of overall molecular shape on distinct solid-state behaviour.

CSP studies have demonstrated their valuable role in providing more confidence in the interpretation of the experimental data. Further CSP studies using better approximation and flexible models for CZPN, AXP<sub>N</sub> and LXP<sub>N</sub> would provide additional insights into the preferences of the packing motifs (Johnston et al., 2011). An obvious area of further work is to investigate if any of the

experimentally observed crystal structure of these molecules can act as a template for the predicted structure of structurally related molecules on the crystal energy landscape (Arlin et al., 2011).

## 9 References

- Aakeröy, C. B.; Forbes, S.; Desper, J. (2009). Using cocrystals to systematically modulate aqueous solubility and melting behavior of an anticancer drug. *Journal of the American Chemical Society*, **131**, 17048-17049.
- Aakeroy, C. B.; Salmon, D. J. (2005). Building co-crystals with molecular sense and supramolecular sensibility. *CrystEngComm*, **7**, 439-448.
- Aaltonen, J.; Allesø, M.; Mirza, S.; Koradia, V.; Gordon, K. C.; Rantanen, J. (2009). Solid form screening – A review. *European Journal of Pharmaceutics and Biopharmaceutics*, **71**, 23-37.
- Abraham, R. J.; Kricka, L. J.; Ledwith, A. (1974). The nuclear magnetic resonance spectra and conformations of cyclic compounds. Part X. Conformational equilibria in 5-substituted 10,11-dihydrodibenz[b,f]azepines. *Journal of the Chemical Society, Perkin Transactions 2*, 1648-1654.
- Accelrys (2010). Pipeline pilot in chemistry collection: Basic chemistry user guide, Accelrys Inc., 5005 Wateridge Vista Drive, San Diego, CA 92121.
- 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.; Ghogale, P. P.; Ghosh, S.; Goswami, P. K.; Goud, N. R.; Jetti, R. R. K. R.; Karpinski, P.; Kaushik, P.; Kumar, D.; Kumar, V.; Moulton, B.; Mukherjee, A.; Mukherjee, G.; Myerson, A. S.; Puri, V.; Ramanan, A.; Rajamannar, T.; Reddy, C. M.; Rodriguez-Hornedo, N.; Rogers, R. D.; Row, T. N. G.; Sanphui, P.; Shan, N.; Shete, G.; Singh, A.; Sun, C. C.; Swift, J. A.; Thaimattam, R.; Thakur, T. S.; Kumar Thaper, R.; Thomas, S. P.; Tothadi, S.; Vangala, V. R.; Variankaval, N.; Vishweshwar, P.; Weyna, D. R.; Zaworotko, M. J. (2012). Polymorphs, salts, and cocrystals: What's in a name? *Crystal Growth & Design*, **12**, 2147-2152.
- Al-Dulaimi, S.; Aina, A.; Burley, J. (2010). Rapid polymorph screening on milligram quantities of pharmaceutical material using phonon-mode Raman spectroscopy. *CrystEngComm*, **12**, 1038-1040.
- Almarsson, Ö.; Hickey, M. B.; Peterson, M. L.; Morissette, S. L.; Soukasene, S.; McNulty, C.; Tawa, M.; MacPhee, J. M.; Remenar, J. F. (2003). High-throughput surveys of crystal form diversity of highly polymorphic pharmaceutical compounds. *Crystal Growth & Design*, **3**, 927-933.
- Almarsson, O.; Zaworotko, M. J. (2004). Crystal engineering of the composition of pharmaceutical phases. Do pharmaceutical co-crystals represent a new path to improved medicines? *Chemical Communications*, 1889-1896.
- Alsenz, J.; Kansy, M. (2007). High throughput solubility measurement in drug discovery and development. *Advanced Drug Delivery Reviews*, **59**, 546-567.

Anderton, C. (2007). A valuable technique for polymorph screening. *American Pharmaceutical Review*, **10**, 36-40.

Angelo, G. (2007). A solid-state chemist's view of the crystal polymorphism of organic compounds. *Journal of Pharmaceutical Sciences*, **96**, 2232-2241.

APEX2 (2005). APEX2 user manual version 1.2.7, Bruker AXS Inc., 5465 East Cheryl Parkway, Madison, WI 53711.

Arlin, J.-B.; Johnston, A.; Miller, G. J.; Kennedy, A. R.; Price, S. L.; Florence, A. J. (2010). A predicted dimer-based polymorph of 10,11-dihydrocarbamazepine (Form IV). *CrystEngComm*, **12**, 64-66.

Arlin, J.-B.; Price, L. S.; Price, S. L.; Florence, A. J. (2011). A strategy for producing predicted polymorphs: Catemeric carbamazepine form V. *Chemical Communications*, **47**, 7074-7076.

Ashby, C. J.; Wang, R. (1996). Pharmacological actions of the atypical antipsychotic drug clozapine: A review. *Synapse*, **24**, 349-394.

Baird, J.; Santiago-Quinonez, D.; Rinaldi, C.; Taylor, L. (2012). Role of viscosity in influencing the glass-forming ability of organic molecules from the undercooled melt state. *Pharmaceutical Research*, **29**, 271-284.

Baird, J. A.; Van Eerdenbrugh, B.; Taylor, L. S. (2010). A classification system to assess the crystallization tendency of organic molecules from undercooled melts. *Journal of Pharmaceutical Sciences*, **99**, 3787-3806.

Balbach, S.; Korn, C. (2004). Pharmaceutical evaluation of early development candidates "the 100 mg-approach". *International Journal of Pharmaceutics*, **275**, 1-12.

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.; van Eijck, B. P.; Facelli, J. C.; Ferraro, M. B.; Grillo, D.; Habgood, M.; Hofmann, D. W. M.; Hofmann, F.; Jose, K. V. J.; Karamertzanis, P. G.; Kazantsev, A. V.; Kendrick, J.; Kuleshova, L. N.; Leusen, F. J. J.; Maleev, A. V.; Misquitta, A. J.; Mohamed, S.; Needs, R. J.; Neumann, M. A.; Nikylov, D.; Orendt, A. M.; Pal, R.; Pantelides, C. C.; Pickard, C. J.; Price, L. S.; Price, S. L.; Scheraga, H. A.; van de Streek, J.; Thakur, T. S.; Tiwari, S.; Venuti, E.; Zhitkov, I. K. (2011). Towards crystal structure prediction of complex organic compounds - A report on the fifth blind test. *Acta Crystallographica Section B*, **67**, 535-551.

Bastin, R. J.; Bowker, M. J.; Slater, B. J. (2000). Salt selection and optimisation procedures for pharmaceutical new chemical entities. *Organic Process Research & Development*, **4**, 427-435.

- Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. (2001). Ritonavir: An extraordinary example of conformational polymorphism. *Pharmaceutical Research*, **18**, 859-866.
- Beckmann, W. (2000). Seeding the desired polymorph: background, possibilities, limitations, and case studies. *Organic Process Research & Development*, **4**, 372-383.
- Bernstein, J. (2002). Polymorphism and patents. Polymorphism in molecular crystals, Oxford University Press Inc., New York, 297-307.
- Bernstein, J. (2005). Cultivating crystal forms. *Chemical Communications*, 5007-5012.
- Bernstein, J. (2011). Polymorphism – A perspective. *Crystal Growth & Design*, **11**, 632-650.
- Bernstein, J. (June 19, 2004). Polymorphism and patents from a chemist's point of view. Diversity Amidst Similarity: A Multidisciplinary Approach to Polymorphs, Solvates and Phase Relationship, Erice, Sicily.
- Bernstein, J.; Davey, R. J.; Henck, J.-O. (1999). Concomitant polymorphs. *Angewandte Chemie International Edition*, **38**, 3440-3461.
- Bhardwaj, R. M.; Price, L. S.; Price, S. L.; Reutzel-Edens, S. M.; Miller, G. J.; Oswald, I. D. H.; Johnston, B. F.; Florence, A. J. (2013). Exploring the experimental and computed crystal energy landscape of olanzapine. *Crystal Growth & Design*, **13**, 1602-1617.
- Bhugra, C.; Shmeis, R.; Pikal, M. J. (2008). Role of mechanical stress in crystallization and relaxation behavior of amorphous indomethacin. *Journal of Pharmaceutical Sciences*, **97**, 4446-4458.
- Billinge, S. J. L.; Dykhne, T.; Juhas, P.; Bozin, E.; Taylor, R.; Florence, A. J.; Shankland, K. (2010). Characterisation of amorphous and nanocrystalline molecular materials by total scattering. *CrystEngComm*, **12**, 1366-1368.
- Blagden, N.; Davey, R. J. (2003). Polymorph selection: Challenges for the future? *Crystal Growth & Design*, **3**, 873-885.
- Blagden, N.; Davey, R. J.; Rowe, R.; Roberts, R. (1998). Disappearing polymorphs and the role of reaction by-products: The case of sulphathiazole. *International Journal of Pharmaceutics*, **172**, 169-177.
- Bond, A. D. (2007). What is a co-crystal? *CrystEngComm*, **9**, 833-834.
- Bond, A. D.; Boese, R.; Desiraju, G. R. (2007). On the polymorphism of aspirin: crystalline aspirin as intergrowths of two "polymorphic" domains". *Angewandte Chemie-International Edition*, **46**, 618-622.



- Boultif, A.; Louer, D. (1991). Indexing of powder diffraction patterns for low-symmetry lattices by the successive dichotomy method. *Journal of Applied Crystallography*, **24**, 987-993.
- Braga, D.; Grepioni, F.; Maini, L. (2010). The growing world of crystal forms. *Chemical Communications*, **46**, 6232-6242.
- Braun, D. E.; Bhardwaj, R. M.; Florence, A. J.; Tocher, D. A.; Price, S. L. (2012). Complex polymorphic system of gallic acid—five monohydrates, three anhydrates, and over 20 solvates. *Crystal Growth & Design*, **13**, 19-23.
- Braun, D. E.; Karamertzanis, P. G.; Arlin, J.-B.; Florence, A. J.; Kahlenberg, V.; Tocher, D. A.; Griesser, U. J.; Price, S. L. (2010). Solid-state forms of  $\beta$ -resorcylic acid: How exhaustive should a polymorph screen be? *Crystal Growth & Design*, **11**, 210-220.
- Breiman, L. (2001). Random forests. *Machine Learning*, **45**, 5-32.
- Bruno, I. J.; Cole, J. C.; Kessler, M.; Luo, J.; Motherwell, W. D. S.; Purkis, L. H.; Smith, B. R.; Taylor, R.; Cooper, R. I.; Harris, S. E.; Orpen, A. G. (2004). Retrieval of crystallographically-derived molecular geometry information. *Journal of Chemical Information and Computer Sciences*, **44**, 2133-2144.
- Bugay, D. E. (2001). Characterization of the solid-state: Spectroscopic techniques. *Advanced Drug Delivery Reviews*, **48**, 43-65.
- Bulow, J. (2004). The gaming of pharmaceutical patents. Innovation Policy and the Economy. A. B. Jaffe, J. Lerner and S. Stern. The MIT Press, Cambridge, **4**, 145-185.
- Bureau, A.; Dupuis, J.; Falls, K.; Lunetta, K. L.; Hayward, B.; Keith, T. P.; Van Eerdewegh, P. (2005). Identifying SNPs predictive of phenotype using random forests. *Genetic Epidemiology*, **28**, 171-182.
- Byrn, S. R. (1982). Polymorphism of drugs. Solid-state chemistry of drugs, Academic Press Inc., New York, 79-188.
- Byrn, S. R.; Pfeiffer, R. R.; Stowell, J. G. (1999). Hydrates and solvates. Solid state chemistry of drugs, SSCI Inc., West Lafayette, 233-247.
- Callear, S. (2008). Preparation, characterisation and structural assessment of salts and co-crystals of organic compounds. Doctor of Philosophy, University of Southampton.
- Capes, J. S.; Cameron, R. E. (2006). Contact line crystallization to obtain metastable polymorphs. *Crystal Growth & Design*, **7**, 108-112.
- Capuano, B.; Crosby, I.; Forsyth, C.; McRobb, F.; Moudretski, V.; Taylor, D.; Vom, A.; Yuriev, E. (2010). New hybrids of clozapine and haloperidol and their

isosteric analogues: synthesis, X-ray crystallography, conformational analysis and preliminary pharmacological evaluation. *Structural Chemistry*, **21**, 613-628.

Capuano, B.; Crosby, I. T.; Egan, S. J.; Fallon, G. D.; Lloyd, E. J.; Neve, J. E. (2005). 4-Chloro-N-[4-(8-chloro-5H-dibenzo[b,e][1,4]diazepin-11-yl)-1-methylpiperazinio]benzamidate dichloromethane solvate. *Acta Crystallographica Section E*, **61**, o20-o22.

Capuano, B.; Crosby, I. T.; Fallon, G. D.; Lloyd, E. J.; Yuriev, E.; Egan, S. J. (2003). 2-Methyl-4-(4-methylpiperazin-1-yl)-10H-thieno[2,3-b][1,5]benzodiazepine methanol solvate monohydrate. *Acta Crystallographica Section E*, **59**, o1367-o1369.

Capuano, B.; Crosby, I. T.; Gable, R. W.; Lloyd, E. J. (2000). N-Piperonyl analogue of the atypical antipsychotic clozapine. *Acta Crystallographica Section C*, **56**, 339-340.

Casimir, A.; Andrew, J. W. (1997) Sertraline polymorph having improved water solubility. Torcan Chemical Ltd, EP 0928784A1.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, **1**, 245-276.

Chemburkar, S. R.; Bauer, J.; Deming, K.; Spiwek, H.; Patel, K.; Morris, J.; Henry, R.; Spanton, S.; Dziki, W.; Porter, W.; Quick, J.; Bauer, P.; Donaubaue, J.; Narayanan, B. A.; Soldani, M.; Riley, D.; McFarland, K. (2000). Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development. *Organic Process Research & Development*, **4**, 413-417.

Chen, S.; Xi, H.; Yu, L. (2005). Cross-nucleation between ROY polymorphs. *Journal of the American Chemical Society*, **127**, 17439-17444.

Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. J.; Refson, K.; Payne, M. C. (2005). First principles methods using CASTEP. *Zeitschrift Fur Kristallographie*, **220**, 567-570.

Clarke, H. D.; Arora, K. K.; Bass, H.; Kavuru, P.; Ong, T. T.; Pujari, T.; Wojtas, L.; Zaworotko, M. J. (2010). Structure–stability relationships in cocrystal hydrates: does the promiscuity of water make crystalline hydrates the nemesis of crystal engineering? *Crystal Growth & Design*, **10**, 2152-2167.

Clarke, H. D.; Hickey, M. B.; Moulton, B.; Perman, J. A.; Peterson, M. L.; Wojtas, L.; Almarsson, Ö.; Zaworotko, M. J. (2012). Crystal engineering of isostructural quaternary multicomponent crystal forms of olanzapine. *Crystal Growth & Design*, **12**, 4194-4201.

Coelho, A. (2003a). Indexing of powder diffraction patterns by iterative use of singular value decomposition. *Journal of Applied Crystallography*, **36**, 86-95.

- Coelho, A. (2003b). TOPAS user manual. *Bruker AXS*, Karlsruhe, Germany.
- Coelho, A. (2007). Users Manual: TOPAS Academic version 4.1.
- Cooper, T. G.; Hejczyk, K. E.; Jones, W.; Day, G. M. (2008). Molecular polarization effects on the relative energies of the real and putative crystal structures of valine. *Journal of Chemical Theory and Computation*, **4**, 1795-1805.
- Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. (2002). New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *The Journal of Chemical Physics*, **117**, 43-54.
- Cosulich, D. B.; Lovell, F. M. (1977). The X-ray crystal structures of loxapine {2-chloro-11-(4-methyl-1-piperazinyl)dibenz[b,f][1,4]oxazepine} and amoxapine {2-chloro-11-(1-piperazinyl)dibenz[b,f][1,4]oxazepine}. *Acta Crystallographica Section B*, **33**, 1147-1154.
- Cross, W. I.; Blagden, N.; Davey, R. J.; Pritchard, R. G.; Neumann, M. A.; Roberts, R. J.; Rowe, R. C. (2002). A whole output strategy for polymorph screening: Combining crystal structure prediction, graph set analysis, and targeted crystallization experiments in the case of diflunisal. *Crystal Growth & Design*, **3**, 151-158.
- Cruz Cabeza, A. J.; Day, G. M.; Motherwell, W. D. S.; Jones, W. (2006). Importance of molecular shape for the overall stability of hydrogen bond motifs in the crystal structures of various carbamazepine-type drug molecules. *Crystal Growth & Design*, **7**, 100-107.
- Davey, R. J. (2004). Crystallization: How come you look so good? *Nature*, **428**, 374-375.
- Davey, R. J.; Allen, K.; Blagden, N.; Cross, W. I.; Lieberman, H. F.; Quayle, M. J.; Righini, S.; Seton, L.; Tiddy, G. J. T. (2002). Crystal engineering - Nucleation, the key step. *CrystEngComm*, **4**, 257-264.
- Davey, R. J.; Blagden, N.; Potts, G. D.; Docherty, R. (1997). Polymorphism in molecular crystals: Stabilization of a metastable form by conformational mimicry. *Journal of the American Chemical Society*, **119**, 1767-1772.
- Davey, R. J.; Schroeder, S. L. M.; ter Horst, J. H. (2013). Nucleation of organic crystals—A molecular perspective. *Angewandte Chemie International Edition*, **52**, 2166-2179.
- David, W. I. F.; Shankland, K.; Shankland, N. (1998). Routine determination of molecular crystal structures from powder diffraction data. *Chemical Communications*, 931-932.

David, W. I. F.; Shankland, K.; van de Streek, J.; Pidcock, E.; Motherwell, W. D. S.; Cole, J. C. (2006). DASH: A program for crystal structure determination from powder diffraction data. *Journal of Applied Crystallography*, **39**, 910-915.

Day, G. M. (2011). Current approaches to predicting molecular organic crystal structures. *Crystallography Reviews*, **17**, 3-52.

Debeljak, Ž.; Škrbo, A.; Jasprica, I.; Mornar, A.; Plečko, V.; Banjanac, M.; Medić-Šarić, M. (2007). QSAR study of antimicrobial activity of some 3-nitrocoumarins and related compounds. *Journal of Chemical Information and Modeling*, **47**, 918-926.

Desiraju, G. R. (1991a). The C-H...O hydrogen bond in crystals: What is it? *Accounts of Chemical Research*, **24**, 290-296.

Desiraju, G. R. (1991b). Hydration in organic crystals: Prediction from molecular structure. *Journal of the Chemical Society, Chemical Communications*, 426-428.

Desiraju, G. R. (2002). Hydrogen Bridges in Crystal Engineering: Interactions without Borders. *Accounts of Chemical Research*, **35**, 565-573.

Desiraju, G. R. (2003). Crystal and co-crystal. *CrystEngComm*, **5**, 466-467.

Desrosiers, P. J. (2004). The potential of preform. *Modern Drug Discovery*, **7**, 40-43.

Di Profio, G.; Tucci, S.; Curcio, E.; Drioli, E. (2007). Selective glycine polymorph crystallization by using microporous membranes. *Crystal Growth & Design*, **7**, 526-530.

Doniger, S.; Hofmann, T.; Yeh, J. (2004). Predicting CNS permeability of drug molecules: Comparison of neural network and support vector machine algorithms *Journal of Computational Biology*, **9**, 849-864.

Dunitz, J. D.; Gavezzotti, A. (2005a). Molecular recognition in organic crystals: directed intermolecular bonds or nonlocalized bonding? *Angewandte Chemie International Edition*, **44**, 1766-1787.

Dunitz, J. D.; Gavezzotti, A. (2005b). Toward a quantitative description of crystal packing in terms of molecular pairs: Application to the hexamorphic crystal system, 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile. *Crystal Growth & Design*, **5**, 2180-2189.

Dunitz, J. D.; Gavezzotti, A.; Schweizer, W. B. (2003). Molecular shape and intermolecular liaison: Hydrocarbons and fluorocarbons. *Helvetica Chimica Acta*, **86**, 4073-4092.

Dupont, L.; Liegeois, J.-F. (2003). 8-chloro-5-(4-methylpiperazin-1-yl)-11H-pyrido[2,3-b][1,5]benzoxazepine. *Acta Crystallographica Section E*, **59**, o1962-o1963.

Dupont, L.; Liegeois, J.-F.; Rogister, F.; Delarge, J. (1996). 8-chloro-5-(4-methylpiperazin-1-yl)-11H-pyrido[2,3-b][1,5]benzodiazepine. *Acta Crystallographica Section C*, **52**, 391-393.

Eckhardt, C. J.; Gavezzotti, A. (2007). Computer simulations and analysis of structural and energetic features of some crystalline energetic materials. *The Journal of Physical Chemistry B*, **111**, 3430-3437.

Eder, B. K.; Davis, J. M.; Bloomfield, P. (1994). An automated classification scheme designed to better elucidate the dependence of ozone on meteorology. *Journal of Applied Meteorology*, **33**, 1182-1199.

Ejgenberg, M.; Mastai, Y. (2012). Crystallization on self assembled monolayers, *Advances in Crystallization Processes*. From <http://www.intechopen.com/books/advances-in-crystallization-processes/crystallization-on-self-assembledmonolayers>.

Etter, M. C. (1990). Encoding and decoding hydrogen-bond patterns of organic compounds. *Accounts of Chemical Research*, **23**, 120-126.

Fabbiani, F. P. A.; Allan, D. R.; Parsons, S.; Pulham, C. R. (2005). An exploration of the polymorphism of piracetam using high pressure. *CrystEngComm*, **7**, 179-186.

Fabian, L. (2009). Cambridge structural database analysis of molecular complementarity in cocrystals. *Crystal Growth & Design*, **9**, 1436-1443.

Farrugia, L. (1999). WinGX suite for small-molecule single-crystal crystallography. *Journal of Applied Crystallography*, **32**, 837-838.

Farrugia, L. (2012). WinGX and ORTEP for windows: An update. *Journal of Applied Crystallography*, **45**, 849-854.

Ferrari, E. S.; Davey, R. J. (2004). Solution-mediated transformation of  $\alpha$  to  $\beta$  l-glutamic acid: Rate enhancement due to secondary nucleation. *Crystal Growth & Design*, **4**, 1061-1068.

Ferré, L. (1995). Selection of components in principal component analysis: A comparison of methods. *Computational Statistics & Data Analysis*, **19**, 669-682.

Fillers, J. P.; Hawkinson, S. W. (1982a). The structure of 4-(2-chlorodibenz[b,f][1,4]oxazepin-11-yl)-1-methyl-1H-piperazinium succinate monohydrate (loxapine succinate monohydrate). *Acta Crystallographica Section B*, **38**, 3041-3045.

Fillers, J. P.; Hawkinson, S. W. (1982b). The structure of 8-chloro-11-(4-methyl-1-piperazinyl)-5H-dibenzo[b,e][1,4]diazepine dihydrobromide, clozapine dihydrobromide. *Acta Crystallographica Section B*, **38**, 1750-1753.

Findlay, W. P.; Bugay, D. E. (1998). Utilization of Fourier transform-Raman spectroscopy for the study of pharmaceutical crystal forms. *Journal of Pharmaceutical and Biomedical Analysis*, **16**, 921-930.

Florence, A. J. (2009a). The solid state. Modern pharmaceuticals basic principles and systems. A. T. Florence and J. Siepmann, Informa Healthcare, New York, **1**, 253-310.

Florence, A. J. (2009b). Approaches to high-throughput physical form screening and discovery. Polymorphism in pharmaceutical solids. H. G. Brittain, Informa Healthcare, New York, **192**, 139-184.

Florence, A. J.; Bardin, J.; Johnston, B.; Shankland, N.; Griffin, T. A. N.; Shankland, K. (2009). Structure determination from powder data: Mogul and CASTEP. *Zeitschrift Fur Kristallographie*, 215-220.

Florence, A. J.; Baumgartner, B.; Weston, C.; Shankland, N.; Kennedy, A. R.; Shankland, K.; David, W. I. F. (2003). Indexing powder patterns in physical form screening: Instrumentation and data quality. *Journal of Pharmaceutical Sciences*, **92**, 1930-1938.

Florence, A. J.; Bedford, C. T.; Fabbiani, F. P. A.; Shankland, K.; Gelbrich, T.; Hursthouse, M. B.; Shankland, N.; Johnston, A.; Fernandes, P. (2008). Two-dimensional similarity between forms I and II of cytenamide, a carbamazepine analogue. *CrystEngComm*, **10**, 811-813.

Florence, A. J.; Johnston, A.; Price, S. L.; Nowell, H.; Kennedy, A. R.; Shankland, N. (2006). An automated parallel crystallisation search for predicted crystal structures and packing motifs of carbamazepine. *Journal of Pharmaceutical Sciences*, **95**, 1918-1930.

Florence, A. J.; Shankland, N.; Shankland, K.; David, W. I. F.; Pidcock, E.; Xu, X.; Johnston, A.; Kennedy, A. R.; Cox, P. J.; Evans, J. S. O.; Steele, G.; Cosgrove, S. D.; Frampton, C. S. (2005). Solving molecular crystal structures from laboratory X-ray powder diffraction data with DASH: The state of the art and challenges. *Journal of Applied Crystallography*, **38**, 249-259.

Frank, A. (2002). The cambridge structural database: A quarter of a million crystal structures and rising. *Acta Crystallographica Section B*, **58**, 380-388.

Fulton, B.; Goa, K. L. (1997). Olanzapine: A review of its pharmacological properties and therapeutic efficacy in the management of schizophrenia and related psychoses. *Drugs*, **53**, 281-298.

Gardner, C. R.; Walsh, C. T.; Almarsson, O. (2004). Drugs as materials: Valuing physical form in drug discovery. *Nature Review Drug Discovery*, **3**, 926-934.

GAUSSIAN03W (1995). Copyright 1995-2004, Gaussian Inc., 340 Quinnipiac St Bldg 40, Wallingford, CT 06492, USA.

Gavezzotti, A. (2002). Calculation of intermolecular interaction energies by direct numerical integration over electron densities. I. Electrostatic and polarization energies in molecular crystals. *The Journal of Physical Chemistry B*, **106**, 4145-4154.

Gavezzotti, A. (2003a). Calculation of intermolecular interaction energies by direct numerical integration over electron densities. 2. An improved polarization model and the evaluation of dispersion and repulsion energies. *The Journal of Physical Chemistry B*, **107**, 2344-2353.

Gavezzotti, A. (2003b). Towards a realistic model for the quantitative evaluation of intermolecular potentials and for the rationalization of organic crystal structures. Part I. Philosophy. *CrystEngComm*, **5**, 429-438.

Gavezzotti, A. (2003c). Towards a realistic model for the quantitative evaluation of intermolecular potentials and for the rationalization of organic crystal structures. Part II. Crystal energy landscapes. *CrystEngComm*, **5**, 439-446.

Gavezzotti, A. (2005). Hierarchies of intermolecular potentials and forces: progress towards a quantitative evaluation. *Structural Chemistry*, **16**, 177-185.

Gelbrich, T.; Hursthouse, M. B. (2005). A versatile procedure for the identification, description and quantification of structural similarity in molecular crystals. *CrystEngComm*, **7**, 324-336.

Gelbrich, T.; Hursthouse, M. B. (2006). Systematic investigation of the relationships between 25 crystal structures containing the carbamazepine molecule or a close analogue: A case study of the XPac method. *CrystEngComm*, **8**, 449-461.

Getsoian, A.; Lodaya, R. M.; Blackburn, A. C. (2008). One-solvent polymorph screen of carbamazepine. *International Journal of Pharmaceutics*, **348**, 3-9.

Gillon, A. L.; Feeder, N.; Davey, R. J.; Storey, R. (2003). Hydration in molecular crystals: a Cambridge structural database analysis. *Crystal Growth & Design*, **3**, 663-673.

Good, A. C.; Hermsmeier, M. A. (2006). Measuring CAMD technique performance. How "druglike" are drugs? Implications of random test set selection exemplified using druglikeness classification models. *Journal of Chemical Information and Modeling*, **47**, 110-114.

- Gould, P. L. (1986). Salt selection for basic drugs. *International Journal of Pharmaceutics*, **33**, 201-217.
- Gracin, S.; Uusi-Penttilä, M.; Rasmuson, Å. C. (2005). Influence of ultrasound on the nucleation of polymorphs of *p*-aminobenzoic acid. *Crystal Growth & Design*, **5**, 1787-1794.
- Greenbla, E.; Osterber, A. (1968). Pharmacodynamic actions of 2-chloro-11-(1-piperazinyl)dibenz [b,f] 1,4 oxazepine: A new psychoactive agent. *Federation Proceedings*, **27**, 438.
- Griffin, T. A. N.; Shankland, K.; van de Streek, J.; Cole, J. (2009). MDASH: A multi-core-enabled program for structure solution from powder diffraction data. *Journal of Applied Crystallography*, **42**, 360-361.
- Grzesiak, A. L.; Lang, M.; Kim, K.; Matzger, A. J. (2003). Comparison of the four anhydrous polymorphs of carbamazepine and the crystal structure of form I. *Journal of Pharmaceutical Sciences*, **92**, 2260-2271.
- Grzesiak, A. L.; Uribe, F. J.; Ockwig, N. W.; Yaghi, O. M.; Matzger, A. J. (2006). Polymer-induced heteronucleation for the discovery of new extended solids. *Angewandte Chemie International Edition*, **45**, 2553-2556.
- Gu, C.-H.; Li, H.; Gandhi, R. B.; Raghavan, K. (2004). Grouping solvents by statistical analysis of solvent property parameters: Implication to polymorph screening. *International Journal of Pharmaceutics*, **283**, 117-125.
- Guillory, J. K. (1999). Generation of polymorphs, hydrates, solvates and amorphous solids. *Polymorphism in Pharmaceutical Solids*. H. G. Brittain, Marcel Dekker Inc., New York, 183-226.
- Ha, J.-M.; Wolf, J. H.; Hillmyer, M. A.; Ward, M. D. (2004). Polymorph selectivity under nanoscopic confinement. *Journal of the American Chemical Society*, **126**, 3382-3383.
- Hamilton, B. D.; Ha, J.-M.; Hillmyer, M. A.; Ward, M. D. (2011). Manipulating crystal growth and polymorphism by confinement in nanoscale crystallization chambers. *Accounts of Chemical Research*, **45**, 414-423.
- Hilfiker, R.; Berghausen, J.; Blatter, F.; Burkhard, A.; De Paul, S. M.; Freiermuth, B.; Geoffroy, A.; Hofmeier, U.; Marcolli, C.; Siebenhaar, B.; Szlagiewicz, M.; Vit, A.; von Raumer, M. (2003). Polymorphism - Integrated approach from high-throughput screening to crystallization optimization. *Journal of Thermal Analysis and Calorimetry*, **73**, 429-440.
- Hiremath, R.; Basile, J. A.; Varney, S. W.; Swift, J. A. (2005). Controlling molecular crystal polymorphism with self-assembled monolayer templates. *Journal of the American Chemical Society*, **127**, 18321-18327.



Hiremath, R.; Varney, S. W.; Swift, J. A. (2004). Selective growth of a less stable polymorph of 2-iodo-4-nitroaniline on a self-assembled monolayer template. *Chemical Communications*, 2676-2677.

Holden, J. R.; Du, Z.; Ammon, H. L. (1993). Prediction of possible crystal structures for C-, H-, N-, O-, and F- containing organic compounds. *Journal of Computational Chemistry*, **14**, 422-437.

Huang, L.-F.; Tong, W.-Q. (2004). Impact of solid state properties on developability assessment of drug candidates. *Advanced Drug Delivery Reviews*, **56**, 321-334.

Huang, X.; Pan, W.; Grindle, S.; Han, X.; Chen, Y.; Park, S.; Miller, L.; Hall, J. (2005). A comparative study of discriminating human heart failure etiology using gene expression profiles. *BMC Bioinformatics*, **6**, 1-15.

Hughes, L. D.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. O. (2008). Why are some properties more difficult to predict than others? A study of QSPR models of solubility, melting point, and log P. *Journal of Chemical Information and Modeling*, **48**, 220-232.

Hulme, A. T.; Johnston, A.; Florence, A. J.; Fernandes, P.; Shankland, K.; Bedford, C. T.; Welch, G. W. A.; Sadiq, G.; Haynes, D. A.; Motherwell, W. D. S.; Tocher, D. A.; Price, S. L. (2007). Search for a predicted hydrogen bonding motif – A multidisciplinary investigation into the polymorphism of 3-azabicyclo[3.3.1]nonane-2,4-dione. *Journal of the American Chemical Society*, **129**, 3649-3657.

Hursthouse, M. (2004). High-throughput chemical crystallography (HTCC): Meeting and greeting the combichem challenge. *Crystallography Reviews*, **10**, 85-96.

Hursthouse, M. B.; Huth, L. S.; Threlfall, T. L. (2009). Why Do Organic Compounds Crystallise Well or Badly or Ever so Slowly? Why Is Crystallisation Nevertheless Such a Good Purification Technique? *Organic Process Research & Development*, **13**, 1231-1240.

Ibberson, R. M.; Telling, M. T. F.; Parsons, S. (2007). Crystal structures and glassy phase transition behavior of cyclohexene. *Crystal Growth & Design*, **8**, 512-518.

Infantes, L.; Chisholm, J.; Motherwell, S. (2003). Extended motifs from water and chemical functional groups in organic molecular crystals. *CrystEngComm*, **5**, 480-486.

Infantes, L.; Fabian, L.; Motherwell, W. D. S. (2007). Organic crystal hydrates: What are the important factors for formation. *CrystEngComm*, **9**, 65-71.

- Jablensky, A.; Sartorius, N.; Korten, A.; Ernberg, G.; Anker, M.; Cooper, J. E.; Day, R. (1987). Incidence worldwide of schizophrenia. *British Journal of Psychiatry*, **151**, 408-409.
- Jahansouz, H.; Thompson, K. C.; Brenner, G. S.; Kaufman, M. J. (1999). Investigation of the polymorphism of the angiotensin II antagonist agent MK-996. *Pharmaceutical Development and Technology*, **4**, 181-187.
- Johnston, A.; Bardin, J.; Johnston, B. F.; Fernandes, P.; Kennedy, A. R.; Price, S. L.; Florence, A. J. (2011). Experimental and predicted crystal energy landscapes of chlorothiazide. *Crystal Growth & Design*, **11**, 405-413.
- Johnston, A.; Florence, A. J.; Shankland, N.; Kennedy, A. R.; Shankland, K.; Price, S. L. (2007). Crystallization and crystal energy landscape of hydrochlorothiazide. *Crystal Growth & Design*, **7**, 705-712.
- Johnston, A.; Johnston, B. F.; Kennedy, A. R.; Florence, A. J. (2008). Targeted crystallisation of novel carbamazepine solvates based on a retrospective Random Forest classification. *CrystEngComm*, **10**, 23-25.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, **20**, 141-151.
- Kandaswamy, K. K.; Pugalenti, G.; Suganthan, P. N.; Gangal, R. (2010). SVMCRY: An SVM approach for the prediction of protein crystallization propensity from protein sequence. *Protein and Peptide Letters*, **17**, 423-430.
- Karamertzanis, P. G.; Pantelides, C. C. (2007). Ab initio crystal structure prediction. II. Flexible molecules. *Molecular Physics*, **105**, 273-291.
- Kastanos, E.; Kyriakides, A.; Hadjigeorgiou, K.; Pitris, C. (2012). A novel method for bacterial UTI diagnosis using Raman spectroscopy. *International Journal of Spectroscopy*, **2012**, 13.
- Kauffman, G. W.; Jurs, P. C. (2001). QSAR and k-nearest neighbor classification analysis of selective cyclooxygenase-2 inhibitors using topologically-based numerical descriptors. *Journal of Chemical Information and Computer Sciences*, **41**, 1553-1560.
- Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C. (2011a). Efficient handling of molecular flexibility in lattice energy minimization of organic crystals. *Journal of Chemical Theory and Computation*, **7**, 1998-2016.
- 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. (2011b). Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction. *International Journal of Pharmaceutics*, **418**, 168-178.

- Kazantsev, A. V.; Karamertzanis, P. G.; Pantelides, C. C.; Adjiman, C. S. (2011c). CrystalOptimizer: An efficient algorithm for lattice energy minimization of organic crystals using isolated-molecule quantum mechanical calculations. *Process Systems Engineering*, Wiley-VCH Verlag GmbH & Co. KGaA, Germany, 1-42.
- Khoshkhoo, S.; Anwar, J. (1993). Crystallization of polymorphs: the effect of solvent. *Journal of Physics D: Applied Physics*, **26**, B90.
- Kitamura, M. (1989). Polymorphism in the crystallization of L-glutamic acid. *Journal of Crystal Growth*, **96**, 541-546.
- Kitamura, M. (2003). Control of polymorphism in crystallization of amino acid. *Developments in Chemical Engineering and Mineral Processing*, **11**, 579-602.
- Kitamura, M. (2005). Thermodynamic stability and transformation of pharmaceutical polymorphs. *Pure and Applied Chemistry*, **77**, 581-591.
- Kitamura, M.; Nakamura, K. (2002). Effects of solvent composition and temperature on polymorphism and crystallization behavior of thiazole-derivative. *Journal of Crystal Growth*, **236**, 676-686.
- Kojima, T.; Onoue, S.; Murase, N.; Katoh, F.; Mano, T.; Matsuda, Y. (2006). Crystalline form information from multiwell plate salt screening by use of Raman microscopy. *Pharmaceutical Research*, **23**, 806-812.
- Kojima, T.; Tsutsumi, S.; Yamamoto, K.; Ikeda, Y.; Moriwaki, T. (2010). High-throughput cocrystal slurry screening by use of in situ Raman microscopy and multi-well plate. *International Journal of Pharmaceutics*, **399**, 52-59.
- Koradia, V.; Fontelonga de Lemos, A. F.; Allesø, M.; Lopez de Diego, H.; Ringkjøbing-Elema, M.; Müllertz, A.; Rantanen, J. (2011). Phase transformations of amlodipine besylate solid forms. *Journal of Pharmaceutical Sciences*, **100**, 2896-2910.
- Kourti, T. (2009). Quality by design in the pharmaceutical industry: Process modelling, monitoring and control using latent variable method. 7<sup>th</sup> IFAC International Symposium on Advanced Control of Chemical Processes, Koç University Campus, Turkey, **7**, 36-41.
- Kurgan, L.; Razib, A.; Aghakhani, S.; Dick, S.; Mizianty, M.; Jahandideh, S. (2009). CRYSTALP2: Sequence-based protein crystallization propensity prediction. *BMC Structural Biology*, **9**, 50.
- Latimer, C. N. (1969). Neuropharmacologic evaluation of oxilapine a potent psychoactive agent. *Journal of Pharmacology and Experimental Therapeutics*, **166**, 151-162.

- Laudise, R. A. (1970). Growth from solution. Solid state physical electronic series. J. Holonyak, N., Prentice-Hall, Inc., USA, 39-73.
- Li, S.; Fedorowicz, A.; Singh, H.; Soderholm, S. C. (2005). Application of the random forest method in studies of local lymph node assay based skin sensitization data. *Journal of Chemical Information and Modeling*, **45**, 952-964.
- Liaw, A.; Wiener, M. (2002). Classification and regression by randomForest. *R News*, **2**, 18-22.
- Llinas, A.; Box, K. J.; Burley, J. C.; Glen, R. C.; Goodman, J. M. (2007). A new method for the reproducible generation of polymorphs: two forms of sulindac with very different solubilities. *Journal of Applied Crystallography*, **40**, 379-381.
- Llinàs, A.; Goodman, J. M. (2008). Polymorph control: Past, present and future. *Drug Discovery Today*, **13**, 198-210.
- López-Mejías, V.; Kampf, J. W.; Matzger, A. J. (2012). Nonamorphism in flufenamic acid and a new record for a polymorphic compound with solved structures. *Journal of the American Chemical Society*, **134**, 9872-9875.
- Lowry, S.; Dalrymple, D.; Song, A.; Rosso, V.; Pommier, C.; Venit, J. (2006). Integrating a Raman microscope into the workflow of a high-throughput crystallization laboratory. *Journal of the Association for Laboratory Automation*, **11**, 75-84.
- Lu, J.; Wang, X. J.; Yang, X.; Ching, C. B. (2007). Characterization and selective crystallization of famotidine polymorphs. *Journal of Pharmaceutical Sciences*, **96**, 2457-2468.
- Lunetta, K.; Hayward, L. B.; Segal, J.; Van Eerdewegh, P. (2004). Screening large-scale association study data: Exploiting interactions using random forests. *BMC Genetics*, **5**, 32.
- Lutker, K. M.; Matzger, A. J. (2010). Crystal polymorphism in a carbamazepine derivative: Oxcarbazepine. *Journal of Pharmaceutical Sciences*, **99**, 794-803.
- Lutker, K. M.; Tolstyka, Z. P.; Matzger, A. J. (2007). Investigation of a privileged polymorphic motif: A dimeric ROY derivative. *Crystal Growth & Design*, **8**, 136-139.
- Macrae, C. F.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Shields, G. P.; Taylor, R.; Towler, M.; van de Streek, J. (2006). Mercury: Visualization and analysis of crystal structures. *Journal of Applied Crystallography*, **39**, 453-457.
- Makretsov, N. A.; Huntsman, D. G.; Nielsen, T. O.; Yorida, E.; Peacock, M.; Cheang, M. C. U.; Dunn, S. E.; Hayes, M.; van de Rijn, M.; Bajdik, C.; Gilks, C. B. (2004). Hierarchical clustering analysis of tissue microarray immunostaining data

identifies prognostically significant groups of breast carcinoma. *Clinical Cancer Research*, **10**, 6143-6151.

Marcus, Y. (1991). Thermodynamics of solvation of ions. Part 5.-Gibbs free energy of hydration at 298.15 K. *Journal of the Chemical Society, Faraday Transactions*, **87**, 2995-2999.

Matousek, P.; Towrie, M.; Parker, A. W. (2002). Fluorescence background suppression in Raman spectroscopy using combined Kerr gated and shifted excitation Raman difference techniques. *Journal of Raman Spectroscopy*, **33**, 238-242.

McCabe, J. F. (2010). Application of design of experiment (DOE) to polymorph screening and subsequent data analysis. *CrystEngComm*, **12**, 1110-1119.

McCrone, W. C. (1965). Physics and chemistry of the organic solid state. D. Fox, M. M. Labes and A. Weissberger, Wiley Interscience, New York, **2**, 725-767.

McKellar, S. C.; Urquhart, A. J.; Lamprou, D. A.; Florence, A. J. (2012). Polymer templating of supercooled indomethacin for polymorph selection. *ACS Combinatorial Science*, **14**, 155-159.

McKinnon, J. J.; Spackman, M. A.; Mitchell, A. S. (2004). Novel tools for visualizing and exploring intermolecular interactions in molecular crystals. *Acta Crystallographica Section B*, **60**, 627-668.

Miglani, R.; Miller, G.; Oswald, I.; Florence, A. J. (2011). Physical form screening of olanzapine and amoxapine. *Acta Crystallographica Section A*, **67**, C569.

Miller, J. M.; Collman, B. M.; Greene, L. R.; Grant, D. J. W.; Blackburn, A. C. (2005). Identifying the stable polymorph early in the drug discovery–development process. *Pharmaceutical Development and Technology*, **10**, 291-297.

Mizianty, M. J.; Kurgan, L. (2009). Meta prediction of protein crystallization propensity. *Biochemical and Biophysical Research Communications*, **390**, 10-15.

Mizianty, M. J.; Kurgan, L. (2011). Sequence-based prediction of protein crystallization, purification and production propensity. *Bioinformatics*, **27**, i24-i33.

MOE (2002). Chemical Computing Group, 1010 Sherbrooke St. W, Montreal, Quebec, H3A 2R7. Qubec, Canada.

Morissette, S. L.; Almarsson, Ö.; Peterson, M. L.; Remenar, J. F.; Read, M. J.; Lemmo, A. V.; Ellis, S.; Cima, M. J.; Gardner, C. R. (2004). High-throughput crystallization: Polymorphs, salts, co-crystals and solvates of pharmaceutical solids. *Advanced Drug Delivery Reviews*, **56**, 275-300.

- Morissette, S. L.; Soukasene, S.; Levinson, D.; Cima, M. J.; Almarsson, Ö. (2003). Elucidation of crystal form diversity of the HIV protease inhibitor ritonavir by high-throughput crystallization. *Proceedings of the National Academy of Sciences*, **100**, 2180-2184.
- Mullin, J. W. (1971). Industrial crystallisation processes. *Crystallization*, Hazell, Watson & Viney Ltd., England, **2**, 258-293.
- Mullin, J. W. (2001). Industrial techniques and equipment. *Crystallization*, Butterworth-Heinemann, Oxford, 315-402.
- Murnane, D.; Marriott, C.; Martin, G. P. (2008). Polymorphic control of inhalation microparticles prepared by crystallization. *International Journal of Pharmaceutics*, **361**, 141-149.
- Mytkolli, H.; Calitoiu, D. (2009). Statistical modelling Using SAS. A short course. From [http://www.oasus.ca/modeling\\_part1.pdf](http://www.oasus.ca/modeling_part1.pdf).
- Nangia, A. (2008). Conformational polymorphism in organic crystals. *Accounts of Chemical Research*, **41**, 595-604.
- Neumann, M. A. (2011). GRACE. Avant-garde Materials Simulation GmbH, Germany. From <http://www.avmatsim.eu>.
- Neumann, M. A.; Leusen, F. J. J.; Kendrick, J. (2008). A major advance in crystal structure prediction. *Angewandte Chemie International Edition*, **47**, 2427-2430.
- Neumann, M. A.; Perrin, M.-A. (2005). Energy ranking of molecular crystals using density functional theory calculations and an empirical van der waals correction. *The Journal of Physical Chemistry B*, **109**, 15531-15541.
- Nicolai, B.; Espeau, P.; Céolin, R.; Perrin, M. A.; Zinke, L.; Giovannini, J.; Leveiller, F. (2007). Polymorph formation from solvate desolvation. *Journal of Thermal Analysis and Calorimetry*, **90**, 337-339.
- Norris, K. H.; Williams, P. C. (1984). Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat .1. Influence of particle-size. *Cereal Chemistry*, **61**, 158-165.
- Nuria Sanchez-Puig; Claude Sauter; Bernard Lorber; Richard Giege; Moreno, A. (2012). Predicting protein crystallizability and nucleation. *Protein & Peptide Letters*, **19**, 725-731.
- O'Connell, M.-L.; Howley, T.; Ryder, A. G.; Leger, M. N.; Madden, M. G. (2005). Classification of a target analyte in solid mixtures using principal component analysis, support vector machines, and Raman spectroscopy. *SPIE Proceedings*, **5826**, 340-350.

Overton, I. M.; Barton, G. J. (2006). A normalised scale for structural genomics target ranking: The OB-Score. *FEBS Letters*, **580**, 4005-4009.

Overton, I. M.; Padovani, G.; Girolami, M. A.; Barton, G. J. (2008). ParCrys: a Parzen window density estimation approach to protein crystallization propensity prediction. *Bioinformatics*, **24**, 901-907.

Palmer, D. S.; O'Boyle, N. M.; Glen, R. C.; Mitchell, J. B. O. (2006). Random forest models to predict aqueous solubility. *Journal of Chemical Information and Modeling*, **47**, 150-158.

Parveen, S.; Davey, R. J.; Dent, G.; Pritchard, R. G. (2005). Linking solution chemistry to crystal nucleation: The case of tetrolic acid. *Chemical Communications*, 1531-1533.

Pawley, G. (1981). Unit-cell refinement from powder diffraction scans. *Journal of Applied Crystallography*, **14**, 357-361.

Petcher, T. J.; Weber, H.-P. (1976). Conformations of some semi-rigid neuroleptic drugs. Part 1. Crystal structures of loxapine, clozapine, and HUF-2046 monohydrate {2-chloro-11-(4-methylpiperazin-1-yl)dibenzo[b,f][1,4]oxazepine, 8-chloro-11-(4-methylpiperazin-1-yl)dibenzo[b,e][1,4]diazepine, and 2-chloro-11(4-methylpiperazin-1-yl)dibenzo[b,e][1,4]diazepine monohydrate}. *Journal of the Chemical Society, Perkin Transactions 2*, 1415-1420.

Peterson, M. L.; Morissette, S. L.; McNulty, C.; Goldsweig, A.; Shaw, P.; LeQuesne, M.; Monagle, J.; Encina, N.; Marchionna, J.; Johnson, A.; Gonzalez-Zugasti, J.; Lemmo, A. V.; Ellis, S. J.; Cima, M. J.; Almarsson, Ö. (2002). Iterative high-throughput polymorphism studies on acetaminophen and an experimentally derived structure for form III. *Journal of the American Chemical Society*, **124**, 10958-10959.

Pfeiffer, R. R.; Yang, K. S.; Tucker, M. A. (1970). Crystal pseudopolymorphism of cephaloglycin and cephalixin. *Journal of Pharmaceutical Sciences*, **59**, 1809-1814.

Pillard, J.; Arnautova, Y. A.; Czaplewski, C.; Gibson, K. D.; Scheraga, H. A. (2001). Conformation-family Monte Carlo: A new method for crystal structure prediction. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 12351-12356.

Polla, G. I.; Vega, D. R.; Lanza, H.; Tombari, D. G.; Baggio, R.; Ayala, A. P.; Filho, J. M.; Fernández, D.; Leyva, G.; Dartayet, G. (2005). Thermal behaviour and stability in Olanzapine. *International Journal of Pharmaceutics*, **301**, 33-40.

Pommier, C. J. S. R., V.; Song, A. (2005). Challenges in data analysis of Raman spectra from crystallization in 96 wellplates. *American Pharmaceutical Review*, **8**, 19-24.

Price, C. P.; Glick, G. D.; Matzger, A. J. (2006). Dissecting the Behavior of a Promiscuous Solvate Former. *Angewandte Chemie International Edition*, **45**, 2062-2066.

Price, C. P.; Grzesiak, A. L.; Matzger, A. J. (2005). Crystalline polymorph selection and discovery with polymer heteronuclei. *Journal of the American Chemical Society*, **127**, 5512-5517.

Price, S. (2013). Why don't we find more polymorphs? *Acta Crystallographica Section B*, **69**, 313-328.

Price, S. L. (2004). The computational prediction of pharmaceutical crystal structures and polymorphism. *Advanced Drug Delivery Reviews*, **56**, 301-319.

Price, S. L. (2008a). Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism. *Accounts of Chemical Research*, **42**, 117-126.

Price, S. L. (2008b). From crystal structure prediction to polymorph prediction: interpreting the crystal energy landscape. *Physical Chemistry Chemical Physics*, **10**, 1996-2009.

Price, S. L. (2009). Computational methodologies: toward crystal structure and polymorph prediction. Polymorphism in pharmaceutical solids. H. G. Brittain, Informa Healthcare, New York, **192**, 52-75.

Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. (2010). Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Physical Chemistry Chemical Physics*, **12**, 8478-8490.

Qi, Y.; Bar-Joseph, Z.; Klein-Seetharaman, J. (2006). Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins: Structure, Function, and Bioinformatics*, **63**, 490-500.

R Development Core Team (2006). R: A language and environment for statistical computing, Version 2.10.1 and 2.11.1, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

R. Barbas; Prohens, R.; Puigjaner, C. (2007). A new polymorph of norfloxacin complete characterization and relative stability of its trimorphic system. *Journal of Thermal Analysis and Calorimetry*, **89**, 687-692.

Remenar, J. F.; MacPhee, J. M.; Larson, B. K.; Tyagi, V. A.; Ho, J. H.; McIlroy, D. A.; Hickey, M. B.; Shaw, P. B.; Almarsson, Ö. (2003). Salt selection and simultaneous polymorphism assessment via high-throughput crystallization: The case of sertraline. *Organic Process Research & Development*, **7**, 990-996.



Reutzel-Edens, S. M. (2006). Achieving polymorph selectivity in the crystallization of pharmaceutical solids: Basic considerations and recent advances. *Curr Opin Drug Discov Devel*, **9**, 806-815.

Reutzel-Edens, S. M.; Bush, J. K.; Magee, P. A.; Stephenson, G. A.; Byrn, S. R. (2003). Anhydrates and hydrates of olanzapine: Crystallization, solid-state characterization, and structural relationships. *Crystal Growth & Design*, **3**, 897-907.

Rietveld, H. (1969). A profile refinement method for nuclear and magnetic structures. *Journal of Applied Crystallography*, **2**, 65-71.

Rodríguez-hornedo, N.; Murphy, D. (1999). Significance of controlling crystallization mechanisms and kinetics in pharmaceutical systems. *Journal of Pharmaceutical Sciences*, **88**, 651-660.

Rodríguez-Spong, B.; Price, C. P.; Jayasankar, A.; Matzger, A. J.; Rodríguez-Hornedo, N. r. (2004). General principles of pharmaceutical solid polymorphism: A supramolecular perspective. *Advanced Drug Delivery Reviews*, **56**, 241-274.

Romero-Torres, S.; Huang, J.; Hernandez-abad, P. E. (2009). Practical considerations in PAT analyzer selection: NIR vs. Raman. American Pharmaceutical Review. From <http://www.americanpharmaceuticalreview.com/Featured-Articles/117780-Practical-Considerations-on-PAT-Analyzer-Selection-Raman-vs-NIR-Spectroscopy/>.

Rose, S. (2002). Statistical design and application to combinatorial chemistry. *Drug Discovery Today*, **7**, 133-138.

Roy, S.; Quiñones, R.; Matzger, A. J. (2012). Structural and physicochemical aspects of Dasatinib hydrate and anhydrate phases. *Crystal Growth & Design*, **12**, 2122-2126.

Sanger, T. M.; Grundy, S. L.; Gibson, P. J.; Namjoshi, M. A.; Greaney, M. G.; Tohen, M. F. (2001). Long-term olanzapine therapy in the treatment of bipolar I disorder: An open-label continuation phase study. *Journal of Clinical Psychiatry*, **62**, 273-281.

Sarma, J. A. R. P.; Desiraju, G. R. (2002). The supramolecular synthon approach to crystal structure prediction. *Crystal Growth & Design*, **2**, 93-100.

Sheldrick, G. (2008). A short history of SHELX. *Acta Crystallographica Section A*, **64**, 112-122.

Sheridan, R.; Nachbar, R.; Bush, B. (1994). Extending the trend vector: The trend matrix and sample-based partial least squares. *Journal of Computer-Aided Molecular Design*, **8**, 323-340.

SIMCA (2012). Multivariate analysis software, version 13.0.0.0., Umetrics Ltd. MKS Instruments UK Ltd., Unit 3-4, Cowley Way, Weston Road, Crewe, Cheshire, CW1 6AG, U. K.

Siva Lakshmi Devi; Prasunamba P. L.; Ravikumar K.; B., S. (2011). Solid state characterization of clozapine monohydrate. *Structural Chemistry Communications*, **2**, 101-104.

Slabinski, L.; Jaroszewski, L.; Rychlewski, L.; Wilson, I. A.; Lesley, S. A.; Godzik, A. (2007). XtalPred: A web server for prediction of protein crystallizability. *Bioinformatics*, **23**, 3403-3405.

Smith, E.; Dent, G. (2005a). Introduction, basic theory and Principles, Modern Raman spectroscopy - A practical approach, John Wiley & Sons, Ltd., England, , 1-21.

Smith, E.; Dent, G. (2005b). Resonance Raman scattering. Modern Raman spectroscopy - A practical approach, John Wiley & Sons, Ltd., England, 93-112.

Spek, A. (2003). Single-crystal structure validation with the program PLATON. *Journal of Applied Crystallography*, **36**, 7-13.

Stahly, G. P. (2007). Diversity in single- and multiple-component crystals. The search for and prevalence of polymorphs and cocrystals. *Crystal Growth & Design*, **7**, 1007-1026.

Steyvers, M. (2006). Multidimensional scaling. Encyclopedia of Cognitive Science, John Wiley & Sons, Ltd., England.

Stoica, C.; Verwer, P.; Meeke, H.; Vlieg, E.; van Hoof, P. J. C. M.; Kaspersen, F. (2005). Heterogeneous 2D nucleation of the stable polymorphic form on the metastable form. *Journal of Crystal Growth*, **275**, e1727-e1731.

Storey, R.; Docherty, R.; Higginson, P.; Dallman, C.; Gilmore, C.; Barr, G.; Dong, W. (2004). Automation of solid form screening procedures in the pharmaceutical industry – How to avoid the bottlenecks. *Crystallography Reviews*, **10**, 45-56.

Suh, C.; Gadzuric, S.; Gaune-Escard, M.; Rajan, K. (2009). Multivariate analysis for chemistry-property relationships in molten salts. *Zeitschrift fur Naturforsch A Journal of Physical Sciences*, **64**, 467-476.

Sun, X.; Garetz, B. A.; Myerson, A. S. (2008). Polarization switching of crystal structure in the nonphotochemical laser-induced nucleation of supersaturated aqueous l-histidine. *Crystal Growth & Design*, **8**, 1720-1722.

Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, **43**, 1947-1958.

- Tao, J.; Jones, K. J.; Yu, L. (2007). Cross-nucleation between d-mannitol polymorphs in seeded crystallization. *Crystal Growth & Design*, **7**, 2410-2414.
- Taskinen, J.; Yliruusi, J. (2003). Prediction of physicochemical properties based on neural network modelling. *Advanced Drug Delivery Reviews*, **55**, 1163-1183.
- Thakuria, R.; Nangia, A. (2011). Polymorphic form IV of olanzapine. *Acta Crystallographica Section C*, **67**, o461-o463.
- Threlfall, T. (2000). Crystallisation of polymorphs: Thermodynamic insight into the role of solvent. *Organic Process Research & Development*, **4**, 384-390.
- Tian, F.; Sandler, N.; Aaltonen, J.; Lang, C.; Saville, D. J.; Gordon, K. C.; Strachan, C. J.; Rantanen, J.; Rades, T. (2007). Influence of polymorphic form, morphology, and excipient interactions on the dissolution of carbamazepine compacts. *Journal of Pharmaceutical Sciences*, **96**, 584-594.
- Tiwari, M.; Chawla, G.; Bansal, A. K. (2007). Quantification of olanzapine polymorphs using powder X-ray diffraction technique. *Journal of Pharmaceutical and Biomedical Analysis*, **43**, 865-872.
- Tollefson, G. D.; Beasley, C. M.; Tran, P. V.; Street, J. S.; Krueger, J. A.; Tamura, R. N.; Graffeo, K. A.; Thieme, M. E. (1997). Olanzapine versus haloperidol in the treatment of schizophrenia and schizoaffective and schizophreniform disorders: Results of an international collaborative trial. *American Journal of Psychiatry*, **154**, 457-465.
- Tong, W. Q.; Alva, G.; Carlton, D.; Viscomi, F.; Adkison, K.; Millar, A.; Dhingra, O. (1999). A strategic approach to the salt selection of an insoluble drug candidate. 13<sup>th</sup> AAPS Meeting, New Orleans.
- Trask, A. V.; Motherwell, W. D. S.; Jones, W. (2006). Physical stability enhancement of theophylline via cocrystallization. *International Journal of Pharmaceutics*, **320**, 114-123.
- Tye, H. (2004). Application of statistical 'design of experiments' methods in drug discovery. *Drug Discovery Today*, **9**, 485-491.
- USFDA, C. (2011). Guidance for Industry: Regulatory Classification of Pharmaceutical Co-Crystals. From <http://www.fda.gov/downloads/Drugs/Guidances/UCM281764.pdf>.
- Uzoh, O. G.; Cruz-Cabeza, A. J.; Price, S. L. (2012). Is the fenamate group a polymorphophore? Contrasting the crystal energy landscapes of fenamic and tolfenamic acids. *Crystal Growth & Design*, **12**, 4230-4239.
- van de Streek, J. (2007). All series of multiple solvates (including hydrates) from the Cambridge Structural Database. *CrystEngComm*, **9**, 350-352.

- Vehring, R. (2008). Pharmaceutical particle engineering via spray drying. *Pharmaceutical Research*, **25**, 999-1022.
- Verwer, P.; Leusen, F. J. J. (2007). Computer simulation to predict possible crystal polymorphs. Reviews in computational chemistry, John Wiley & Sons, Inc., New York, 327-365.
- Vidmar, G.; Pohar, M. (2005). Augmented convex hull plots: Rationale, implementation in R and biomedical applications. *Computer Methods and Programs in Biomedicine*, **78**, 69-74.
- Vippagunta, S. R.; Brittain, H. G.; Grant, D. J. W. (2001). Crystalline solids. *Advanced Drug Delivery Reviews*, **48**, 3-26.
- Vishweshwar, P.; McMahon, J. A.; Oliveira, M.; Peterson, M. L.; Zaworotko, M. J. (2005). The predictably elusive form II of aspirin. *Journal of the American Chemical Society*, **127**, 16802-16803.
- Wawrzycka-Gorczyca, I.; Borowski, P.; Osypiuk-Tomasik, J.; Mazur, L.; Koziol, A. E. (2007). Crystal structure of olanzapine and its solvates. Part 3. Two and three-component solvates with water, ethanol, butan-2-ol and dichloromethane. *Journal of Molecular Structure*, **830**, 188-197.
- Wawrzycka-Gorczyca, I.; Koziol, A. E.; Glice, M.; Cybulski, J. (2004a). Polymorphic form II of 2-methyl-4-(4-methyl-1-piperazinyl)-10H-thieno[2,3-b][1,5]benzodiazepine. *Acta Crystallographica Section E*, **60**, o66-o68.
- Wawrzycka-Gorczyca, I.; Mazur, L.; Koziol, A. E. (2004b). 2-Methyl-4-(4-methyl-1-piperazinyl)-10H-thieno[2,3-b][1,5]benzodiazepine methanol solvate. *Acta Crystallographica Section E*, **60**, o69-o71.
- Wiklund, S. (2008). Multivariate data analysis for Omics. Umetrics Ltd. An MKS Company. From [http://www.metabolomics.se/Courses/MVA/MVA%20in%20Omics Handouts Exercises Solutions Thu-Fri.pdf](http://www.metabolomics.se/Courses/MVA/MVA%20in%20Omics%20Handouts%20Exercises%20Solutions%20Thu-Fri.pdf).
- Wold, S.; Esbensen, K.; Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **2**, 37-52.
- Wood, P. A.; Francis, D.; Marshall, W. G.; Moggach, S. A.; Parsons, S.; Pidcock, E.; Rohl, A. L. (2008). A study of the high-pressure polymorphs of L-serine using ab initio structures and PIXEL calculations. *CrystEngComm*, **10**, 1154-1166.
- Wood, P. A.; Oliveira, M. A.; Zink, A.; Hickey, M. B. (2012). Isostructurality in pharmaceutical salts: How often and how similar? *CrystEngComm*, **14**, 2413-2421.
- Xu, D.; Redman-Furey, N. (2007). Statistical cluster analysis of pharmaceutical solvents. *International Journal of Pharmaceutics*, **339**, 175-188.

Yang, X.; Sarma, B.; Myerson, A. S. (2012). Polymorph Control of Micro/Nano-Sized Mefenamic Acid Crystals on Patterned Self-Assembled Monolayer Islands. *Crystal Growth & Design*, **12**, 5521-5528.

Yu, L.; Ng, K. (2002). Glycine crystallization during spray drying: The pH Effect on Salt and Polymorphic Forms. *Journal of Pharmaceutical Sciences*, **91**, 2367-2375.

Yu, L.; Reutzel-Edens, S. M.; Mitchell, C. A. (2000). Crystallization and polymorphism of conformationally flexible molecules: Problems, patterns, and strategies. *Organic Process Research & Development*, **4**, 396-402.

Zeidan, T. A.; Trotta, J. T.; Chiarella, R. A.; Oliveira, M. A.; Hickey, M. B.; Almarsson, Ö.; Remenar, J. F. (2013). Polymorphism of Dehydro-aripiprazole, the active metabolite of the antipsychotic drug aripiprazole (abilify). *Crystal Growth & Design*, **13**, 2036-2046.

Zhang, Q.-Y.; Aires-de-Sousa, J. (2006). Random forest prediction of mutagenicity from empirical physicochemical descriptors. *Journal of Chemical Information and Modeling*, **47**, 1-8.

## **Appendix**

The appendix is available on a disk.

## List of Publications

### 1) Exploring the Experimental and Computed Crystal Energy Landscape of Olanzapine

Rajni M. Bhardwaj, Louise S. Price, Sarah L. Price, Susan M. Reutzel-Edens, Gary J Miller, Iain D. H. Oswald, Blair F. Johnston and Alastair J. Florence, *Cryst. Growth Des.*, 2013, 13, 1602–17.

**Abstract.** An extensive experimental search for solid forms of the antipsychotic compound olanzapine identified 60 distinct solid forms including three non-solvated polymorphs, 56 crystalline solvates and an amorphous phase. XPac analysis of the 35 experimental crystal structures (30 from this work and 5 from the CSD) containing olanzapine show that they contain a specific, dispersion-bound, dimer structure which can adopt various arrangements and accommodate diverse solvents to produce structures with a similar moderate packing efficiency to form I. The crystal energy landscape confirms the inability of olanzapine to pack with an efficiency of more than 70%, explains the role of solvent in stabilizing the solvate structures, and identifies a hypothetical structural type that offers an explanation for the inability to obtain the metastable forms II and III separately. The calculations find that structures that do not contain the observed dimer are thermodynamically feasible, suggesting that kinetic effects are responsible for all the observed structures being based on the dimer. Thus this extensive screen probably has not found all possible physical forms of olanzapine, and further form diversity could be targeted with a better understanding of the role of kinetics in its crystallization.

### 2) A complementary experimental and computational study of loxapine succinate and its monohydrate

Rajni M. Bhardwaj, Blair F. Johnston, Iain Oswald and Alastair J. Florence, *Acta Cryst. C*, 2013, 69, 1273-1278.

The crystal structures of loxapine succinate [systematic name: 4-(2-chlorodibenzo[*b,f*][1,4]oxazepin-11-yl)-1-methylpiperazin-1-ium 3-carboxypropanoate],  $C_{18}H_{19}ClN_3O^+ \cdot C_4H_5O_4^-$ , and loxapine succinate monohydrate {systematic name: bis[4-(2-chlorodibenzo[*b,f*][1,4]oxazepin-11-yl)-1-methylpiperazin-1-ium] succinate succinic acid dihydrate},  $2C_{18}H_{19}ClN_3O^+ \cdot C_4H_4O_4^{2-} \cdot C_4H_6O_4 \cdot 2H_2O$ , have been determined using X-ray powder diffraction and single-crystal X-ray diffraction, respectively. Fixed cell geometry optimization calculations using density functional theory confirmed that the global optimum powder diffraction derived structure also matches an energy minimum structure. The energy calculations proved to be an effective tool in locating the positions of the H atoms reliably and verifying the salt configuration of the structure determined from powder data. Crystal packing analysis of these structures revealed that the loxapine

succinate structure is based on chains of protonated loxapine molecules while the monohydrate contains dispersion stabilized centrosymmetric dimers. Incorporation of water molecules within the crystal lattice significantly alters the molecular packing and protonation state of the succinic acid.

### 3) 2,4-Diamino-6-methyl-1,3,5-triazine 1,4-butanediol solvate

Rajni M. Bhardwaj, Iain Oswald and Alastair J. Florence, *Acta Cryst.* 2012, E68, o3377.

**Abstract.** 2,4-diamino-6-methyl-1,3,5-triazine (DMT, acetoguanamine) forms a 1:1 solvate with 1,4-butanediol ( $C_4H_7N_5 \cdot C_4H_{10}O_2$ ). The compound crystallizes with one molecule of DMT and one of 1,4-butanediol in the asymmetric unit. DMT molecules forms ribbons involving  $R_2^2(8)$  dimer motif between DMT molecule along the direction of the *c*-axis and these ribbons are H-bonded to each other through 1,4-butanediol to form sheets parallel to (121) plane.

### 4) 2-Methyl-4-(4-methylpiperazin-1-yl)-10H-thieno[2,3-*b*][1,5] benzodiazepine (olanzapine) propan-2-ol disolvate

Rajni M. Bhardwaj and Alastair J. Florence, *Acta Cryst.*, 2013, E69, o752-o753

**Abstract.** In the title solvate,  $C_{17}H_{20}N_4S_2C_3H_8O$ , pairs of olanzapine molecules related by a centre of inversion stack along the *a* axis, forming columns, which are packed parallel to each other along the *b* axis, forming a sheet arrangement. The columns within these sheets are hydrogen bonded to each other through the propan-2-ol solvent molecules. The diazepine ring of the olanzapine exists in a puckered conformation with the thiophene and phenyl rings making a dihedral angle of 57.66 (7)° and the piperazine ring adopts a chair conformation with the methyl group in an equatorial position.

### 5) Complex Polymorphic System of Gallic Acid—Five Monohydrates, Three Anhydrates, and over 20 Solvates

Doris E. Braun, Rajni M. Bhardwaj, Alastair J. Florence, Derek A. Tocher, and Sarah L. Price, *Cryst. Growth Des.*, 2013, 13,19–23

**Abstract.** We report the structure of the fifth monohydrate of gallic acid and two additional anhydrate polymorphs and evidence of at least 22 other solvates formed, many containing water and another solvent. This unprecedented number of monohydrate polymorphs and diversity of solid forms is consistent with the anhydrate and monohydrate crystal energy landscapes, showing both a wide range of packing motifs and also some structures differing only in proton positions. By



aiding the solution of structures from powder X-ray diffraction data and guiding the screening, the computational studies help explain the complex polymorphism of gallic acid. This is industrially relevant, as the three anhydrates are stable at ambient conditions but hydration/dehydration behavior is very dependent on relative humidity and phase purity.

**6) The Absorbing a little water: the structural, thermodynamic and kinetic relationship between pyrogallol and its tetarto-hydrate**

Doris E. Braun· Rajni M. Bhardwaj, Jean-Baptiste Arlin, Alastair J. Florence, Volker Kahlenberg, Ulrich J. Griesser, Derek A. Tocher and Sarah L. Price, *Cryst. Growth Des.* 2013, 13, 4071–4083.

**Abstract.** The anhydrate and the stoichiometric tetarto-hydrate of pyrogallol (0.25 mole water per mole pyrogallol) are both storage stable at ambient conditions, provided that they are phase pure, with the system being at equilibrium at  $a_w$  (water activity) = 0.15 at 25 °C. Structures have been derived from single and powder X-ray diffraction data for the anhydrate and hydrate, respectively. It is notable that the tetarto-hydrate forms a tetragonal structure with water in channels, a framework that although stabilized by water, is found as a higher energy structure on a computationally generated crystal energy landscape, which has the anhydrate crystal structure as the most stable form. Thus a combination of slurry experiments, X-ray diffraction, spectroscopy, moisture (de)sorption, and thermo-analytical methods with the computationally generated crystal energy landscape and lattice energy calculations provides a consistent picture of the finely balanced hydration behavior of pyrogallol. In addition two monotropically related dimethyl sulfoxide monosolvates were found in the accompanying solid form screen.