Electronic Thesis and Dissertation Repository

8-8-2023 2:00 PM

Application of Crystal Engineering in Multicomponent Pharmaceutical Crystals: A Study of Theory and Practice

Soroush Ahmadi Nasrabadi, *Western University*

Abstract

Multicomponent crystallization, a prominent strategy in crystal engineering, offers the ability to modify the physicochemical properties of crystals by introducing a secondary component to their lattice structure. Such multicomponent crystals have found widespread application in the pharmaceutical industry. This thesis explores the experimental screening, characterization, application, and theoretical prediction of multicomponent crystals of Active Pharmaceutical Ingredients (APIs).

The first case study investigates a new solvate of Dasatinib which exhibits high instability at room temperature and transforms into a different polymorph upon desolvation. The crystal structure of this compound is obtained, revealing insights into its transient nature and the potential application of desolvation for particle size reduction.

Another case study focuses on synthesizing a new cocrystal of zinc-phenylacetate (Zn-PA) with isonicotinamide (INAM). The resulting Zn-PA-INAM ionic cocrystal resolves the hydrophobicity issue of Zn-PA, enhancing solubility and dissolution rate. The crystal structure of Zn-PA-INAM, lattice energy comparison, and crystal morphology studies provide scientific explanations for these alterations.

Additionally, this thesis proposes computational prediction strategies to discover new multicomponent crystals. Quantitative predictive approaches based on hydrogen bonding strength are investigated, employing DFT-derived electrostatic potential (ESP) maps, hydrogen bond energy (HBE) and propensity (HBP) calculations. We demonstrate the enhanced classification capability achieved by combining HBE and HBP through multivariate logistic regression.

Expanding on cocrystal prediction strategies, we performed DFT calculations for a comprehensive database of 6,388 cocrystals from literature reports of both successful and unsuccessful experimental attempts. The extracted ESP surfaces were utilized to develop robust machine learning models that demonstrated exceptional discriminatory performance and achieved up to 94% accuracy on unseen test data.

Lastly, an investigation is conducted on the crystal morphology of Rufinamide (RUF), utilizing temperature cycling, solvent screening, and additive selection to modify its thread-like morphology into a more isometric shape. The crystal structures of three RUF polymorphs are determined, and a connection between the microscopic structure and the macroscopic morphologies is established through face indexing.

This thesis provides valuable insights into the application and systematic discovery of multicomponent crystals. By combining experimental screening, characterization, and predictive tools, it contributes to advancing the field's understanding and utilization of multicomponent crystals.

Keywords

Crystal engineering, Pharmaceuticals, Multicomponent solid forms, Cocrystal, Crystal structure, Hydrogen bond, Density Functional Theory (DFT), Machine Learning, Virtual screening, Crystal morphology.

Summary for Lay Audience

From common medications like aspirin and paracetamol to a more specialized pill, they are all manufactured with small crystals of an Active Pharmaceutical Ingredient (API). But what if we could make these crystals more effective by adding another safe substance to them while preserving their chemical nature? By adding a secondary safe substance to an API, we can improve important properties like solubility, which affects how well the drug can dissolve in our bodies and reach the targeted areas. In this research, our goal is to discover new multicomponent crystals by testing different combinations of ingredients and studying their properties. We also use chemistry-based knowledge and computer simulations to predict the outcome of different combinations before even making them in the lab.

One interesting finding was with Dasatinib, where we discovered that when crystallized with methanol, the crystals would crack immediately, creating many smaller crystals. This entails particle size reduction that can be advantageous in certain cases. Another discovery we made in the lab was the significance of adding isonicotinamide (INAM) to zinc phenylacetate (Zn-PA), an ammonia-scavenging drug. The crystal we made addressed the issue of Zn-PA being water-repellent, making it dissolve better and work more effectively as a medicine.

But we did not stop at lab work. We also developed computer models to quickly select the secondary compound without having to conduct an extensive experimental search. Using advanced statistical techniques and analyzing the strength of chemical interactions, we were able to accurately predict the formation of new multicomponent crystals.

We also determined the crystal structure of Rufinamide (RUF) in order to investigate the root of its problematically thin crystals. Our goal was to promote the growth of a more symmetrical crystal of RUF. We employed temperature cycling, solvent screening, and additive screening in our study.

This research provides insight into how crystals can be modified to create new materials with improved properties. By combining experimental testing, computer modeling, and predictive tools, we contribute to the advancement of multicomponent crystals, especially in medicine.

Co-Authorship Statement

Chapter 2: Soroush Ahmadi contributed to the experimental characterization, computer simulations, and drafted the manuscript. Pradip Kumar Mondal conducted the crystallographic experiment and helped in preparing the manuscript. Mahmoud Mirmehrabi and Sohrab Rohani were project supervisors who also provided financial support and reviewed the manuscript.

A version of this chapter was published as:

Ahmadi, S., Mondal, P. K., Mirmehrabi, M., & Rohani, S. (2021). Desolvation of dasatinib methanolate: an improved anhydrous polymorph. *CrystEngComm*, *23*(24), 4272–4283.

Chapter 3: Soroush Ahmadi designed the analysis, collected the data, and wrote the manuscript draft for this chapter. Sohrab Rohani supervised the project, conducted project administration, acquired funding, and edited the manuscript.

A version of this chapter was published as:

Ahmadi, S., & Rohani, S. (2023). Overcoming the Hydrophobic Nature of Zinc Phenylacetate Through Co-Crystallization with Isonicotinamide. *Journal of Pharmaceutical Sciences*, *112*(7), 1929–1938.

Chapter 4: Soroush Ahmadi conducted the computational screening and experimental analysis of Rufinamide and wrote the manuscript draft. Pradip Kumar Mondal contributed to the experimental/crystallographic analysis and reviewed the manuscript. Yuanyi Wu helped with the data processing. Weizhong Gong conducted the experimental screening of Olanzapine. Mahmoud Mirmehrabi and Sohrab Rohani were project supervisors.

A version of this chapter was published as:

Ahmadi, S., Mondal, P. K., Wu, Y., Gong, W., Mirmehrabi, M., & Rohani, S. (2021). Virtual Multicomponent Crystal Screening: Hydrogen Bonding Revisited. *Crystal Growth & Design*, *21*(10), 5862–5872

Chapter 5: Soroush Ahmadi conducted the DFT calculations, data analysis, and drafted the manuscript. Mohammad Amin Ghanavati contributed to the development of machine learning algorithms and assisted with the writing of the manuscript. Sohrab Rohani acted as the main project supervisor, provided funding for the research, and revised the manuscript.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my Ph.D. supervisor, Prof. Sohrab Rohani, for his unwavering support, guidance, and mentorship throughout my research. His expertise, encouragement, and valuable insights have played a pivotal role in shaping my academic and professional growth.

I would also like to extend my sincere appreciation to Prof. Bernhard Spingler for hosting me as a visiting student at the University of Zurich and generously sharing his expertise with me. Being part of his research group allowed me to deepen my understanding of crystallography and develop practical skills required for crystal structure determination.

I would like to warmly thank my family for their unwavering love, encouragement, and constant support throughout my academic pursuits. They have been a constant source of motivation, and I am truly grateful for their presence in my life.

I am grateful to all my labmates, Yuanyi, Pradip, Amin, Weizhong, Mengxing, Ehsan, and Hafezeh for their camaraderie, intellectual discussions, and support during my Ph.D. journey.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1 Introduction

1.1 Solid Active Pharmaceutical Ingredients

During the process of discovering and developing Active Pharmaceutical Ingredients (APIs), pre-formulation is an important stage as it aims to optimize the performance of the identified drug candidates (Almarsson & Vadas, 2015; Baghel et al., 2016). APIs are commonly formulated in the solid state, although liquid or gas formulations (such as nitrous oxide) are also possible. The preference for solid-state formulation stems from its enhanced physical and chemical stability, ease of handling, packaging, and downstream processing (Shan & Zaworotko, 2008; Vippagunta et al., 2001). Consequently, solid-state formulations have been widely employed to ensure the successful translation of developed drugs into the market and real-world application.

Solid-state formulations based on the arrangement of constituent particles can be divided into two classes of amorphous and crystalline solids. Amorphous solids lack order in their molecular arrangements, exhibiting liquid state disorder while having solid rheological properties. A molecule in amorphous form is in a higher energy state that may be advantageous and lead to an increase in solubility and faster drug release. However, there is a major issue with amorphous formulations as they are thermodynamically unstable and tend to convert to a low-energy crystalline solid form during storage. Thus, most APIs consist of crystalline solids, which have their own diverse landscape, as depicted in Figure 1-1 (Sakamoto & Uekusa, 2020).
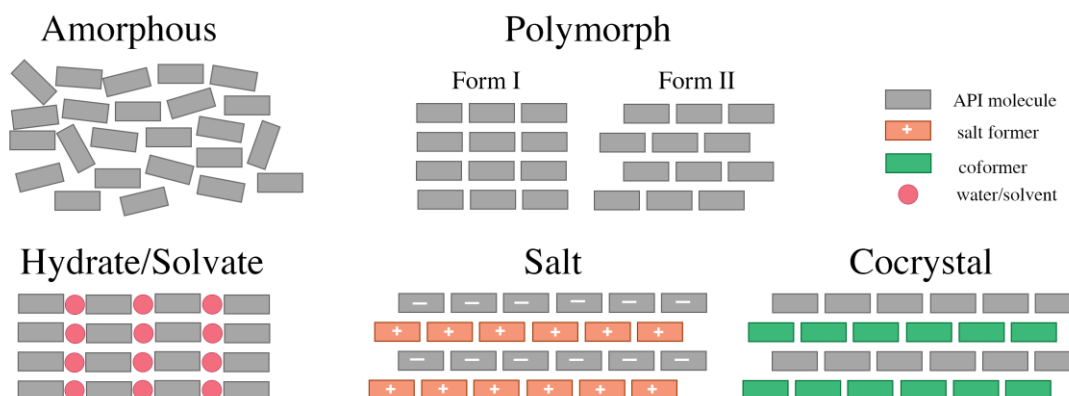
Figure 1-1 Illustration of different solid forms of an API

API molecules exhibit polymorphism, a phenomenon characterized by the tendency to assemble into different crystal structures (Bernstein, 2002). Polymorphs are of two kinds: packing polymorphism in which rigid molecules are packed differently, and conformation polymorphism in which flexible molecules bend in different conformations, changing the crystal structure (Morissette et al., 2004). API molecules can also exist as multicomponent crystals by incorporating other molecules in the lattice. The inclusion of water or solvent molecules in the lattice leads to the formation of hydrates and solvates, respectively. If both compounds are solid at room temperature, the resulting multicomponent crystal is either called a salt or a cocrystal (Cerreia Vioglio et al., 2017). The distinctive feature of cocrystals is that they are composed of two neutral molecules, while salts are formed when the API donates or receives a proton from another molecule (Desiraju et al., 2011).

1.2   Addressing Physicochemical Challenges

Solid formulations are generally intended for oral delivery, during which the API must first dissolve, avoid precipitation during gastrointestinal transit, and finally get absorbed at the absorption site (Khadka et al., 2014). Unfortunately, 40% of marketed drugs and more than 80% of pipeline drugs have solubility issues which may prevent the drug from reaching the minimum required therapeutic concentrations (Good & Rodríguez-Hornedo, 2009; Kawabata et al., 2011). Besides solubility, hygroscopicity (Newman et al., 2008), chemical or photo instability (Zhu et al., 2016), and brittleness (Sun & Grant, 2004) are some other major undesirable physicochemical properties that solid-state APIs may suffer from. Hygroscopic drugs pose significant production costs to dry the API during processing, storage, and the need for sealed packaging (Visalakshi et al., 2005). Brittleness poses difficulties for milling, filling, and compaction (Sun & Grant, 2001). Chemical or photo instability can also strongly impact the safety, quality, and efficacy of an API (Gupta et al., 2018; Putra et al., 2016). To tackle these unfavorable physicochemical properties, three main strategies are usually considered: formulation, particle engineering, and crystal engineering, as illustrated in Figure 1-2.
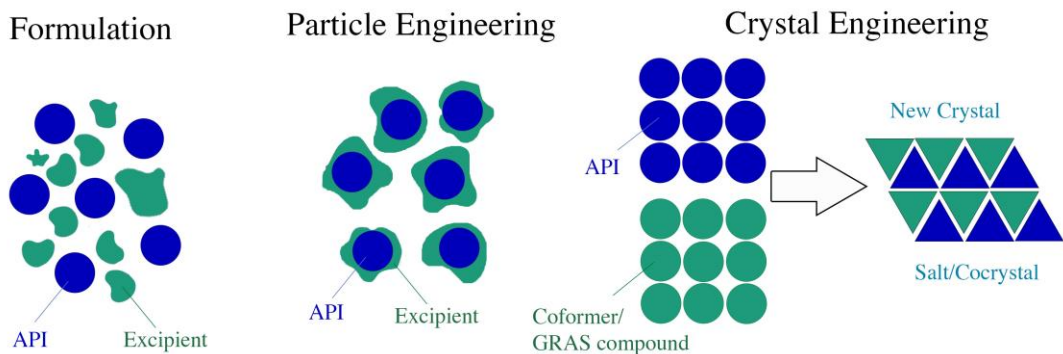
Figure 1-2 Schematic illustration of three main strategies to enhance the physicochemical properties of an API.

A formulation scientist selects and combines inert substances, known as excipients, with the API. These excipients serve various functions, including absorption enhancers, fillers, emulsifiers, extenders, diluents, flavors, colouring agents, preservatives, wetting agents, solvents, and sustained release matrices (Abrantes et al., 2016). By carefully selecting and combining these excipients, the formulation scientist can optimize various properties such as dissolution rate, taste, appearance, and shelf-life.

Particle Engineering is concerned with altering particle size, size distribution, morphology, and surface modification, with or without the addition of excipients (Chattoraj & Sun, 2018; Iacocca et al., 2010). Particle engineering can address some of the physicochemical issues.

And lastly, crystal engineering concerns the strategic production of new classes of crystalline products, using molecules as building blocks and considering their ability in the formation of supramolecular structures (Desiraju et al., 2011). In other words, crystal engineering can be regarded as the application of supramolecular chemistry and self-assembly upon crystalline solids, as a result of weak but directional intermolecular events (Almarsson & Zaworotko, 2004). Multicomponent crystals introduced in Section 1.1 fit under this crystal engineering strategies.

1.2.1 Crystal Engineering of Pharmaceutical Solids

Although we have previously discussed three distinct categories of multicomponent crystals (solvates, salts, and cocrystals), it is important to note that there are overlapping regions that significantly expand the diversity of multicomponent crystals. Phenomena such as salt solvates, cocrystal solvates, ionic cocrystals (cocrystal salts), and even cocrystal salt solvates, have all been observed and studied (Grothe et al., 2016).

Moreover, cocrystals can exhibit polymorphism as well as stoichiometric diversity. For instance, cocrystals of carbamazepine with nicotinamide and saccharin are polymorphic (Porter III et al., 2008), and both 2:1 and 1:1 cocrystals of carbamazepine with 4-aminobenzoic acid are discovered (Jayasankar et al., 2009).

The extensive structural diversity encountered in multicomponent crystals offers a unique opportunity to tailor the solid forms of APIs and fine-tune their properties. Despite such diversity and complexity, all multicomponent systems are typically stabilized through non-covalent interactions, designed using crystal engineering principles. The rationalization of these crystal formations is mostly based on interactions between functional groups present in the molecular skeletons to generate *supramolecular synthons* (Desiraju, 1995). Examples of homosynthon (synthons between the same functional groups) and heterosynthons (synthons between different functional groups) are shown in Figure 1-3. These synthons play a crucial role in dictating the crystal packing and properties of multicomponent crystals.

homosynthon   heterosynthon
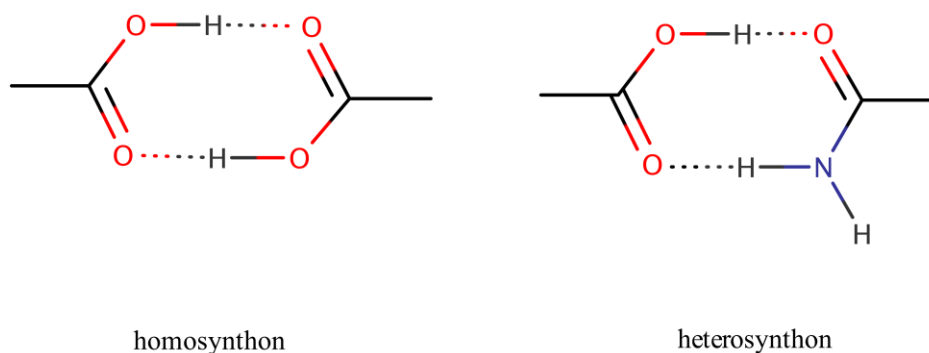
Figure 1-3 Examples of supramolecular synthons as the crystal engineering alphabet.

Cocrystals, although relatively new in the spotlight, are gaining increasing attention as an attractive multicomponent crystal. As a result, a growing number of commercially available cocrystals are entering the market. Beta chlor® Depakote®, Entresto®, Lexapro® (ESIX-10®), Steglatro®, Suglat® , Cafcit®, Zafatek®, and Lamivudine/zidovudine Teva®, are some commercially available cocrystal-based pharmaceutical products in 2023 (Singh et al., 2023). The following table provides a summary of the six commercial cocrystals and the how they could improve some physicochemical properties of the target API.

Table 1-1 Commercial pharmaceutical cocrystals (Kumar Bandaru et al., 2021).

Commercial name	API	Coformer	Improved property
Beta chlor®	Chloral hydrate	Betaine	Improved thermal stability
Depakote®	Valproic acid	Valproate sodium	Solid phase stability and less hygroscopicity
Entresto®	Valsartan	Sacubitril	Improved pharmacokinetics and bioavailability of valsartan
Lexapro®	Escitalopram	Oxalate	Improved stability of API
Steglatro®	Ertugliflozin	Z-Pyroglutamic acid	Improved stability
Suglat®	Ipragliflozin	L-Proline	Stability against hydrate formation

In summary, the field of multicomponent crystals encompasses a vast array of intriguing phenomena and structural possibilities, offering a rich landscape for exploring and designing novel solid forms with tailored properties. Gaining understanding and control over non-covalent interactions and supramolecular synthons entails the development of innovative strategies in pharmaceutical research and beyond.

1.3   Cocrystal Preparation

With the growing interest in cocrystals, researchers have been actively exploring various approaches and techniques to synthesize cocrystals. Cocrystal production methods are broadly classified into three groups: solid-state and solution-based methods, which have received extensive attention in the literature, and the use of supercritical fluids, primarily CO_2, which has also been explored for cocrystal preparation (Karimi-Jafari et al., 2018).

Solid-state methods	Solution-based methods	Supercritical fluid methods
Contact formation	Evaporative cocrystallization	Cocrystallization with supercritical solvent
Dry (neat) grinding	Cooling Crystallization	
Liquid-assisted grinding	Antisolvent cocrystallization	Supercritical antisolvent cocrystallization
Extrusion	Isothermal slurry conversion	
Hot melt extrusion	Spray drying	Supercritical assisted spray drying
High shear wet granulation	Reaction cocrystallization	

Figure 1-4 Summary of cocrystal preparation techniques.

1.3.1 Solid-state Methods

Cocrystals may form by simply mixing the powders of the API and coformer. Despite the potential acceleration achieved by pre-milling the compounds, the conversion process requires a significant amount of time, ranging from days to months. For instance, a cocrystal of carbamazepine and nicotinamide was successfully formed after 80 days of contact between un-milled pure compounds, while in the pre-milled case, the cocrystal was achieved after 12 days (Maheshwari et al., 2009). A more effective form of contact formation approach involves hot-stage microscopy. This technique involves melting and subsequent solidification of one compound, followed by introducing the second compound to the solidified material. As a

result, a portion of the first compound dissolves, and upon recrystallization of the mixture, a zone of mixing is formed (Berry et al., 2008). Additionally, melt crystallization can be conducted using a differential scanning calorimetry (DSC) instrument. In this method, a binary physical mixture of an API and a coformer is heated, and the observed phase changes in the DSC thermogram can provide insights into cocrystal formation (Saganowska & Wesolowski, 2018).

Solid-state grinding is a popular method that has proven to be a very successful for cocrystal preparation. This method can be performed either in dry (neat) conditions or with the assistance of a liquid (wet grinding). By applying mechanical force via a mortar and pestle or a ball milling machine, the cocrystal formation process can be accelerated to just a matter of minutes. Solvents such as methanol, ethanol, and nitromethane have been found to accelerate the cocrystal formation by wetting the solid surface during grinding when added in small amounts (10-50 µL). (Karimi-Jafari et al., 2018) Liquid-assisted grinding also reduces the risk of incomplete conversion, crystalline defects, and generation of amorphous content which are commonly observed in neat grinding.

Low-temperature extrusion (using single- or twin-screw extruders) has emerged as another technique for cocrystal synthesis. Moreoever, hot melt extrusion has also been proposed, a process involving both melting and mixing within a heated screw extruder. The continuous nature of the extrusion process offers significant potential for the scale-up of pharmaceutical cocrystal manufacturing (Moradiya et al., 2014). And finally, high-shear wet granulation, typically used for drug product formulation by agglomerating powder particles via a liquid medium in the presence of a binder, has also been used for cocrystal manufacturing (Rehder et al., 2013).

1.3.2 Solution-based Methods

Solution-based methods of cocrystallization follow the principles of single-component crystallization which is driven by supersaturation. However, for cocrystals, the concentrations of both API and coformer must be considered. Ternary phase diagrams are usually developed and used to control the region of concentrations which lead to cocrystallization.

Evaporative cocrystallization achieves supersaturation by removing the solvent by evaporation. Antisolvent cocrystallization involves mixing a concentrated cocrystal solution with a solvent of limited solubility, causing the system to reach a state of supersaturation. These methods are commonly used to generate single crystals suitable for diffraction studies and crystal structure elucidation of cocrystals (Karimi-Jafari et al., 2018; Spingler et al., 2012). Cooling cocrystallization, sometimes combined with seeding, has been studied to establish a scalable cocrystallization strategy (Sheikh et al., 2009).

Isothermal slurry conversion is another solution-based method that offers an alternative approach to cocrystal formation. It involves the addition of the API to a solution or suspension of the coformer, without the need for a clear starting solution. The conversion times are reported between 15 minutes and 5 hours, depending on solution concentration and solvent choice (Zhang & Rasmuson, 2013).

1.3.3 Supercritical Fluid Methods

Supercritical CO_2 offers three distinct properties that can be utilized for cocrystal preparation: solvent, antisolvent, and atomization (Padrela et al., 2009). When using the solvent aspect of supercritical fluid, API and coformers that are soluble in supercritical CO_2 can be used. Since the density and solvent power of CO_2 can be fine-tuned by changing the thermodynamic conditions, this method provides additional control over the cocrystallization (Neurohr et al., 2015). On the other hand, if CO_2 reduces the solubility of an API, supercritical antisolvent cocrystallization can be applied. In a batch gas antisolvent setup, a cocrystal-saturated solution is placed in a high-pressure vessel with CO_2, allowing for the formation of cocrystals as CO_2 diffuses into the solution (Padrela et al., 2009). Lastly, by expanding and depressurizing a saturated supercritical solution through a nozzle into a drying chamber, microparticles of the cocrystal can be produced (Müllers et al., 2015).

In summary, the combination of solid-based, solution-based, and supercritical fluid-based techniques provides a diverse range of approaches for the successful synthesis of pharmaceutical cocrystals, enabling researchers to successfully synthesize novel cocrystals and study their applications in the drug development process.

1.4   Cocrystal Characterization

After attempting to generate a new multicomponent drug form, such as cocrystals, it is essential to employ a range of characterization methods. These methods serve two main objectives. Firstly, they aim to confirm whether a new crystalline form has been formed, or the experiment has yielded no result. Secondly, the physicochemical properties of the newly obtained crystalline compound must be determined, allowing for a comparison with the starting API to determine if the desired target has been achieved. The following characterization techniques enable scientists to acquire valuable insights into the success rate of the multicomponent development process and identify the potential advantages of the newly synthesized crystals.

1.4.1   Crystallographic Studies

Single crystal X-ray diffraction (SCXRD) is the gold standard method to determine the crystal structure and prove a new cocrystal is formed. Crystal structure determination requires growing large enough single crystals that diffract. In this method, a carefully selected single crystal will be exposed to X-ray irradiation at different orientations and the resulting diffraction patterns are recorded on a detector. The experimentally measured diffraction intensities can then be used to derive the structure factors from which a model of periodic electron density distribution can be obtained. The determined structure provides information about the chemical composition in the unit cell, the molecular arrangement, intermolecular interactions, bond lengths, and bond angles.

Since most cocrystal screening methods yield microcrystalline powders, SCXRD is not a common first characterization step. Instead, powder X-ray diffraction (PXRD) emerges as the preferred method of choice for initial assessment. PXRD generates one-dimensional diffraction intensity plots of the powdered sample as a function of 2θ. PXRD patterns are distinctive and correspond to a unique crystal structure. Thus, PXRD serves as a "fingerprint" of a crystalline phase. In the case of cocrystallization, a successful formation would result in a new diffraction pattern distinct from the two starting materials, rather than a simple overlay of their patterns (physical mixture). This distinction makes PXRD a perfect tool for the initial identification of a successful cocrystallization attempt. Additionally, an expert user of PXRD can go beyond pattern matching and extract further information from the PXRD pattern. This includes

determining unit cell dimensions, the space group, and even an approximate structural model using Rietveld refinement.

1.4.2 Spectroscopic Characterization

There are two categories of spectroscopy methods used for solid-state cocrystal studies: vibrational spectroscopy (FT-IR and Raman), and solid-state nuclear magnetic resonance (ssNMR).

The mid-IR region (4000-400 cm^{-1}) is relevant for identifying the stretching and bending of bonds within a structure. Intermolecular interactions (such as hydrogen bonding) can cause shifts in the IR bands and result in the appearance of a distinct pattern. These shifts serve as evidence for a new crystal structure and enable the discussion of probable interactions within the crystal lattice. The near-IR (14,000-4000 cm^{-1}) and far-IR (400-10cm^{-1}) regions have also been used for additional confirmation on intermolecular vibrations such as hydrogen bonds and Van der Waals interactions. Moreover, the Raman spectrum of a cocrystal also differs from that of the starting materials, providing another means of characterization.

It has been shown that advanced 1D and 2D ssNMR analysis that can detect dipolar connectivity can provide strong confirmatory structural information about a cocrystal by unraveling ^{1}H shifts, locating hydrogen bonding groups, and identifying close atomic connectivity using different dipolar couplings (Vogt et al., 2009).

1.4.3 Thermal Analysis

Thermal analysis encompasses a group of techniques that monitor a sample's property under programmed temperature conditions and within a controlled atmosphere. The three main techniques commonly employed for cocrystal studies are differential scanning calorimetry (DSC), thermogravimetric analysis (TGA), and hot-stage microscopy (HSM).

DSC is a widely used thermal analysis method that measures the heat flow difference between a sample and a reference pan as a function of temperature (Pindelska et al., 2017). It enables the identification of important thermal events including melting point, glass transition,

decomposition temperature, cold crystallization, and can also be utilized for conducting physical and oxidative stability tests as well as purity control.

TGA measures the mass change (loss or gain) of a substance as it is subjected to varying temperatures. When combined with DSC, it provides a comprehensive understating of the observed phase changes in a thermograph (Pyramides et al., 1995). TGA can characterize the sorption/desorption of volatiles, oxidation, decomposition, and reduction events. TGA is particularly used for assessing the improved thermal stability of a cocrystal compared to the starting API.

HSM combines thermal analysis with microscopy to visually follow the physical properties as temperature increases, detecting small changes in the sample that may be missed by DSC and TGA (Kumar et al., 2020). HSM is used in pharmaceutical studies to study solid-state phase transformation, interactions between different compounds, sublimation, melting range, crystal growth, and polymorph screening.

All these thermal analyses provide valuable insight into the behavior and characteristics of pharmaceuticals, aiding their development and stability assessment.

1.4.4 Solubility and Dissolution Studies

After confirming the presence of a new crystalline phase, it is crucial to evaluate its properties. The most sought-after property enhancement, as discussed before, is solubility and dissolution profile. To measure solubility, solution and 'excess solid' are mixed until equilibrium is reached. The saturated solution is then filtered and sampled, and the concentration is determined. Concentration measurement can be performed by various methods including UV-vis spectroscopy, HPLC, NMR, or simple methods such as titration, turbidity, and gravimetric methods (Black et al., 2013).

To understand the dissolution process and track the concentration fluctuations as a function of time, kinetic studies are conducted. Powder dissolution tests typically use the paddle or rotating basket apparatus (Pindelska et al., 2017). Alternatively, intrinsic dissolution rate (IDR) can be measured which tracks the dissolution rate of a drug under constant surface area, stirring

rate, pH, and ionic strength of the dissolution medium (Stuart & Box, 2005). These tests provide insight into the dissolution behavior of a new crystalline phase.

1.4.5   Stability Studies

After solubility, stability is the most important property which can be enhanced with cocrystallization. Food and Drug Administration (FDA) states that a stability test should encompass those attributes of the API that are susceptible to change during storage and may influence the quality, safety, and efficacy of the API. A stable drug must maintain its physical, chemical, biological, and microbial properties during the storage and usage periods (FDA, 2003).

Temperature, moisture, and pH are among the key parameters that can affect the stability of an API. Thermal analysis, mainly DSC and TGA are the main techniques employed to determine the stability of a compound to temperature, as discussed earlier. Moisture sorption, which relates to hygroscopicity, is another important aspect to consider during API development.

The hygroscopicity of a compound can be determined using various methods. The simple approach involves using a desiccator. In a desiccator, the sample is exposed to different relative humidity (RH) levels created by saturated solutions of various salts (Greenspan, 1977). The weight change of the sample at different RH can be determined over specified time periods. Alternatively, an automated moisture sorption technique, known as dynamic vapour sorption (DVS) can be utilized. DVS is equipped with a microbalance and the experiment involves subjecting the sample to a continuous flow of nitrogen gas with a user-defined RH. The experiment typically involves ramping up and down the RH to observe the weight change (sorption/desorption behavior of the sample) under different conditions (Garbalińska et al., 2017). Lastly, PXRD or ssNMR techniques can be employed to explore potential phase transformations due to exposure to moisture.

Overall, the characterization of solid pharmaceuticals had a crucial role in advancing the field of multicomponent pharmaceutical solids. Through advancements in crystallography, vibrational spectroscopy, and ssNMR, valuable insights into the molecular-level details of different solid forms are gained. These analytical techniques, along with approaches such as

HPLC for solubility measurement, DCS and TGA for thermal stability assessment, and DVS for moisture stability evaluation have played a significant role in studying the properties of new solid forms.

1.5 Cocrystal Screening: Computational Tools for Design

In this section, we highlight the necessity of complementary predictive tools alongside experimental methods for cocrystal screening and introduce several methodologies explored in the literature. There is a vast number of potential coformers, as evident from the Generally Recognized as Safe (GRAS) (FDA, 2022) and Everything Added to Foods in the United States (EAFUS) (FDA, 2018) lists. With a vast range of potential coformers available, the sheer number of combinations makes complete experimental screening impractical.

In addition to the property-enhancement aspect of cocrystals, regulatory guidelines of the United States Food and Drug Administration (FDA) and European Medicines Agency (EMA) state that cocrystals are eligible for intellectual property protection rights. Thus, cocrystals' patentability has further intensified competition among innovative and generic manufacturing companies to secure potential patents.

Considering the time and resource demands of brute-force experimental approaches, the need for reliable *in-silico* coformer screening techniques to narrow down the scope and identify the most likely candidates becomes crucial (Kumar & Nanda, 2021).

1.5.1 Cambridge Structural Database (CSD)

CSD, established by Cambridge Crystallographic Data Center (CCDC) in 1965, contains over 1.2 million curated entries as of 2023. It is the largest repository for accurate crystal structures of organic and organo-metallic compounds. In the context of cocrystal screening, it enables the analysis of crystal packing features and hydrogen bond donors and acceptors competitions. These knowledge-based approaches are based on synthon design and involves the assessment of possible synthons between API and coformers.

Particularly, for hydrogen bond analysis, the CSD Mercury program offers a tool called hydrogen bond propensity (HBP). Originally designed for identifying polymorphic structures,

HBP has demonstrated its potential as a knowledge-based strategy for cocrystal design. The HBP analysis consists of four steps: I) identifying functional groups within molecular structures, II) extracting hundreds of reported crystal structures in CSD with similar functional groups, III) assigning a propensity value for all potential donor-acceptor pairs after analyzing the extracted structures, and IV) assessing the competition in propensity values between donors and acceptors present the API, coformer, and cocrystal. By subtracting the propensity of the most likely interaction between API:API or coformer:coformer from that of the API:coformer, a multicomponent (MC) score is determined. The higher positive values of the MC score indicate a greater potential for cocrystal formation.

CSD Mercury also features a molecular complementarity module, designed for cocrystal screening (Fábián, 2009). This module was developed after analysis of 1949 molecules over 131 molecular descriptors, which revealed that molecular shape and polarity are the key shared similarities among molecules within cocrystal structures. To capture molecular shape similarities, three shape factors, namely the length of the short axis, the short/long, and medium/long axis ratios were identified and selected for prediction. For polarity considerations, the fraction of nitrogen and oxygen atoms, as well as the dipole moment were selected as relevant descriptors. One can conduct a coformer screening by leveraging any combination of these five molecular features within the Mercury program.

1.5.2 Hansen Solubility Parameters (HSP) for Cocrystal Prediction

Hansen in the late 1960s developed a conceptual framework to predict the solubility and compatibility of substances. The second edition of the handbook was released in 2007 (Hansen, 2007). HSP considers three forces: dispersion, polar, and hydrogen bonding. These parameters are commonly determined by group contribution which only requires a two-dimensional molecular structure to calculate.

Besides the original intent to predict solubilities, HSP has found applications in various fields including cocrystal prediction. One view of cocrystals is the "miscibility" of API and coformer at a molecular level. Therefore, the two solids are "soluble" in each other and HSP can be applicable to identify molecular compatibilities.

Mohammad et al. tested this postulate and used HSPs to predict cocrystals of indomethacin. Among thirty candidate coformers, the differences between the HSP of the API and coformers were calculated. Based on these calculations, the miscibilities of the combinations were predicted. The subsequent experimental investigation confirmed that all except one of the combinations that were predicted miscible indeed formed cocrystals (Mohammad et al., 2011). This study showed that HSP has significant potential in cocrystal screening.

1.5.3 Quantum-mechanical Views on Cocrystal Screening

Quantum chemistry is a powerful branch of science that applies the principles and equations of quantum mechanics to investigate molecules, their interactions, material properties, spectra, and thermodynamic properties. Sophisticated molecular and crystal representations derived from quantum calculations offer promising strategies to study and better understand cocrystal formation.

A particularly promising approach is the use of molecular electrostatic potential maps (ESP). From a gas-phase optimized structure of a molecule, a three-dimensional ESP surface can be generated that shows the charge distribution of a molecule. These calculations are usually performed using Density-functional theory (DFT). Pioneering work by (Musumeci et al., 2011) introduced a strategy to use ESP maps to determine the probability of cocrystal formation. The prediction involved obtaining the local maxima and minima of the ESP maps and converting them into lists of hydrogen bond donors (α) and acceptors (β). These lists, referred to as surface site interaction points (SSIP), enable the quantification of pairing energy of possible intermolecular interactions by multiplying corresponding pairs of (α, β). By comparing the site pairing energies of the pure compounds of the API and the coformer with the pairing energy in a cocrystal, we can estimate the likelihood of cocrystal formation.

Closely related to the ESP method, the conductor-like screening model for real solvents (COSMO-RS) theory combines quantum mechanical calculations with liquid phase thermodynamics to predict properties in the liquid state (Klamt, 2005). For cocrystal prediction, COSMO-RS assumes that the miscibility observed in the supercooled melt phase of cocrystal components closely represents the behavior of the cocrystal in its solid state

(Loschen & Klamt, 2015). The excess enthalpy (H_{ex}) of the cocrystal relative to that of the pure API and coformer is used as a measure to estimate the cocrystallization tendency, with more negative values indicating a more favorable cocrystal. H_{ex} is obtained after DFT calculation at BP-TZVP-COSMO level theory within the COSMOtherm package.

Cocrystal prediction based on lattice energy comparison goes beyond molecular descriptors and incorporates information from crystal structures. Since the crystal structure of the cocrystal is not determined during the screening process, a three-step methodology is proposed (Karamertzanis et al., 2009) to predict the cocrystal structure. It starts with generating a large number of conformations and estimating the stability of each configuration using the CrystalPredictor program. Inter- and intra-molecular energies are then estimated with the help of precomputed quantum calculations, and configurations are adjusted to minimize the lattice energy. After finding the most stable structure, the energy minimization step is conducted again to improve the data quality. Finally, a complex quantum mechanical calculation is conducted using the DMAflex algorithm (Karamertzanis & Price, 2006) to accurately calculate lattice energies.

If crystal structures are available, the computationally demanding task of crystal structure prediction can be bypassed, and the lattice energies can be computed using the PIXEL method (Gavezzotti, 2005). This approach combines classical formulas with electron density information obtained by quantum chemistry programs to calculate lattice energies. PIXEL uses the symmetry elements available in the Crystallographic Information File (CIF) to form a lattice and calculate the intermolecular interactions including Coulombic, dispersion, polarization, and repulsive interactions.

In conclusion, optimizing screening approaches is crucial to reduce the number of experiments needed for cocrystal discovery. Virtual screening techniques based on knowledge-based and theoretical approaches both aim to identify potential cocrystals. Knowledge-based tools like CSD uses the availability of previous cocrystals to predict the formation of supramolecular structures in yet-unknown cocrystals, while methods such as HSP, ESP, COSMO-RS, and lattice energy comparison provide insights into the thermodynamic landscape of cocrystal

formation. Overall, virtual screening guides subsequent experimental investigations and accelerates the cocrystal discovery process.

1.6 Research Objectives and Organization of Thesis

The overall objective of this thesis is to investigate the experimental and computational aspects of multicomponent crystals, specifically focusing on cocrystals. Additionally, it explores the control of crystal morphology as an addition topic of interest. The objectives include:

- Conduct a comprehensive search for multicomponent crystals of selected APIs (Zinc phenylacetate, Dasatinib, Rufinamide, and Olanzapine) using various experimental techniques such as liquid-assisted grinding, ball milling, evaporative, and cooling crystallization.

- Analyze the newly discovered multicomponent crystals to assess their properties compared to the stating API, searching for improvements in solubility, dissolution rate, hydrophilicity, and particle size.

- Develop *in-silico* methods for the identification of cocrystals by utilizing techniques such as ESP maps, HBP knowledge-based model, and advanced machine learning algorithms.

- Improve the thread-like crystal morphology of Rufinamide by obtaining its crystal structure, studying its crystalline faces, and employing various experimental techniques for crystal morphology engineering.

By addressing these objectives, this thesis aims to contribute to the understanding and advancement of multicomponent crystals, particularly cocrystals, and their potential applications in the pharmaceutical industry.

1.6.1 Thesis Organization

This thesis is organized in the integrated-article format.

Chapter 1 introduces the research topic and provides background information on the role of multicomponent crystals in the pharmaceutical industry. Various methods of experimental cocrystal screening are briefly discussed followed by the characterization techniques identification of the content in the new crystals and explore their physicochemical advantages. An overview of the prediction strategies to complement the experimental screening is also introduced.

Chapters 2 and 3 are experimentally focused publications. In Chapter 2, a new crystal structure of methanol solvate of Dasatinib is investigated. Using SCXRD, PXRD, HSM, and DSC, this new crystal is characterized. The effect of various crystallization techniques, on the final product has also been discussed in this chapter.

Chapter 3 centers around Zinc phenylacetate (Zn-PA) and its novel cocrystal with isonicotinamide (Zn-PA-INAM). A comprehensive analysis is conducted on this cocrystal using PXRD, SCXRD, FTIR, DSC, TGA, contact angle measurement, BFDH morphology analysis, lattice energy calculation, and Hirshfeld surfaces. These investigations provide a deeper understanding of how the cocrystal addresses the hydrophobic nature of Zn-PA and highlight the properties of the cocrystal system.

Chapters 4 and 5 are concerned with publications related to *in-silico* cocrystal prediction. In Chapter 4, the emphasis is on the application of hydrogen bonding analysis using ESP maps and the HBP method. Additionally, a multivariable logistic regression model is developed as a linear machine learning algorithm, which combines the strengths of individual prediction models to achieve superior classification performance.

Chapter 5 builds upon the application of ESP maps for cocrystal prediction and explores various machine learning models, including the deep learning PointNet network, artificial neural network, random forest, and ensemble learning. A large dataset of positive and negative cocrystallization experiments in literature is compiled. The molecular constituents of all these

cocrystals are extracted and their ESP map are calculated and fed into the machine learning models. The findings demonstrate the potential of ESP as a robust cocrystal prediction strategy.

Chapter 6 is concerned with explanation and control of the morphology of Rufinamide. Utilizing temperature cycling, solvent screening, and additive selection, we attempt to modify its thread-like morphology into a more isometric shape. The crystal structures of three RUF polymorphs are determined, and a connection between the microscopic structure and the macroscopic morphologies is established through face indexing. These findings provide valuable insights applicable to other pharmaceutical compounds with similar morphology-related challenges.

Chapter 7 summarizes the research work and provides suggestions for potential research ideas for future works.

1.7   References

Abrantes, C. G., Duarte, D., & Reis, C. P. (2016). An Overview of Pharmaceutical Excipients: Safe or Not Safe? *Journal of Pharmaceutical Sciences*, *105*(7), 2019–2026.

Almarsson, Ö., & Vadas, E. B. (2015). Molecules, Materials, Medicines (M3): Linking Molecules to Medicines through Pharmaceutical Material Science. *Crystal Growth & Design*, *15*(12), 5645–5647.

Almarsson, Ö., & 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*, *17*, 1889–1896.

Baghel, S., Cathcart, H., & O'Reilly, N. J. (2016). Polymeric Amorphous Solid Dispersions: A Review of Amorphization, Crystallization, Stabilization, Solid-State Characterization, and Aqueous Solubilization of Biopharmaceutical Classification System Class II Drugs. *Journal of Pharmaceutical Sciences*, *105*(9), 2527–2544.

Bernstein, J. (2002). *Polymorphism in Molecular Crystals*. Clarendon Press.

Berry, D. J., Seaton, C. C., Clegg, W., Harrington, R. W., Coles, S. J., Horton, P. N., Hursthouse, M. B., Storey, R., Jones, W., Friščić, T., & Blagden, N. (2008). Applying Hot-Stage Microscopy to Co-Crystal Screening: A Study of Nicotinamide with Seven Active Pharmaceutical Ingredients. *Crystal Growth & Design*, *8*(5), 1697–1712.

Black, S., Dang, L., Liu, C., & Wei, H. (2013). On the Measurement of Solubility. *Organic Process Research & Development*, *17*(3), 486–492.

Cerreia Vioglio, P., Chierotti, M. R., & Gobetto, R. (2017). Pharmaceutical aspects of salt and cocrystal forms of APIs and characterization challenges. *Advanced Drug Delivery Reviews*, *117*, 86–110.

Chattoraj, S., & Sun, C. C. (2018). Crystal and Particle Engineering Strategies for Improving Powder Compression and Flow Properties to Enable Continuous Tablet Manufacturing by Direct Compression. *Journal of Pharmaceutical Sciences*, *107*(4), 968–974.

Desiraju, G. R. (1995). Supramolecular Synthons in Crystal Engineering—A New Organic Synthesis. *Angewandte Chemie International Edition in English*, *34*(21), 2311–2327.

Desiraju, G. R., Vittal, J. J., & Ramanan, A. (2011). *Crystal Engineering: A Textbook*. World Scientific Publishing Co. Pte. Ltd.

Fábián, L. (2009). Cambridge Structural Database Analysis of Molecular Complementarity in Cocrystals. *Crystal Growth & Design*, *9*(3), 1436–1443.

FDA. (2003). *Guidance for Industry Q1A(R2) Stability Testing of New Drug Substances and Products*. U.S. Department of Health and Human Services Food and Drug Administration. www.fda.gov/media/71707/download

FDA. (2018). *Substances Added to Food (formerly EAFUS)*. U.S. Department of Health and Human Services Food and Drug Administration. www.fda.gov/food/food-additives-petitions/substances-added-food-formerly-eafus

FDA. (2022). *Generally Recognized as Safe (GRAS)*. U.S. Department of Health and Human Services Food and Drug Administration. www.fda.gov/food/food-ingredients-packaging/generally-recognized-safe-gras

Garbalińska, H., Bochenek, M., Malorny, W., & von Werder, J. (2017). Comparative analysis of the dynamic vapor sorption (DVS) technique and the traditional method for sorption isotherms determination — Exemplified at autoclaved aerated concrete samples of four density classes. *Cement and Concrete Research*, *91*, 97–105.

Gavezzotti, A. (2005). Calculation of lattice energies of organic crystals: the PIXEL integration method in comparison with more traditional methods. *Zeitschrift für Kristallographie - Crystalline Materials. 220*(5–6), 499–510.

Good, D. J., & Rodríguez-Hornedo, N. (2009). Solubility Advantage of Pharmaceutical Cocrystals. *Crystal Growth & Design*, *9*(5), 2252–2264.

Greenspan, L. (1977). Humidity fixed points of binary saturated aqueous solutions. *Journal of Research of the National Bureau of Standards Section A: Physics and Chemistry*, *81A*(1), 89.

Grothe, E., Meekes, H., Vlieg, E., ter Horst, J. H., & de Gelder, R. (2016). Solvates, Salts, and Cocrystals: A Proposal for a Feasible Classification System. *Crystal Growth & Design*, *16*(6), 3237–3243.

Gupta, D., Bhatia, D., Dave, V., Sutariya, V., & Varghese Gupta, S. (2018). Salts of Therapeutic Agents: Chemical, Physicochemical, and Biological Considerations. *Molecules*, *23*(7).

Hansen, C. M. (2007). *Hansen solubility parameters: a user's handbook*. CRC press.

Iacocca, R. G., Burcham, C. L., & Hilden, L. R. (2010). Particle engineering: A strategy for establishing drug substance physical property specifications during small molecule development. *Journal of Pharmaceutical Sciences*, *99*(1), 51–75.

Jayasankar, A., Reddy, L. S., Bethune, S. J., & Rodríguez-Hornedo, N. (2009). Role of Cocrystal and Solution Chemistry on the Formation and Stability of Cocrystals with Different Stoichiometry. *Crystal Growth & Design*, *9*(2), 889–897.

Karamertzanis, P. G., Kazantsev, A. V, Issa, N., Welch, G. W. A., Adjiman, C. S., Pantelides, C. C., & Price, S. L. (2009). Can the Formation of Pharmaceutical Cocrystals Be Computationally Predicted? 2. Crystal Structure Prediction. *Journal of Chemical Theory and Computation*, *5*(5), 1432–1448.

Karamertzanis, P. G., & Price, S. L. (2006). Energy Minimization of Crystal Structures Containing Flexible Molecules. *Journal of Chemical Theory and Computation*, *2*(4), 1184–1199.

Karimi-Jafari, M., Padrela, L., Walker, G. M., & Croker, D. M. (2018). Creating Cocrystals: A Review of Pharmaceutical Cocrystal Preparation Routes and Applications. *Crystal Growth & Design*, *18*(10), 6370–6387.

Kawabata, Y., Wada, K., Nakatani, M., Yamada, S., & Onoue, S. (2011). Formulation design for poorly water-soluble drugs based on biopharmaceutics classification system: Basic approaches and practical applications. *International Journal of Pharmaceutics*, *420*(1), 1–10.

Khadka, P., Ro, J., Kim, H., Kim, I., Kim, J. T., Kim, H., Cho, J. M., Yun, G., & Lee, J. (2014). Pharmaceutical particle technologies: An approach to improve drug solubility, dissolution and bioavailability. *Asian Journal of Pharmaceutical Sciences*, *9*(6), 304–316.

Klamt, A. (2005). *COSMO-RS From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design* (1st ed.). Elsevier Science.

Kumar, A., & Nanda, A. (2021). In-silico methods of cocrystal screening: A review on tools for rational design of pharmaceutical cocrystals. *Journal of Drug Delivery Science and Technology*, *63*, 102527.

Kumar, A., Singh, P., & Nanda, A. (2020). Hot stage microscopy and its applications in pharmaceutical characterization. *Applied Microscopy*, *50*(1), 12.

Kumar Bandaru, R., Rout, S. R., Kenguva, G., Gorain, B., Alhakamy, N. A., Kesharwani, P., & Dandela, R. (2021). Recent Advances in Pharmaceutical Cocrystals: From Bench to Market. *Frontiers in Pharmacology*, *12*.

Loschen, C., & Klamt, A. (2015). Solubility prediction, solvate and cocrystal screening as tools for rational crystal engineering. *Journal of Pharmacy and Pharmacology*, *67*(6), 803–811.

Maheshwari, C., Jayasankar, A., Khan, N. A., Amidon, G. E., & Rodríguez-Hornedo, N. (2009). Factors that influence the spontaneous formation of pharmaceutical cocrystals by simply mixing solid reactants. *CrystEngComm*, *11*(3), 493–500.

Mohammad, M. A., Alhalaweh, A., & Velaga, S. P. (2011). Hansen solubility parameter as a tool to predict cocrystal formation. *International Journal of Pharmaceutics*, *407*(1), 63–71.

Moradiya, H. G., Islam, M. T., Halsey, S., Maniruzzaman, M., Chowdhry, B. Z., Snowden, M. J., & Douroumis, D. (2014). Continuous cocrystallisation of carbamazepine and trans-cinnamic acid via melt extrusion processing. *CrystEngComm*, *16*(17), 3573–3583.

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*(3), 275–300.

Müllers, K. C., Paisana, M., & Wahl, M. A. (2015). Simultaneous Formation and Micronization of Pharmaceutical Cocrystals by Rapid Expansion of Supercritical Solutions (RESS). *Pharmaceutical Research*, *32*(2), 702–713.

Musumeci, D., Hunter, C. A., Prohens, R., Scuderi, S., & McCabe, J. F. (2011). Virtual cocrystal screening. *Chemical Science*, *2*(5), 883–890.

Neurohr, C., Marchivie, M., Lecomte, S., Cartigny, Y., Couvrat, N., Sanselme, M., & Subra-Paternault, P. (2015). Naproxen–Nicotinamide Cocrystals: Racemic and Conglomerate Structures Generated by CO_2 Antisolvent Crystallization. *Crystal Growth & Design*, *15*(9), 4616–4626.

Newman, A. W., Reutzel-Edens, S. M., & Zografi, G. (2008). Characterization of the 'hygroscopic' properties of active pharmaceutical ingredients. *Journal of Pharmaceutical Sciences*, *97*(3), 1047–1059.

Padrela, L., Rodrigues, M. A., Velaga, S. P., Matos, H. A., & de Azevedo, E. G. (2009). Formation of indomethacin–saccharin cocrystals using supercritical fluid technology. *European Journal of Pharmaceutical Sciences*, *38*(1), 9–17.

Pindelska, E., Sokal, A., & Kolodziejski, W. (2017). Pharmaceutical cocrystals, salts and polymorphs: Advanced characterization techniques. *Advanced Drug Delivery Reviews*, *117*, 111–146.

Porter III, W. W., Elie, S. C., & Matzger, A. J. (2008). Polymorphism in Carbamazepine Cocrystals. *Crystal Growth & Design*, *8*(1), 14–16.

Putra, O. D., Yoshida, T., Umeda, D., Gunji, M., Uekusa, H., & Yonemochi, E. (2016). Crystallographic Analysis of Phase Dissociation Related to Anomalous Solubility of Irsogladine Maleate. *Crystal Growth & Design*, *16*(12), 6714–6718.

Pyramides, G., Robinson, J. W., & William Zito, S. (1995). The combined use of DSC and TGA for the thermal analysis of atenolol tablets. *Journal of Pharmaceutical and Biomedical Analysis*, *13*(2), 103–110.

Rehder, S., Christensen, N. P. A., Rantanen, J., Rades, T., & Leopold, C. S. (2013). High-shear granulation as a manufacturing method for cocrystal granules. *European Journal of Pharmaceutics and Biopharmaceutics*, *85*(3, Part B), 1019–1030.

Saganowska, P., & Wesolowski, M. (2018). DSC as a screening tool for rapid co-crystal detection in binary mixtures of benzodiazepines with co-formers. *Journal of Thermal Analysis and Calorimetry*, *133*(1), 785–795.

Sakamoto, Masami., & Uekusa, Hidehiro. (2020). Advances in Organic Crystal Chemistry Comprehensive Reviews 2020. In *Springer Nature eBook* (1st ed. 2020.). Springer Singapore.

Shan, N., & Zaworotko, M. J. (2008). The role of cocrystals in pharmaceutical science. *Drug Discovery Today*, *13*(9), 440–446.

Sheikh, A. Y., Rahim, S. A., Hammond, R. B., & Roberts, K. J. (2009). Scalable solution cocrystallization: case of carbamazepine-nicotinamide I. *CrystEngComm*, *11*(3), 501–509.

Singh, M., Barua, H., Jyothi, V. G. S. S., Dhondale, M. R., Nambiar, A. G., Agrawal, A. K., Kumar, P., Shastri, N. R., & Kumar, D. (2023). Cocrystals by Design: A Rational Coformer Selection Approach for Tackling the API Problems. *Pharmaceutics*, *15*(4), 1161.

Spingler, B., Schnidrig, S., Todorova, T., & Wild, F. (2012). Some thoughts about the single crystal growth of small molecules. *CrystEngComm*, *14*(3), 751–757.

Stuart, M., & Box, K. (2005). Chasing equilibrium: Measuring the intrinsic solubility of weak acids and bases. *Analytical Chemistry*, *77*(4), 983–990.

Sun, C., & Grant, D. J. W. (2001). Influence of Crystal Structure on the Tableting Properties of Sulfamerazine Polymorphs. *Pharmaceutical Research*, *18*(3), 274–280.

Sun, C., & Grant, D. J. W. (2004). Improved Tableting Properties of p-Hydroxybenzoic Acid by Water of Crystallization: A Molecular Insight. *Pharmaceutical Research*, *21*(2), 382–386.

Vippagunta, S. R., Brittain, H. G., & Grant, D. J. W. (2001). Crystalline solids. *Advanced Drug Delivery Reviews*, *48*(1), 3–26.

Visalakshi, N. A., Mariappan, T. T., Bhutani, H., & Singh, S. (2005). Behavior of Moisture Gain and Equilibrium Moisture Contents (EMC) of Various Drug Substances and Correlation with Compendial Information on Hygroscopicity and Loss on Drying. *Pharmaceutical Development and Technology*, *10*(4), 489–497.

Vogt, F. G., Clawson, J. S., Strohmeier, M., Edwards, A. J., Pham, T. N., & Watson, S. A. (2009). Solid-State NMR Analysis of Organic Cocrystals and Complexes. *Crystal Growth & Design*, *9*(2), 921–937.

Zhang, S., & Rasmuson, Å. C. (2013). Thermodynamics and Crystallization of the Theophylline–Glutaric Acid Cocrystal. *Crystal Growth & Design*, *13*(3), 1153–1161.

Zhu, B., Wang, J.-R., Zhang, Q., & Mei, X. (2016). Improving Dissolution and Photostability of Vitamin K3 via Cocrystallization with Naphthoic Acids and Sulfamerazine. *Crystal Growth & Design*, *16*(1), 483–492.

Chapter 2

Desolvation of Dasatinib Methanolate: An Improved Anhydrous Polymorph

A version of this chapter was published as:

2 Desolvation of Dasatinib Methanolate: An Improved Anhydrous Polymorph

Abstract

Transient solvates are unstable crystals that readily desolvate on harvesting but can significantly alter the outcome of the crystallization process. In this study, we investigated a new crystal structure of methanol solvate of Dasatinib that shows transient characteristics. Using SCXRD, PXRD, HSM, and DSC, this new crystal was characterized to shed light on the mechanism of achieving different anhydrous polymorphs of Dasatinib upon desolvation. The single-crystal X-ray structure of this methanol solvate was determined for the first time in this work. Via Hirshfeld surface analysis and molecular electrostatic potential map (ESP), the intermolecular interactions are discussed. In addition, void map analysis shed light on the transient nature of DAS-MeOH solvate crystals. The effect of various crystallization techniques, including slow evaporation and grinding/milling techniques, on the product of desolvation, has also been discussed. Finally, with the aid of particle size distributions measurements, we illustrated the potential benefit of such transient crystals on particle size reduction.

2.1 Introduction

Multicomponent crystallization is an effective technique in hands of a crystal engineer to fine-tune the properties of a molecule, without changing its chemical structure (Desiraju et al., 2011; Mir et al., 2019; Mondal et al., 2017). Among the most common subcategories of multicomponent crystals (solvates, salts, cocrystals, solid solutions, and inclusion compounds) (Duggirala et al., 2016), solvate formation is frequently observed among Active Pharmaceutical Ingredients (API) and generates a challenging yet rewarding area for both scientific research and industry (Griesser, 2006; Healy et al., 2017; Stieger, 2012).

Historically speaking, solvates were observed as a natural outcome of various crystallization techniques, and, therefore, were labeled as accidental and serendipitous results (Almarsson & Zaworotko, 2004; Mohamed & Li, 2018). This trend, particularly in the pharmaceutical

industry, has changed; since regulatory authorities (EMA, 2000; FDA, 2007), nowadays, require an extensive report of crystal forms of an API, solvate discoveries are preferentially obtained during polymorph screening (Byrn et al., 1995; Griesser, 2006).

Solvates are commonly subdivided into stoichiometric and non-stoichiometric classes (Vippagunta et al., 2001). A common differentiating factor of these two classes is that solvent molecules in a non-stoichiometric solvate are often located in structural voids and act as a space filler, whereas in stoichiometric solvates, solvent molecules are often an integral part of the crystal structure and are necessary for maintaining the molecular network. This classification helps us to distinguish the outcome of the ubiquitous concern of solvates: desolvation. Desolvation of stoichiometric solvates usually results in a different crystal structure or an amorphous state while a non-stoichiometric solvate usually tends to retain its crystal structure (Griesser, 2006).

Dasatinib, hereinafter abbreviated as DAS, is a second-generation tyrosine kinase inhibitor that its monohydrate is marketed under the brand name of Sprycel by Bristol-Myers Squibb. In a series of patents, more than 60 solvated forms (Simo et al., 2009; Vraspir et al., 2010; Yan et al., 2010) and 3 anhydrous polymorphs (N6, B, and BM) (Gore et al., 2010; Simo et al., 2009) of DAS have been reported.

The variety of crystalline forms of DAS is partly the consequence of possessing numerous hydrogen bond donor/acceptor moieties. DAS is recognized as a "promiscuous solvate former" (Price et al., 2006) which means the low packing efficiency of the anhydrous form of DAS gives rise to indiscriminate solvate formation. Roy et al., in 2012, for the first time published the crystal structure of the monohydrate and an anhydrate form of DAS (N6) (Roy et al., 2012). During the crystallization experiments and comparison of lattice parameters, they observed that DAS crystallizes in isostructural forms, which is congruent with the model of promiscuous solvate formers.

Note that despite having a large number of crystal forms patented, a few 3D crystal structures of DAS have ever successfully been solved and uploaded to the Cambridge Structural Database (CSD). Besides Roy et al. (Roy et al., 2012), only one other publication (Sarceviča et al., 2016)

has successfully obtained 8 solvates of DAS. With the aid of hierarchical cluster analysis, they showed that most of the DAS solvates were isostructural and fell into the same group (group I) (Sarceviča et al., 2016). This study provided compelling evidence that DAS is an extraordinary promiscuous solvate former.

In the sequential order of reported crystal structures of alcohol solvates (ethanol, isopropanol, and *n*-butanol), methanol (MeOH) is unaccounted for. We suspected that DAS-MeOH crystallization had yielded unstable solvates which instantly desolvated on harvesting. The associated solvent, in such cases, possesses a high degree of volatility. Hence, non-ambient conditions are required to characterize these transient solvates. These challenging scenarios are often overlooked (Griesser, 2006).

Although these unstable solvates may not possess any direct industrial significance, the outcome of their desolvation might be an attractive polymorph. Thus, obtaining polymorphs from solution crystallization of different solvents does not necessarily imply that the polymorph of interest directly nucleates in the solvent. Griesser (2006), in a book chapter on the importance of solvates, pointed out that more attention has to be devoted to these "transient" solvates in order to understand the principle of polymorph formations. DAS-MeOH is especially interesting since it has been reported that grinding of the anhydrous form of DAS (form N6) with MeOH results in a different anhydrous form of DAS, known as form B (Sarceviča et al., 2016).

Another potential benefit of transient solvates is with regard to particle size reduction (Sekiguchi et al., 1968). Transient solvates often easily break into very fine, highly crystalline, and homogeneous powder; therefore, studying transient solvates can offer an alternative technique for particle size reduction where milling processes are problematic (Griesser, 2006). Moreover, desolvation becomes easier when the solvate is unstable.

In this work, we investigated the formation, stability, and characterization of DAS-MeOH solvate and argued the importance of this crystal structure in obtaining the anhydrous polymorphs of DAS as well as its relevance in particle size reduction. In addition, this crystal

has also been compared with the previously reported ethanol solvate of DAS due to its similar characteristics.

2.2   Experimental Section

2.2.1 Materials

Apotex PharmaChem Inc. (Ontario, Canada) generously donated the anhydrous (form N6) of Dasatinib. Solvents, with analytical reagent purity, were purchased from Fisher Chemical (Ontario, Canada) and were used as received.

2.2.2 Sample Preparation

Single crystal of DAS-MeOH for structure identification was obtained by slow evaporation technique. About 5 mg of DAS was dissolved in methanol to create a saturated solution. The vial was then stored in a refrigerator with a controlled temperature of 3-5℃ for 2 weeks until single crystals, kept in the mother liquor, were grown to a suitable size ($\sim$>50 μm for the smallest edge).

The powder samples were produced by 2 techniques: mortar and pestle grinding and ball milling. About 200 mg of DAS was transferred to an agate mortar and pestle and a few droplets of MeOH were added at 10 minutes intervals. The procedure of manual grinding lasted 30 minutes. For ball milling, 400 mg of DAS with one bulb of MeOH ($\sim$2 ml) was mixed; the resultant paste was put in an air-tight ball mill jar (PTFE SmartSnap Jar – 5 ml) with two 5 mm Zirconia grinding balls. Retsch Mixer Mill MM 200 (Haan, Germany) was then used to mix this setup for 30 mins at 12 Hz frequency.

2.2.3 Powder X-ray Diffraction Characterization

The Powder X-ray Diffraction (PXRD) spectra were collected using a Rigaku-MiniFlex II (Carlsbad, California) benchtop diffractometer using Cu Kα radiation (1.54059 Å) at 30 kV and 15 mA The diffractograms were analyzed with JADE 7.0 software (*JADE 7.0*, 2020).

2.2.4 Single-crystal X-ray Diffraction

The data was collected on a Bruker APEX-II CCD diffractometer, using Mo Kα radiation (λ = 0.71073 Å) with a graphite monochromator, at 110 K. APEX2 and SAINT software (Bruker, 2012) was utilized for data integration and reduction, respectively. The crystal structure of DAS-MeOH was solved by direct methods using SIR2014 (Burla et al., 2015) program. Structure refinement was done in the program package of WinGX (Farrugia, 2012) and *OLEX2* (Dolomanov et al., 2009). Anisotropic refinement of all non-hydrogen atoms was performed on SHELXL (Sheldrick, 2015) by full matrix-least squares calculations based on F2. All protons connected to C, N, O were fixed. Mercury 2020.1 (Macrae et al., 2020) was utilized for the crystal structures and packing analysis and visualization.

2.2.5 Differential Scanning Calorimetry

To study the phase transitions of the polymorphic crystals, a Mettler Toledo DSC 822e differential scanning calorimeter (Greifensee, Switzerland) was used. About 5 mg of samples were accurately weighed and placed in a standard 40 μl aluminum crucible, covered with a pinhole lid. An identical empty pan was used as a reference. The data was collected for the temperature range of 25 to 300℃ with a constant heating rate of 5℃/min, under an ultra-high purity nitrogen purge.

To capture desolvation, we needed to use a different setup: the standard 40 μl aluminum crucibles were kept open. 1-hour isotherm step helped us to keep the crucibles at -50℃. Single crystals with trace amounts of mother liquor were added to the sample crucible. After reaching thermal equilibrium, we gently removed the excess methanol with a Kimwipe, waited another 15 minutes, and then started the heating process. Samples were heated from -50 to 300℃ at a 1℃/min rate.

2.2.6 Hot-stage Optical Microscopy

Hot-stage Microscopy (HSM) was performed on a Linkam THMS 600 hot-stage (Tadworth, UK) and temperature was controlled with a Linkam TNS 94 programmable heater and a Linkam LNP (liquid nitrogen pump) cooler. The hot-stage was coupled with a Zeiss Axioskop

40 microscope (Oberkochen, Germany) with an attached CCD image sensor. Crystals were scanned from -40 to 20℃ at a 0.2℃/min rate.

2.2.7 Particle Size Distribution

The particle size distributions (PSD) were determined using the Malvern Mastersizer 2000 (Worcestershire, UK) laser particle size analyzer. An adequate amount of DAS was dissolved in distilled water to obtain a saturated solution. After 6 hours of mixing with a magnetic stirrer, the saturated solution was filtered, and fresh powder were used for PSD measurement. First, the background noise was measured, and then, about 20 mg of fresh DAS powder was added to 300 ml of saturated solution and the measurements were taken. The amount of added sample was based on the in-built diagram of the Malvern software for the recommended levels of obscuration. The measurement was repeated several times to ensure the results are reproduceable.

2.2.8 Crystal Structure Analyses

CrystalExplorer17 (Turner et al., 2017) was used to generate Hirshfeld surface and two-dimensional fingerprint plot to study the intermolecular interactions. Hirshfeld surface (Spackman & Byrom, 1997) is a powerful tool for studying interactions while preserving a whole-of-molecule approach. The distance of the closest atom (d_i for inside, and d_e for outside atoms), to each point on the Hirshfeld surface can also be used to generate a two-dimensional fingerprint plot that summarizes the interactions (McKinnon et al., 2007).

The molecular electrostatic potential map (ESP) of DAS in the crystal structure is mapped to $0.002 \, e^-/Å^3$ electron density isosurface, computed by Gaussian 16 (Frisch et al., 2016) using DFT B3LYP level of theory and 6-311++G** basis set.

2.3 Results and Discussion

2.3.1 Obtaining New Solvate Form (DAS-MeOH)

The starting material for DAS was an anhydrous polymorph, referred to as the N6 form. During preliminary attempts to obtain DAS-MeOH single crystals by slow evaporation in a refrigerator of a supersaturated solution of DAS in MeOH, single crystals turned into a white

powder, immediately after leaving the mother liquor at room temperature. After several attempts, however, we were able to obtain the single crystal data of DAS-MeOH by immediately mounting the crystal from the mother liquor, as shown in Figure 2-1.
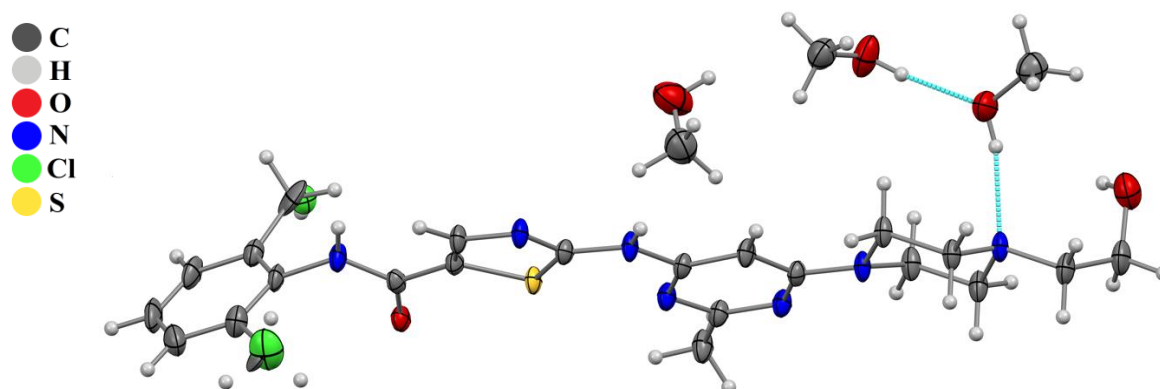
Figure 2-1 ORTEP diagram of the asymmetric unit of DAS-MeOH solvate with 50% ellipsoid probability.

DAS-MeOH crystallizes in $P2_1/n$ and it constitutes of one DAS and three MeOH molecules in the asymmetric unit. The unit cell dimensions are $a = 17.673(4)$ Å, $b = 8.2835(16)$ Å, and $c = 21.405(3)$ Å, with the β of 109.922(12)°. More detailed information about this crystal structure can be found in Table 2-1.

Table 2-1 Crystal data and structure refinement for DAS-MeOH.

Identification code	DAS-MeOH
CCDC number	2055291
Empirical formula	$C_{25}H_{38}ClN_7O_5S$
Formula weight	584.13
Temperature	110(2) K
Wavelength	0.71073 Å
Crystal system	Monoclinic
Space group	$P2_1/n$
Unit cell dimensions	a = 17.673(4) Å
	b = 8.2835(16) Å
	c = 21.405(3) Å
	α = 90°
	β = 109.922(12)°
	γ = 90°
Volume	2946.1(10) $Å^3$
Z	4
Density (calculated)	1.317 Mg/m^3
Absorption coefficient	0.247 mm^{-1}
F(000)	1240
Crystal size	0.410 x 0.360 x 0.060 mm^3
Theta range for data collection	2.024 to 26.426°.
Index ranges	-22≤h≤22, -10≤k≤10, -26≤l≤26
Reflections collected	54112
Independent reflections	6060 [R(int) = 0.0664]
Completeness to theta = 25.242°	99.9 %
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	6060 / 0 / 374
Goodness-of-fit on F^2	1.041
Final R indices [I>2sigma(I)]	R1 = 0.0492, wR2 = 0.1223
R indices (all data)	R1 = 0.0749, wR2 = 0.1392
Largest diff. peak and hole	0.855 and -0.430 e $Å^{-3}$
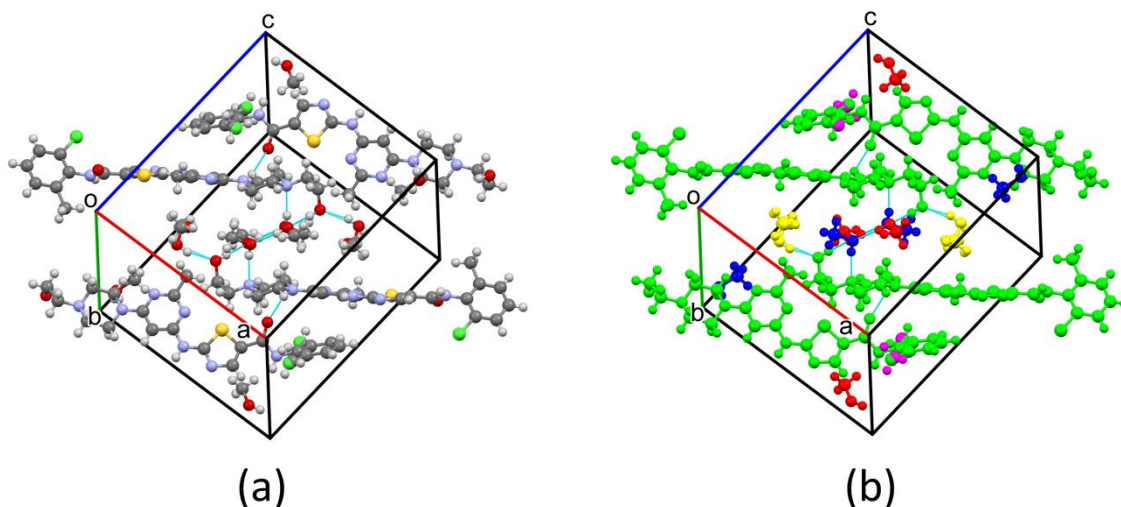
Figure 2-2 Packing of DAS and MeOH molecules in the crystal structure coloured by (a) elements and (b) symmetry equivalence.

The list of all intermolecular interactions of DAS-MeOH crystal are summarized in Table 2-2. One of the methanol molecules, coloured by blue in Figure 2-2.b, interacts with DAS by an O-H···N hydrogen bond (Interaction I). The hydrogen donor of the second methanol, coloured by yellow in Figure 2-2.b, is connected to the alcohol group of DAS (O-H···O) (Interaction II). The red methanol in Figure 2-2.b is attached to the blue methanol as the hydrogen bond donor via O-H···O interaction (Interaction III) while its hydrogen bond acceptor moiety is linked to DAS via O-H···O (Interaction IV). DAS molecules are bound to each other via two main interactions: First is the N-H···O chain along the b axis (Interaction V), and the second is an N-H···N dimer (Interaction VI).

Table 2-2 List of Intermolecular Interactions of DAS-MeOH

Interaction	D-X···A	Symmetry Code	X···A (Å)	D···A (Å)	∠D-X···A (°)
I	O4-H4 ···N1	x, y, z	2.05	2.87	166
II	O3-H3 ···O1	1-x, 1-y, 1-z	1.96	2.77	162
III	O5-H5C ···O4	x, y, z	1.90	2.74	174
IV	O1-H1···O5	1-x, 1-y, 1-z	1.94	2.76	165
V	N7-H7··· O2	-½-x, ½+y, ½-z	1.94	2.77	158
VI	N5-H5··· N6	-x, 1-y, 1-z	2.07	2.95	178

2.3.2 Characterization of Interactions and Packing Efficiencies

Close contacts of DAS in DAS-MeOH crystal are depicted in red on Hirshfeld surface (Figure 2-3.a) and characterized by spikes on the 2D fingerprint plot (Figure 2-3.b). The percentage of $O \cdots H$, $N \cdots H$, $C \cdots H$, and $H \cdots H$ atom pair contributions in the total interactions has also been shown. These figures stress the strong intermolecular interactions between $O \cdots H$ and $N \cdots H$.
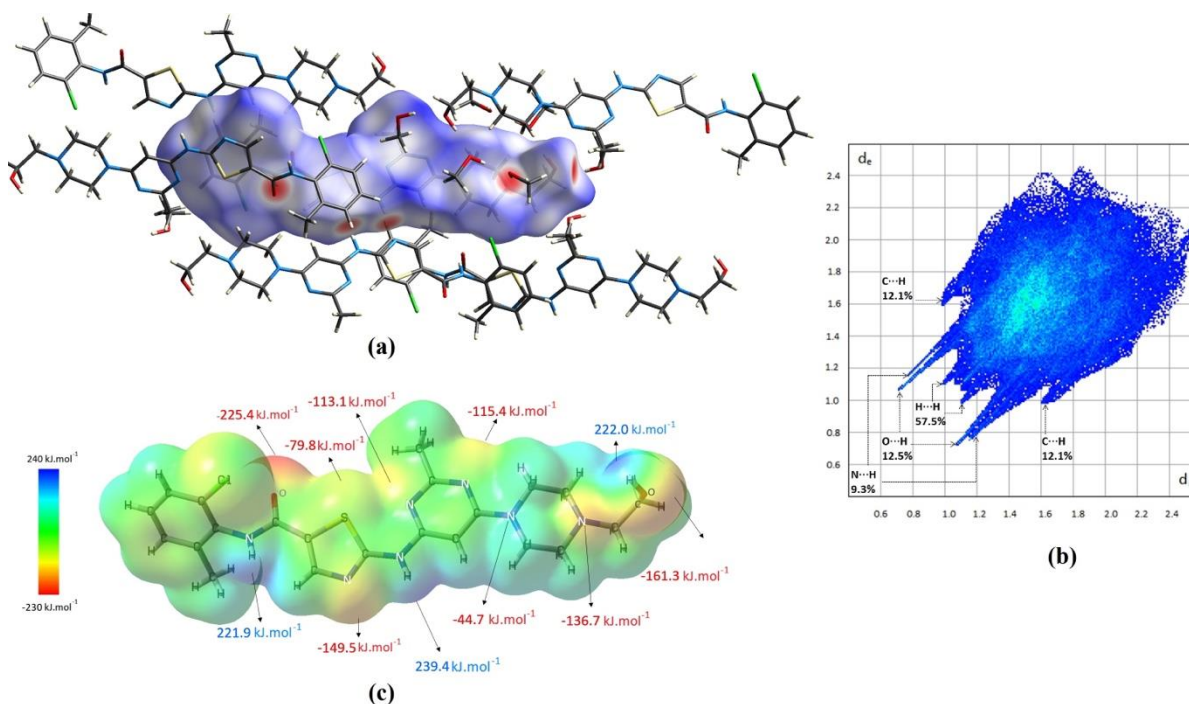
Figure 2-3 Hirshfeld surface (a), fingerprint plot (b), and molecular electrostatic potential map (c) of DAS in DAS-MeOH crystal structure obtained from CrystalExplorer17 and Gaussian 16 (Frisch et al., 2016; Turner et al., 2017).

The molecular electrostatic potential map (ESP) of DAS in the crystal structure is illustrated in Figure 2-3.c. One can see that DAS possesses 3 hydrogen bond donors (coloured in blue) and 8 hydrogen bond acceptors (coloured in red). The hydroxyl (OH) functional group contains the second strongest hydrogen bond donor (222.0 kJ.mol^{-1}) and acceptor (-161.3 kJ.mol^{-1}) of DAS but does not participate in the known crystal structure of the anhydrous form of DAS (Refcode: RAVPUB) (Roy et al., 2012). The tertiary amine nitrogen (-136.7 kJ.mol^{-1}) is another strong hydrogen bond acceptor that is not involved in a hydrogen bond. These 3

hydrogen bonding sites are satisfied by methanol molecules in DAS-MeOH crystal which favors the formation of the solvate over the anhydrous form.

Solvent molecules in a solvate improve the packing efficiency. In order to display and compare the voids in the crystal structures (Figure 2-4), Mercury (Macrae et al., 2020) was used. With the probe radius of 0.8 Å, approximate grid spacing of 0.7 Å, and contact surface method, it was determined that 7.6% of the unit cell volume (188.6 Å^3) of the anhydrous form of DAS was taken up by voids. On the other hand, only 1.5% (43.9 Å^3) of DAS-MeOH was unoccupied that is an indicator of improved packing efficiency.

A striking observation was that after manually removing the atom information of solvent (MeOH) molecules from the CIF file, a well-defined pattern of void channels was observed (Figure 2-4.c) that occupied up 25.9% (762.1 Å^3) of the unit cell. This helps us to vividly observe the spatial distribution of solvent molecules in the crystal packing. Consequently, desolvation of DAS-MeOH should be markedly facile due to the existence of these organized methanol channels.

<div align="center">(a)   (b)   (c)</div>
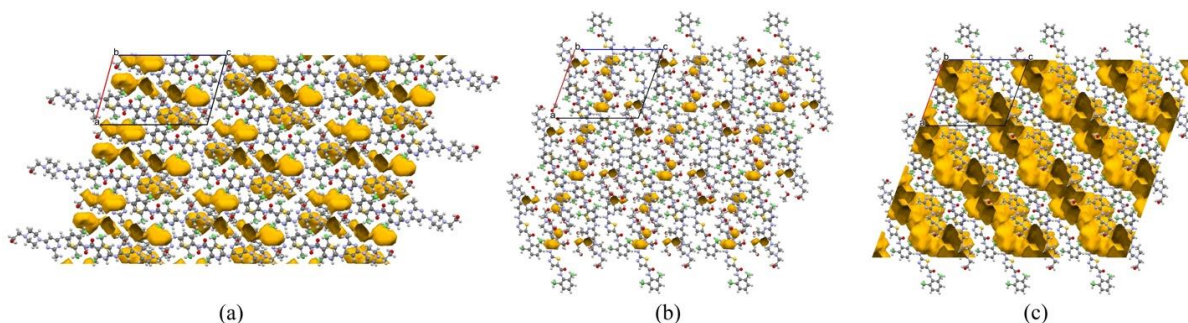
Figure 2-4 Void maps along the b-axis of (a) anhydrous DAS, (b) DAS-MeOH, and (c) DAS-MeOH after deleting atom information of solvent molecules from the CIF file, on a 3*3*3 unit cell packing, obtained from Mercury (Macrae et al., 2020).

DAS-MeOH resembles the DAS-EtOH crystals, showing 3-dimensional isostructurality. DAS-EtOH crystal (Refcode: IGAHUV01) (Sarceviča et al., 2016) unit cell dimensions (a = 17.0090(8) Å, b = 8.8401(4) Å, c = 21.3823(7) Å, β=111.008(2)°) matches DAS-MeOH crystal (Table 2-1). They both belong to the same crystal system (monoclinic) and occupy a

comparable volume (2946.1 $Å^3$ for DAS-MeOH and 3001.37 $Å^3$ for DAS-EtOH). To quantify the degree of isostructurality of DAS-MeOH and DAS-EtOH, we utilized CrystalCMP (Rohliček et al., 2016). This program uses distance/angle displacement of the largest molecules to find the best overlay of crystal structures. Inclusion of all non-hydrogen atoms of DAS molecule showed that the combined RMSD (root-mean-squared deviation) between these two crystal structures is 0.9182. These isostructural packings are visualized in Figure 2-5 along the *b*-axis.
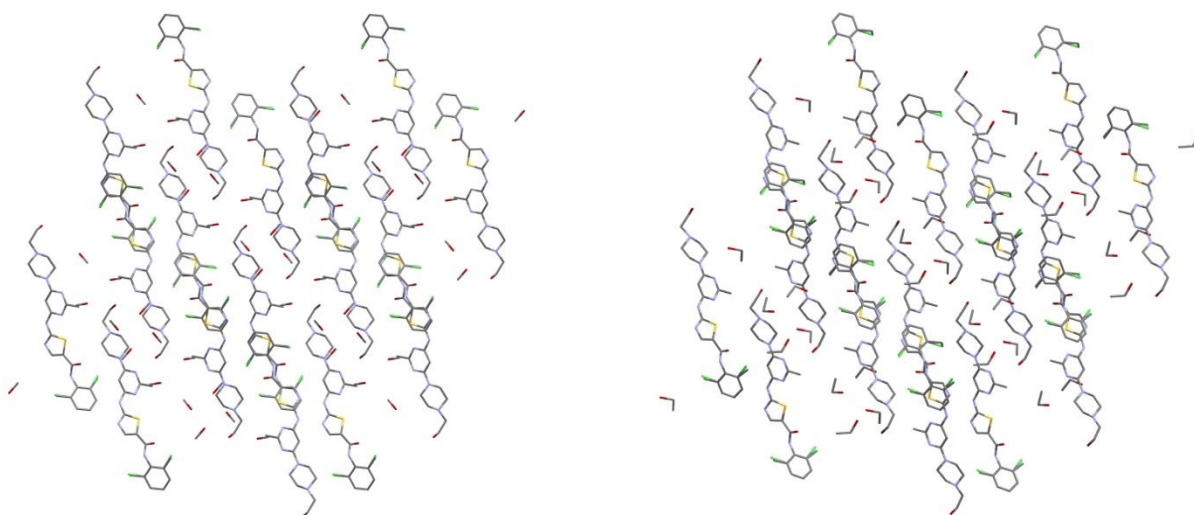
Figure 2-5 Packing similarity of DAS-MeOH (left) and DAS-EtOH (right) crystals.

2.3.3 Stability Analysis of DAS-MeOH Crystals

It was evident that the DAS-MeOH crystal was not stable at room temperature. To further characterize this crystal and determine its thermal stability, hot stage microscopy (HSM) was performed to determine at which temperature the desolvation started and how it evolved as temperature rose. Crystals with trace amounts of mother liquor were mounted on the HSM plate and kept at -35℃ for half an hour. After ensuring all of the surface methanol was evaporated and crystals were dried, we heated the crystal at a rate of 0.2℃/min.
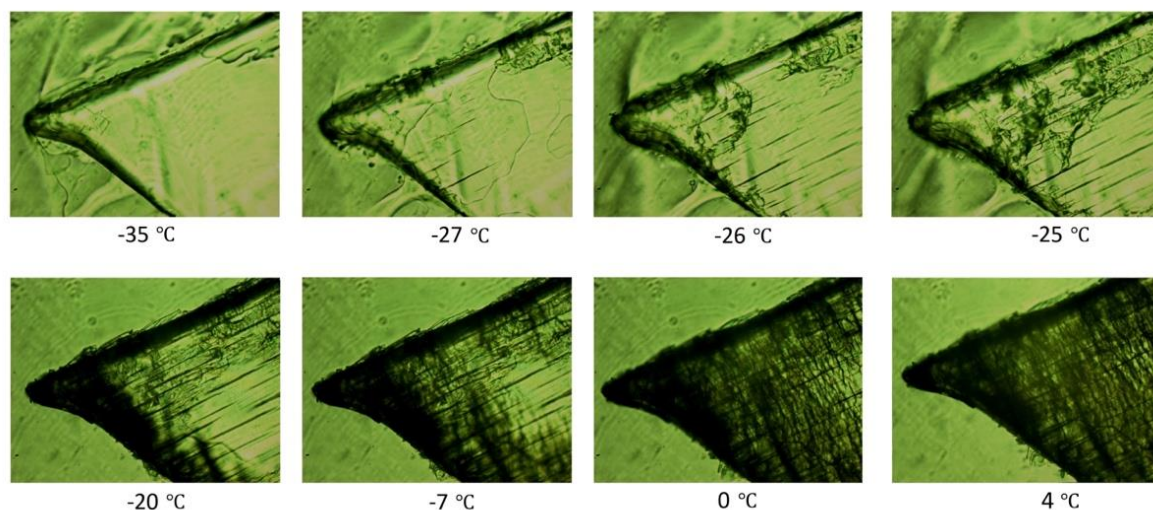
Figure 2-6 The evolution of DAS-MeOH crystal in HSM experiment under magnification power of 50X. These microphotographs demonstrate the desolvation upon heating.

As shown in Figure 2-6, DAS-MeOH crystals were stable until the first crack started at -27℃. These directional fractures were quickly evolved over a span of 2 degrees. At -20℃, another series of crystal breakage, perpendicular to the first pattern, started occurring, causing the crystal to gradually become opaque. As mentioned before, pronounced structural changes are associated with desolvation of stoichiometric solvates; primarily, recrystallization into minute crystals. One can see that at temperatures higher than 0℃, the methanolate crystals are fully converted to DAS form B. We also performed the same experiment with DAS-EtOH crystals to compare the stability of these isostructural crystals.
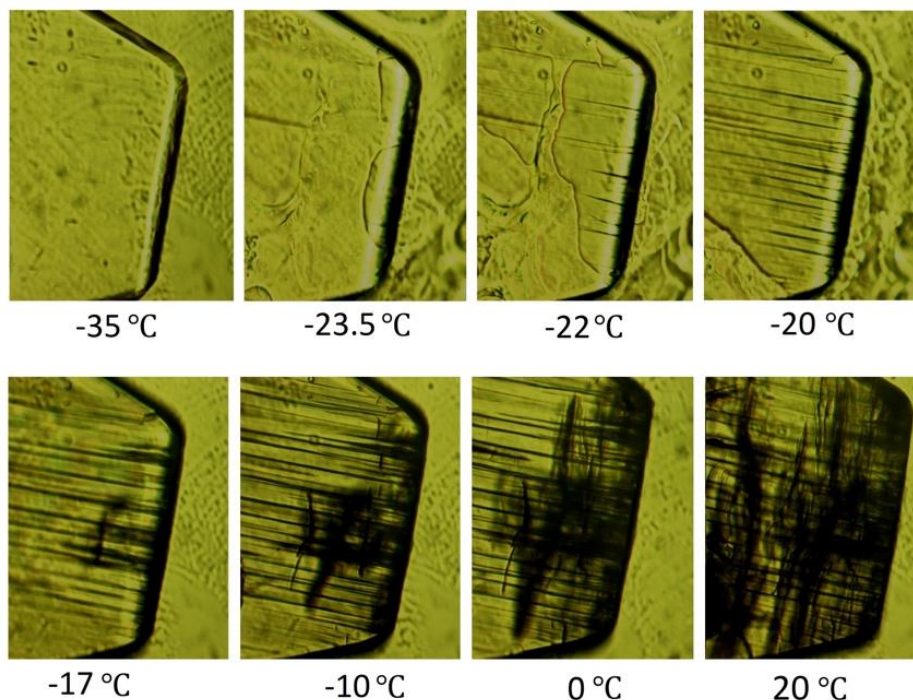
Figure 2-7 The evolution of DAS-EtOH crystal in HSM experiment under magnification power of 50X. These microphotographs demonstrate the desolvation upon heating.

The same development as DAS-MeOH was observed for DAS-EtOH crystals, with a clear shift in the crystal stability temperature range. As depicted in Figure 2-7, deformation started at -23.5℃, and small pockets of ethanol leaving the crystal are vividly observed. These directional surface desolvation continued until -20℃. Further removal of solvate molecules was spotted at -17℃ and continued up to the point that the desolvation terminated.

DSC can also capture the desolvation process. The samples of DAS-MeOH and DAS-EtOH were prepared according to the procedure described in the experimental section. Samples were slowly (1℃/min) heated from -50 to 300℃ and the graphs are shown in Figure 2-8. DAS-MeOH shows three peaks for the desolvation process; it starts with a broad and subtle peak, followed by two more marked peaks. Just below 10℃, the plot reaches a plateau and finally decomposes at 277℃. Note that DAS does not melt, and the main DSC peak at 277℃ represents the irreversible decomposition of the compound. DAS-EtOH desolvates with a similar pattern but at higher temperatures. The main desolvation peak of DAS-EtOH ends at about 30℃.
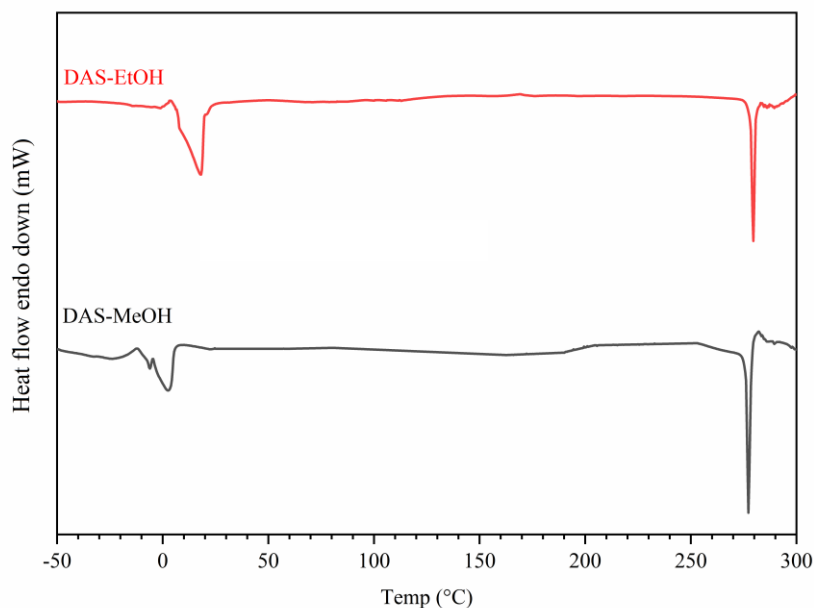
Figure 2-8 DAS-MeOH and DAS-EtOH solvates DSC graphs from -50℃ to 300℃ at the rate of 1℃/min.

2.3.4 Determining the Outcome of Desolvation

The real interest of the transient solvate of DAS-MeOH is due to the final state of the DAS crystals after complete desolvation. In fact, the instability of the methanolate crystal can be an advantage by entailing the high purity of the final desolvated product.

As previously mentioned, grinding of DAS with MeOH results in a new anhydrous crystalline form (form B) of DAS (Sarceviča et al., 2016). To illustrate if the mechanism of this polymorphic transition is DAS-MeOH desolvation or direct nucleation of form B from the paste, we employed different crystal preparation techniques. First and foremost, wet single crystals were located on a PXRD sample holder and their pattern was obtained (red line on Figure 2-9). We observe that the experimental and simulated (black line) pattern from Mercury (Macrae et al., 2020) share the same general pattern, particularly the unique peak at 5.6°. Immediately after desolvation, the pattern significantly changes, and the crystals convert to the anhydrous form B (blue curve).
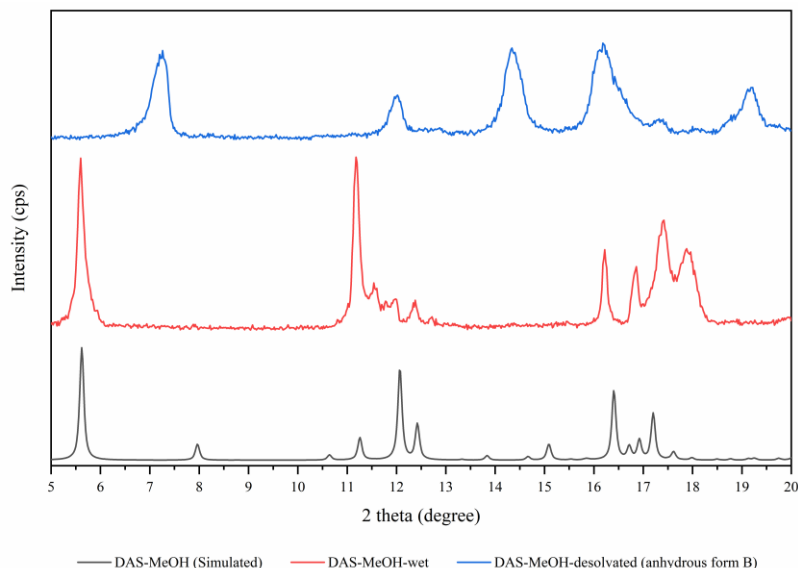
Figure 2-9 PXRD patterns of simulated DAS-MeOH (black), experimental wet crystals (red), and desolvated crystals (blue).

We further investigated the interactions of DAS and MeOH in mortar and pestle grinding and ball milling. The outcome of mortar and pestle grinding, even though a new pattern (form B) was observed, showed broad peaks with relatively low intensities. The PXRD pattern is illustrated by the blue line in Figure 2-10. To improve the peak strength of polymorph B, liquid assisted ball-mill grinding was carried out. After drying the sample under vacuum for 48 hours, the PXRD was collected and the same polymorph B with notably sharper peaks and crystal uniformity was obtained (green line in Figure 2-10). Using the profile fitting option in JADE software, the diffraction peaks were fitted to the pseudo-Voigt profile. FWHM (Full width at half maximum) of each peak was used to calculate the crystallite size using the Scherrer (Scherrer, 1918) equation. The average crystallite size of form B of DAS obtained from mortar and pestle grinding and ball milling were 19 and 25 nm, respectively. In principle, smaller crystallite size, not to be confused with particle size, results in broadening of PXRD peaks that can explain why mortar and pestle grinding exhibited smaller and blunter peaks. Comparably, the raw material (DAS form N6) showed an average crystallite size of 47 nm.

Another noteworthy finding was that the PXRD pattern of the ball mill experiment could also result in another anhydrous polymorph of DAS (form BM), provided that the drying is done

immediately. The obtained PXRD pattern for form BM is shown with the purple line in Figure 2-10.
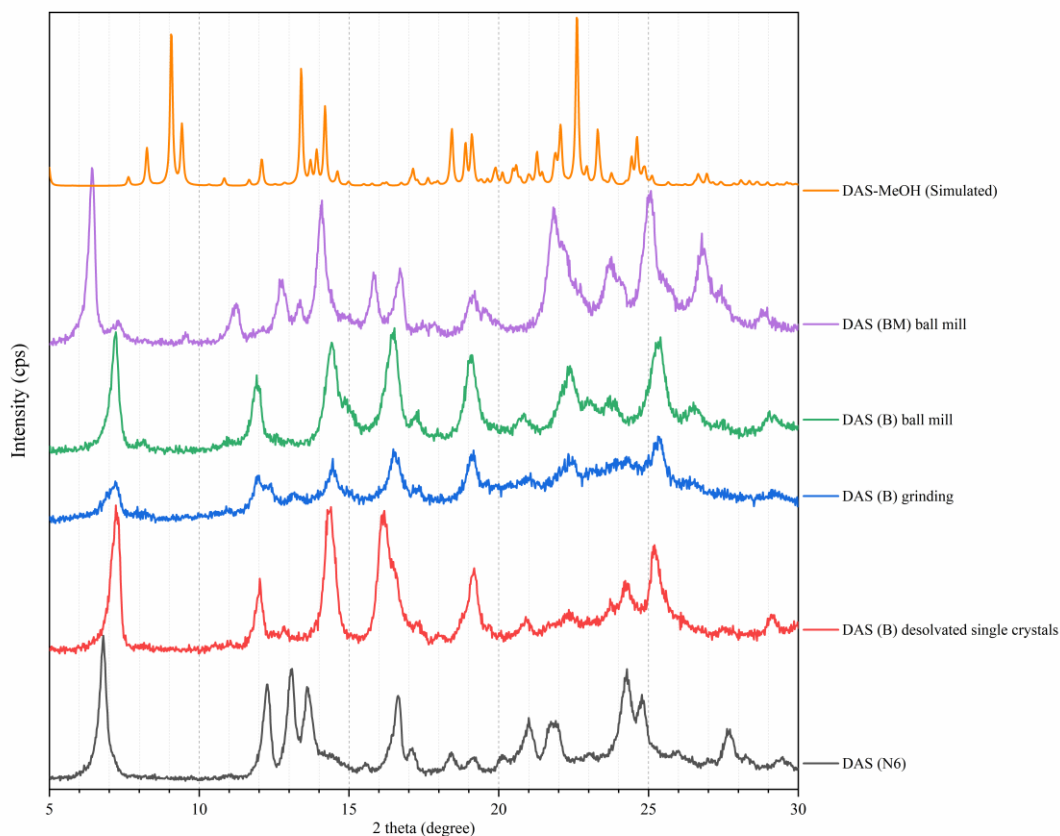
Figure 2-10 PXRD patterns of anhydrous forms of DAS and the simulated pattern of DAS-MeOH.

Since our ball mill experiment ended up with two different polymorphs, we attempted to find out the stability of form BM and captured its conversion to form B.

Four characteristic peaks of form BM (at 6.4, 12.7, 14.0, and 25.0 degrees) and four characteristic peaks of form B (at 7.1, 11.8, 14.3, and 25.2 degrees) were selected and shown by vertical lines in Figure 2-11. The associated diffraction angles with each polymorph are differentiated with colours for clarity. One can see that the initial pattern (red) to a large extent lacks the characteristic peaks of form B but has pronounced intensities of form BM. Over the course of 3 hours, gradually the solid-state phase transformation occurs and conversion to form

B completes. Note that after a week the PXRD pattern of the form B crystal remained unchanged. Therefore, form B can be considered as the ultimate desolvation product of DAS-MeOH crystals.
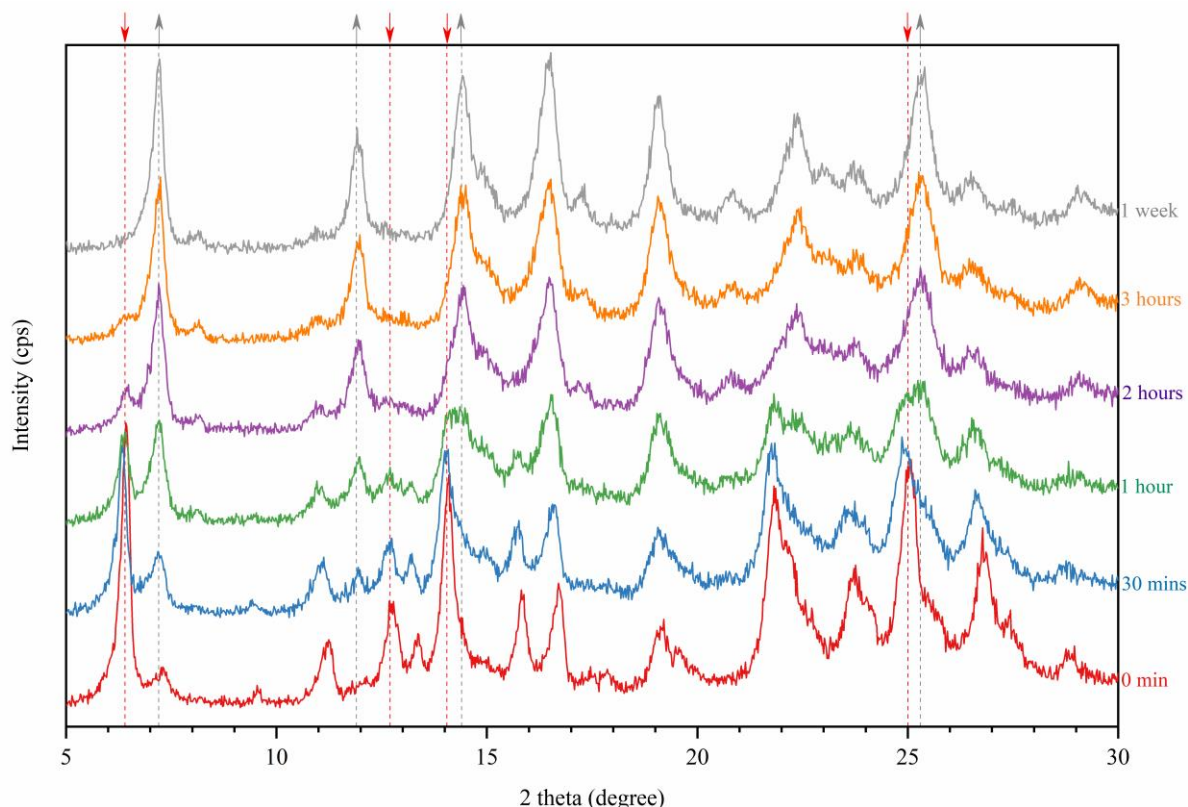
Figure 2-11 PXRD patterns showing the solid-state phase transition of DAS form BM (red line) to form B (gray line).

DSC analysis was performed to provide additional support for polymorph determination. To reemphasize, DAS does not melt, and the main DSC peaks shown in Figure 2-12 represent the irreversible decomposition of the compound. DAS form N6 (starting compound) decomposes at the highest temperature (286℃) compared to others. Decomposition of form B, obtained from either manual grinding or ball milling, onsets at 280℃. The unstable form BM decomposition temperature lies between the other two forms, starting at about 283℃.
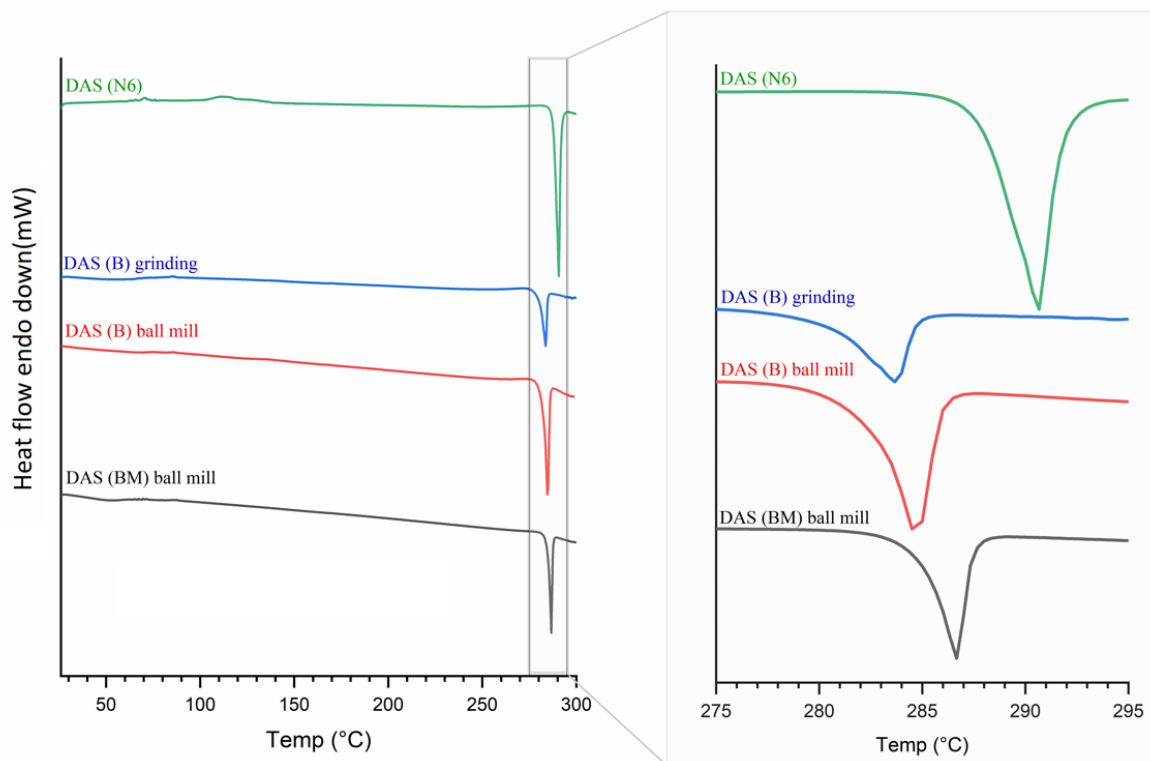
Figure 2-12 DSC results of the anhydrous polymorphs of DAS (N6, B, and BM).

To explain why different milling techniques underwent different paths of polymorphic transformations, access to the crystal structure of all forms, including unavailable structures of form B and BM, is needed. However, there are some fundamental differences between mortar and pestle grinding and ball milling that may cause the observed difference: mortar and pestle grinding induces molecular diffusion and, depending on the exerted pressure, increases the local temperature well above the melting point. On the other hand, ball milling is a more controlled process in which the heat generated by the mechanical shaking can be dissipated (Braga et al., 2013). In addition, ball mill jars are sealed and solvent does not escape during the process while in manual grinding methanol evaporates.

2.3.5 Particle Size Distribution (PSD)

One advantage of the ball mill technique was the uniformity of the particles compared to the mortar and pestle grinding. To accent one potential use of transient solvates, we mention that transient solvates can offer an alternative technique for particle size reduction where dry milling processes are problematic. To quantify our qualitative observation, particle size distribution was determined using the Malvern Mastersizer 2000 laser particle size analyzer.

As one can see in Figure 2-13, the commercial form of DAS (N6) has a monomodal distribution with a median of about 25 μm. Grinding in a mortar and pestle caused a significant polydispersity in PSD with a median of 1.5 μm. This undesired effect has been significantly improved by the ball mill technique; the monomodal distribution of crystals was recovered and the PSD was shifted to smaller particles with a median of less than half ($\sim$ 12 μm) of the commercial DAS. Since DAS has a low bioavailability, proposed particle size reduction can provide improved physicochemical properties of this drug, without compromising its uniformity.
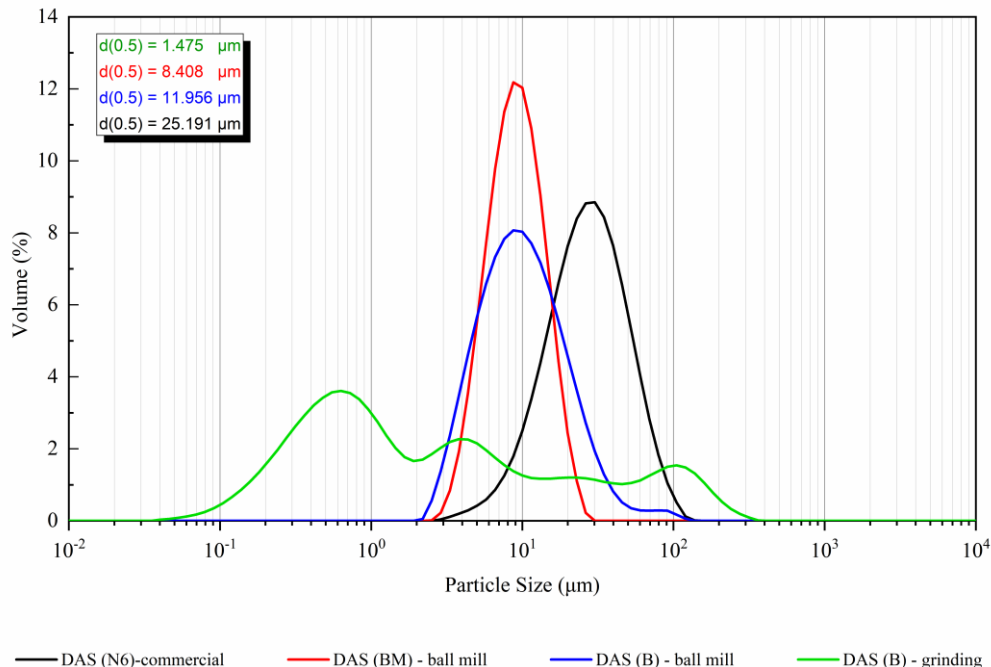
Figure 2-13 The particle size distribution of DAS polymorphs. Median, d(0.5), of each distribution is also shown.

Since the product of desolvation is a different polymorph than the starting material, particle size reduction cannot be justified without a discussion on the relative stability of these two polymorphs. In a 2009 patent (Simo et al., 2009), it has been noted that for temperatures 20°C or below, form B is relatively stable. They have utilized the slurry crystallization technique, in which form N6 was converted to form B over the course of two days at temperature of 20°C to 0°C. Consequently, we can conclude that there is no risk associated with form B being converted to form N6 during downstream processes.

2.4 Conclusions

DAS-MeOH has been thoroughly studied in this research. This transient solvate immediately breaks down outside of the mother liquor, except for low temperatures ($< -30°C$). After obtaining the crystal structure of DAS-MeOH for the first time by SCXRD, Hirshfeld surface and ESP map were used to study the interactions. Additionally, void map analysis was utilized to disclose the transient nature of this crystal. The structure was discovered to be favorable in energy but the arrangement of solvent molecules in well-defined channels facilitated the desolvation process at room temperature. Employing HSM and DSC, the stability of this solvate has been determined. Upon desolvation of single crystals, very fine but crystalline powder of an anhydrous form of DAS (form B) was achieved. Mechanical grinding of DAS with MeOH also eventually yielded form B; however, the characteristics of the crystals were significantly different. Mortar and pestle grinding led to polydisperse size distribution whereas ball milling allowed very fine (~ 12 μm) and highly crystalline product. The path which eventually ended to form B in ball milling, nonetheless, was not directly from desolvation: another anhydrous form of DAS (form BM) was initially formed that spontaneously transformed to form B over the course of 3 hours.

2.5 References

Almarsson, Ö., & 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*, *17*, 1889–1896.

Braga, D., Maini, L., & Grepioni, F. (2013). Mechanochemical preparation of co-crystals. *Chemical Society Reviews*, *42*(18), 7638–7648.

Bruker. (2012). *APEX2, SAINT*. Bruker AXS Inc., Madison, Wisconsin, USA.

Burla, M. C., Caliandro, R., Carrozzini, B., Cascarano, G. L., Cuocci, C., Giacovazzo, C., Mallamo, M., Mazzone, A., & Polidori, G. (2015). Crystal structure determination and refinement via SIR2014. *Journal of Applied Crystallography*, *48*(1), 306–309.

Byrn, S., Pfeiffer, R., Ganey, M., Hoiberg, C., & Poochikian, G. (1995). Pharmaceutical Solids: A Strategic Approach to Regulatory Considerations. *Pharmaceutical Research*, *12*(7), 945–954.

Desiraju, G. R., Vittal, J. J., & Ramanan, A. (2011). Multi-component Crystals. In *Crystal Engineering* (pp. 131–153). World Scientific Publishing Company.

Dolomanov, O. V, Bourhis, L. J., Gildea, R. J., Howard, J. A. K., & Puschmann, H. (2009). OLEX2: a complete structure solution, refinement and analysis program. *Journal of Applied Crystallography*, *42*(2), 339–341.

Duggirala, N. K., Perry, M. L., Almarsson, Ö., & Zaworotko, M. J. (2016). Pharmaceutical cocrystals: along the path to improved medicines. *Chemical Communications*, *52*(4), 640–655.

EMA. (2000). *ICH Topic Q 6 A Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances*. European Medicines Agency (EMA). www.ema.europa.eu/en/documents/scientific-guideline/ich-q-6-test-procedures-acceptance-criteria-new-drug-substances-new-drug-products-chemical_en.pdf

Farrugia, L. J. (2012). WinGX and ORTEP for Windows: an update. *Journal of Applied Crystallography*, *45*(4), 849–854.

FDA. (2007). *ANDAs:Pharmaceutical Solid Polymorphism: Chemistry, Manufacturing, and Controls Information (Guidance for Industry)*. Food and Drug Administration (FDA). www.fda.gov/regulatory-information/search-fda-guidance-documents/andaspharmaceutical-solid-polymorphism-chemistry-manufacturing-and-controls-information

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. a., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. a., Nakatsuji, H., Li, X., Caricato, M., Marenich, a. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., … Fox, D. J. (2016). *Gaussian 16, Revision b.01* (Gaussian 16, Revision b01). Gaussian, Inc., Wallin CT.

Gore, V. G., Patkar, L., Bagul, A., Vijaykar, P. S., & Edake, Mahesh. (2010). *Preparation of crystalline forms of dasatinib.* (Patent No. WO2010139981A2). World Intellectual Property Organization.

Griesser, U. J. (2006). The Importance of Solvates. In *Polymorphism: in the Pharmaceutical Industry* (pp. 211–233).

Healy, A. M., Worku, Z. A., Kumar, D., & Madi, A. M. (2017). Pharmaceutical solvates, hydrates and amorphous forms: A special emphasis on cocrystals. *Advanced Drug Delivery Reviews*, *117*, 25–46.

JADE 7.0. (2020). Materials Data: Livermore, CA.

Macrae, C. F., Sovago, I., Cottrell, S. J., Galek, P. T. A., McCabe, P., Pidcock, E., Platings, M., Shields, G. P., Stevens, J. S., Towler, M., & Wood, P. A. (2020). Mercury 4.0: from visualization to analysis, design and prediction. *Journal of Applied Crystallography*, *53*(1), 226–235.

McKinnon, J. J., Jayatilaka, D., & Spackman, M. A. (2007). Towards quantitative analysis of intermolecular interactions with Hirshfeld surfaces. *Chemical Communications*, *37*, 3814–3816.

Mir, N. A., Dubey, R., & Desiraju, G. R. (2019). Strategy and Methodology in the Synthesis of Multicomponent Molecular Solids: The Quest for Higher Cocrystals. *Accounts of Chemical Research*, *52*(8), 2210–2220.

Mohamed, S., & Li, L. (2018). From serendipity to supramolecular design: assessing the utility of computed crystal form landscapes in inferring the risks of crystal hydration in carboxylic acids. *CrystEngComm*, *20*(39), 6026–6039.

Mondal, P. K., Rao, V., Mittapalli, S., & Chopra, D. (2017). Exploring Solid State Diversity and Solution Characteristics in a Fluorine-Containing Drug Riluzole. *Crystal Growth & Design*, *17*(4), 1938–1946.

Price, C. P., Glick, G. D., & Matzger, A. J. (2006). Dissecting the Behavior of a Promiscuous Solvate Former. *Angewandte Chemie International Edition*, *45*(13), 2062–2066.

Rohliček, J., Skovrepová, E., Babor, M., & Čejka, J. (2016). CrystalCMP: an easy-to-use tool for fast comparison of molecular packing. *Journal of Applied Crystallography*, *49*(6), 2172–2183.

Roy, S., Quiñones, R., & Matzger, A. J. (2012). Structural and Physicochemical Aspects of Dasatinib Hydrate and Anhydrate Phases. *Crystal Growth & Design*, *12*(4), 2122–2126.

Sarceviča, I., Grante, I., Belyakov, S., Rekis, T., Bērziņš, K., Actiņš, A., & Orola, L. (2016). Solvates of Dasatinib: Diversity and Isostructurality. *Journal of Pharmaceutical Sciences*, *105*(4), 1489–1495.

Scherrer, P. (1918). Bestimmung der Größe und der inneren Struktur von Kolloidteilchen mittels Röntgenstrahlen. *Nachrichten von Der Gesellschaft Der Wissenschaften Zu Göttingen, Mathematisch-Physikalische Klasse*, *2*, 98–100.

Sekiguchi, K., Horikoshi, I., & Himuro, I. (1968). Studies on the method of size reduction of medicinal compounds. 3. Size reduction of griseofulvin by solvation and desolvation method using chloroform. (3). *Chemical & Pharmaceutical Bulletin*, *16*(12), 2495–2502.

Sheldrick, G. M. (2015). Crystal structure refinement with SHELXL. *Acta Crystallographica Section C*, *71*(1), 3–8.

Simo, O., Filipcik, J., Martaus, A., Jegorov, A., Gavenda, A., Aronhime, J., Vraspir, P., Koltai, T., Faustmann, J., & Gabriel, Roman. (2009). *Polymorphs of dasatinib and process for preparation thereof.* (Patent No. WO2009053854A2). World Intellectual Property Organization.

Spackman, M. A., & Byrom, P. G. (1997). A novel definition of a molecule in a crystal. *Chemical Physics Letters*, *267*(3), 215–220.

Stieger, N. (2012). Recrystallization of Active Pharmaceutical Ingredients. In *Crystallization-Science and Technology* (pp. 183–204). IntechOpen.

Turner, M. J., McKinnon, J. J., Wolff, S. K., Grimwood, D. J., Spackman, P. R., Jayatilaka, D., & Spackman, M. A. (2017). *CrystalExplorer17*. University of Western Australia.

Vippagunta, S. R., Brittain, H. G., & Grant, D. J. W. (2001). Crystalline solids. *Advanced Drug Delivery Reviews*, *48*(1), 3–26.

Vraspir, P., Gavenda, A., Gabriel, R., & Jegorov, Alexandr. (2010). *Polymorphs of dasatinib solvates.* (Patent No. WO2010062715A2). World Intellectual Property Organization.

Yan, R., Yang, H., & Xu, Yongxiang. (2010). *Dasatinib polymorphic substances and their preparation process.* (Patent No. CN101891738A). China National Intellectual Property Administration.

Chapter 3

Overcoming the Hydrophobic Nature of Zinc Phenylacetate through Co-crystallization with Isonicotinamide

3 Overcoming the Hydrophobic Nature of Zinc Phenylacetate through Co-crystallization with Isonicotinamide

Abstract

Zinc phenylacetate (Zn-PA), a substitute for sodium phenylacetate as an ammonia-scavenging drug is hydrophobic, which poses problems for drug dissolution and solubility. We were able to co-crystallize the zinc phenylacetate with isonicotinamide (INAM) and produce a novel crystalline compound (Zn-PA-INAM). The single crystal of this new crystal was obtained, and its structure is reported here for the first time. Zn-PA-INAM was characterized computationally by *ab initio*, Hirshfeld calculations, CLP-PIXEL lattice energy calculation, and BFDH morphology analysis, and experimentally by PXRD, SCXRD, FTIR, DSC, and TGA analyses. Structural and vibrational analyses showed a major modification in intermolecular interaction of Zn-PA-INAM compared to Zn-PA. The dispersion-based pi-stacking in Zn-PA is replaced by coulomb-polarization effect of hydrogen bonds. As a result, Zn-PA-INAM is hydrophilic, improving the wettability and powder dissolution of the target compound in an aqueous solution. Morphology analysis revealed, unlike Zn-PA, Zn-PA-INAM has polar groups exposed on its prominent crystalline faces, reducing the hydrophobicity of the crystal. The shift in average water droplet contact angle from 128.1° (Zn-PA) to 27.1° (Zn-PA-INAM) is strong evidence of a marked decrease in hydrophobicity of the target compound. Finally, HPLC was used to obtain the dissolution profile and solubility of Zn-PA-INAM compared to Zn-PA.

3.1 Introduction

Liver is the primary organ responsible for the nitrogen balance in human body, excreting the excess nitrogen via the urea cycle. Patients with genetic liver failure may experience toxic elevated levels of ammonia which severely affects their central nervous system. Ammonia-scavenging drugs (e.g. benzoate, phenylacetate, and phenylbutyrate) provide alternative pathways (Batshaw et al., 1982) for nitrogen, which regulate hepatic nitrogen metabolism, and lower blood ammonia levels (De Las Heras et al., 2017).

Sodium salts of benzoate, phenylacetate, and phenylbutyrate have been the primary formulation for these drugs. Sodium benzoate is conjugated with glycine which can potentially remove 1 mole of waste nitrogen for every mole of benzoate administered. Sodium phenylacetate and sodium phenylbutyrate, on the other hand, are conjugated with glutamine, and have the potential to remove 2 moles of waste nitrogen for every mole of administered drug (Walker, 2009).

Sodium benzoate and sodium phenylacetate are administered intravenously with the recommended dose of 250 mg/kg infused over 90 min, followed by another 250 mg/kg infused over 24 h (Praphanproj et al., 2000). Significant sodium load of the therapeutic dose of these drugs can exacerbate comorbid cardiac dysfunction (Brouillard et al., 2019; Morrison & Ness, 2011) and ascites (Wong, 2012) that are common in patients with liver failure. Therefore, substituting Na^+ with counterions such as Mg^{2+} and Zn^{2+} can lower the sodium intake and its side effects.

Zinc is one of the most biologically essential trace elements, taking part in many biological processes, ranging from cell growth, division, and activation to DNA synthesis and RNA transcription (Chasapis et al., 2012). Therefore, zinc phenylacetate (Zn-PA) (Ali et al., 2014; Liu et al., 2018) can potentially be an ideal substitute for sodium phenylacetate. However, our experimental observations during the course of this project confirmed that Zn-PA is highly hydrophobic with low solubility in water.

To tackle the hydrophobicity issue, crystal engineering can provide a viable solution. The field of crystal engineering focuses on alteration of the solid-state chemistry by means of intermolecular interactions in the design strategy (Ahmadi et al., 2021; Desiraju, 2007, 2013) Within the pharmaceutical industry (Berry & Steed, 2017), addition of a second ionic component as pharmaceutical salts has been commonplace for over 80 years (Berge et al., 1977; Bharate, 2021). In the last two decades (Almarsson et al., 2012), addition of a non-ionic second component has also become an established practice, forming pharmaceutical cocrystals (Almarsson & Zaworotko, 2004). The focus of this article is on ionic cocrystals, where three components (a cation, an anion, and a neutral molecule) bind in a unique crystal structure (T. Wang et al., 2018).

Multicomponent crystals have found various applications in pharmaceutical industry (Bolla et al., 2022). Bolla & Nangia classified these applications in three categories: physicochemical properties (solubility, stability, hydration, and the melting point), mechanical properties (hardness and compressibility), and pharmacokinetic properties (bioavailability and permeability) (Bolla & Nangia, 2016).

Bioavailability improvement, including solubility and membrane permeability, is one of the major areas of research where multicomponent crystals offer significant advantages. Theophylline-aminobenzoic acid (Saikia et al., 2015), meloxicam-aspirin (Cheney et al., 2011), quercetin-caffeine (Smith et al., 2011) are some examples of cocrystals that introduced marked changes in solubility, dissolution rate, or pharmacokinetic parameters (such as C_{max}, T_{max}, and AUC) of the parent API (Bolla et al., 2022).

Hydration stability is another criterion that cocrystals can offer positive changes. Some APIs spontaneously form hydrates under humid conditions but their cocrystals may inhibit the hydration process and improve stability of the API (Bolla et al., 2022). Acemetacin (Sanphui et al., 2014), etoricoxib (Mittapalli et al., 2016), niclosamide (Sanphui et al., 2012), and nitrofurantoin (Cherukuvada et al., 2011) are all examples of APIs that their hydration was controlled through cocrystallization.

By forming stable layered crystalline solids, cocrystals can also improve tableting properties of APIs (Bolla et al., 2022). Paracetamol form 1 when cocrystallized with theophylline constructs a layered structure with high tensile strength and favourable tablet-forming properties (Karki et al., 2009). Temozolomide-hesperetin (J. Wang et al., 2021), nicorandil-suberic acid (Mannava et al., 2021), and nicorandil-oxalic acid (Mannava et al., 2021) are other examples of multicomponent crystals which show better tabletability. In this article, preparation, and properties of a new multicompetent crystal (Zn-PA-INAM) are discussed, in which Zinc (Zn) is the cation, phenylacetate (PA) is the anion, and isonicotinamide (INAM) is the neutral molecule. Addition of INAM resulted in significant decrease in hydrophobicity of the Zn-PA.

3.2 Methodology

3.2.1 Sample Preparation

Zinc Phenylacetate (Zn-PA) was received from TDC Research Inc. Isonicotinamide (INAM) (>99%) was purchased from Alfa Aesar and was used as purchased. Liquid-assisted grinding (LAG) was used by mechanical grinding of the Zn-PA and a list of coformers in an agate mortar and pestle with the addition of 1 bulb of methanol. Ball milling with a Retsch Mixer Miller 200 (on 12 Hz vibrational frequency) in a 5mL PTFE SmartSnap Jar with two 5 mm Zirconia balls was performed parallelly to prepare additional powder. In both LAG approaches, grinding was done for 30 minutes.

In total, 15 organic compounds have been considered for synthesis of a new multicomponent crystal with Zn-PA: catechol, resorcinol, hydroquinone, benzoic acid, fumaric acid, malic acid, maleic acid, malonic acid, succinic acid, oxalic acid, nicotinamide, isonicotinamide, isonicotinic acid, melamine, and imidazole. These compounds were selected based on synthon considerations. Even though more rigorous binary cocrystal prediction algorithms have been developed based on hydrogen bonding propensity (Delori et al., 2013) and energy (Musumeci et al., 2011) calculations, as well as machine learning algorithms (Ahmadi et al., 2021) such as graph convolutional neural networks (Devogelaer et al., 2020), prediction and validation of ionic cocrystals remains an underexplored area of research (Shunnar et al., 2020). Coformers, in this work, are selected based on their potential to form various supramolecular synthons, defined as structural units formed by arrangement of intermolecular interactions (Desiraju, 1995). The potential synthons (Kavuru et al., 2010; Li et al., 2018; Nanubolu et al., 2013; Ortega et al., 2018) that the carboxylate group of the API can form with the coformers listed in this work are shown in Figure 3-1.
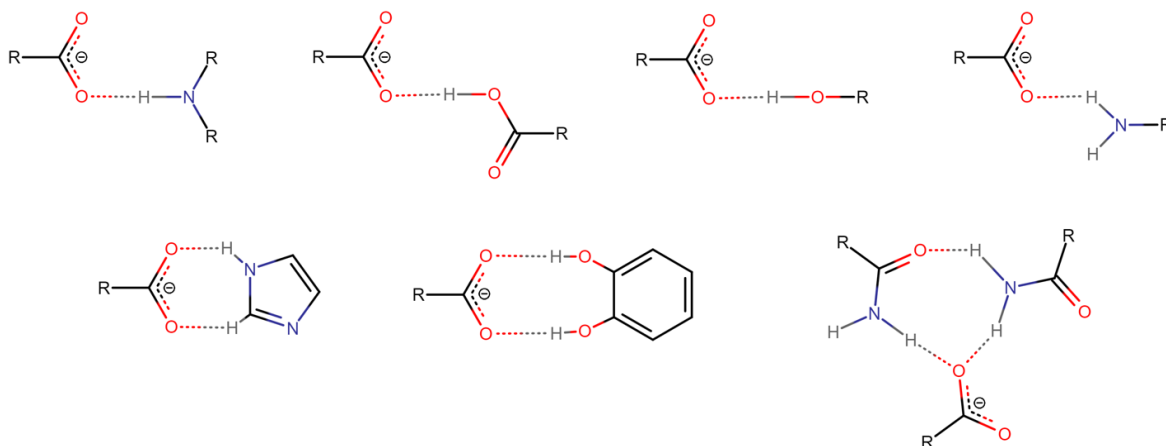
Figure 3-1 Potential supramolecular synthons carboxylate group can form with the selected coformers in this work.

The chemical composition of the result of each LAG experiment was analyzed with PXRD and DSC and was classified as:

a) physical mixture (neither Zn-PA nor the coformer was altered),

b) amorphous/gelation (reaction led to gelation and crystalline powder could not be isolated),

c) ligand substitution (the coformer replaced phenylacetate, forming phenylacetic acid),

d) co-crystallization (both Zn-PA and coformer were combined in a new unit cell).

3.2.2 Single Crystal X-ray Diffraction

To obtain single crystals of Zn-PA-INAM, slow evaporation, liquid-liquid diffusion, and vapour diffusion were considered with 11 solvents: water, methanol, ethanol, isopropanol, tetrahydrofuran (THF), acetonitrile, toluene, hexane, chloroform, dimethylformamide (DMF), and dimethyl sulfoxide (DMSO). Successful crystals were only obtained by slow evaporation of THF. 5 mg of the compound was dissolved in 5 ml of THF and was kept in the fridge (4℃) for a month until block crystals were formed.

Bruker APEX-II CCD diffractometer with Mo Kα radiation ($\lambda = 0.7107$ Å) was used for data collection at 110(2) K, APEX2 software (Bruker, 2012) for data integration, and SAINT program (Bruker, 2012) for data reduction. Crystal structure was solved by direct methods

using SIR 2014 (Burla et al., 2015) and then refined in Olex2 1.3 (Dolomanov et al., 2009), with full-matrix least-squares method SHELXL (Sheldrick, 2015). TWINABS-Version 2012/1-Bruker AXS scaling was used for deconvolution of data from a twinned single crystal. All the non-hydrogen atoms were refined anisotropically while hydrogen atoms were placed in the calculated positions. Structural and packing analyses were performed on the CCDC Mercury program (Macrae et al., 2020).

3.2.3 Powder X-ray Diffraction

Powder X-ray diffractograms were collected on a Rigaku MiniFlex II benchtop diffractometer with a Cu Kα source (1.54059Å) at 30 kV and 15 mA. About 100 mg powder of each sample was placed on a front load glass sample holder and scanned for 2θ range of 5°-50° at the rate of 3°/min. Data was analyzed on JADE 7.0 software (*JADE 7.0*, 2020).

3.2.4 Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) was used for thermal analysis of the samples. 5-10 mg of each sample was put in a 40 μl aluminum crucible, covered with a pinhole lid. The experiments were carried out for the temperature range of 25−350°C with a 5°C/min heating rate on a Mettler Toledo DSC 822e under nitrogen gas atmosphere.

3.2.5 High-performance Liquid Chromatography

High-performance liquid chromatography (HPLC) technique was used the determine the solution concentration of phenylacetate for various compounds. For each run, samples were withdrawn from the vial, filtered with Chromspec syringe filters (13 mm PTFE, 0.2μm pore size), diluted with deionized water, and injected (10 μl) into the HPLC instrument. The stationary phase was a C18 Varian (Chrompack) column (250×4.6 mm) and the mobile phase was HPLC grade water (Caledon Laboratories) at a fixed flowrate of 1.2 mL/min. The UV detector was set to 215 nm to characterize phenylacetate. A standard calibration curve was generated by running the HPLC experiment for a series of known concentrations of phenylacetate.

3.2.6 Solubility and Powder Dissolution Setup

A phosphate buffer (0.1 M) was prepared by dissolving 24.237 g of potassium phosphate dibasic (K_2HPO_4) and 8.281 g of potassium phosphate monobasic (KH_2PO_4) in 2 liters of HPLC grade water. The resulting buffer had a pH of 7.3. An excess amount (~ 400 mg) of powder samples was added to 15 mL of aqueous buffer solution. The vials were gently mixed with an AC Tech Intelligent Drive MC Series M1103SB at 3.2 Hz throughout the experiment and kept at 37°C in an Endocal RTE-220 heat bath. Samples were withdrawn periodically, filtered with Chromspec syringe filters (13 mm PTFE, 0.2μm pore size), diluted with deionized water, and their concentrations were measured with HPLC.

3.2.7 Fourier-transform Infrared Spectroscopy

The functional group interactions and modes of carboxylate binding to the Zinc were studied with FTIR. Powder samples were analyzed using a Bruker Tensor II spectrometer by placing them on the diamond crystal of the Platinum® attenuated total reflectance (Pt-ATR) accessory. The spectra were collected from 4000-400 cm^{-1}, with 16 scans and a resolution of 4 cm^{-1}. The spectra were corrected for the contribution from water vapour and carbon dioxide and were baseline corrected.

3.2.8 Contact Angle Measurement

Optical contact angle measurement, using the sessile drop method, was performed on an OCA 30 (DataPhysics Instruments) instrument, consisting of a sample stage, light source, lens, and camera. First, about 2g of each sample was hydraulically pressed to 6 tons (Carver Laboratory Press) to create tablets with 5 cm^2 smooth surfaces and avoid the interference of surface roughness in the measurements. Tablets were immediately transferred for contact angle measurement. 10 μl of water was deposited on the surface by the automatic droplet dispensing unit of the OCA 30 instrument and the photos were captured immediately by a high-speed CCD camera connected to a computer. The droplet contact profiles on the collected images were analyzed using the SCA 20 software, which comes with the instrument. The experiment was repeated three times, and the contact angles were averaged.

3.2.9 Computational Crystal Structure Analyses

CrystalExplorer17 (Turner et al., 2017) was used to generate a Hirshfeld surface and two-dimensional fingerprint plot to study the intermolecular interactions. *ab initio* calculations were performed on Gaussian 16 (Frisch et al., 2016). The atomic coordinates were obtained from the experimental crystal structure, then geometry optimization was done only on hydrogen atoms to maintain the integrity of the crystal structure, followed by frequency calculation on the optimized structure. The optimized structure had no imaginary frequency. Second-order Møller-Plesset perturbation theory with 6-31G** basis set (MP2/6-31G**) was used (Frisch et al., 1990; Head-Gordon et al., 1988). The molecular electrostatic potential map (ESP) is mapped to a 0.002 $e^- \, Å^{-3}$ electron density isosurface (Ahmadi et al., 2021).

CLP-PIXEL (Gavezzotti, 2005) is a program that calculates the lattice energy of a crystal, dividing the intermolecular interactions into 3 categories: Coulomb-polarization, London dispersion, and repulsion (Pauli exclusion). A discretized molecular electron grid, generated by Gaussian 16 (MP2/6-31G** level of theory), was used as an input to CLP-PIXEL to calculate the intermolecular interactions in the crystal structure. The PIXEL method based on MP2/6-31G** calculations has proven to reproduce the experimental heats of sublimation of organic crystal structures and has the great advantages of in principle being applicable throughout the periodic table of elements while remaining computationally efficient (Gavezzotti, 2005).

3.3 Result and Discussion

3.3.1 Structural Analysis of Zn-PA-INAM

From the list of organic compounds selected for the initial screening with PXRD analysis, summarized in Table 3-1, only isonicotinamide (INAM) proved to form a new multicomponent crystal (Zn-PA-INAM), visualized in Figure 3-2. Methanol was used as the LAG solvent for all coformers below. Since each mole of Zn-PA has 2 moles of PA, the molar ratio of 1:2 was selected as the most expected API:coformer molar ratio. For dicarboxylic acids, the ratio of 1:1 was also tested.

Table 3-1 Experimental Screening of 15 organic coformers with Zn-PA

Coformer	Molar Ratio (API:coformer)	Outcome
benzoic acid	1:2	a) physical mixture
catechol	1:2	a) physical mixture
fumaric acid	1:1, 1:2	a) physical mixture
hydroquinone	1:2	a) physical mixture
imidazole	1:2	b) amorphous/gelation
isonicotinamide	1:2	**d) co-crystallization**
isonicotinic acid	1:2	a) physical mixture
melamine	1:2	a) physical mixture
maleic acid	1:1, 1:2	c) ligand substitution
malic acid	1:1, 1:2	b) amorphous/gelation
malonic acid	1:1, 1:2	b) amorphous/gelation
nicotinamide	1:2	b) amorphous/gelation
oxalic acid	1:1, 1:2	c) ligand substitution
resorcinol	1:2	a) physical mixture
succinic acid	1:1, 1:2	a) physical mixture
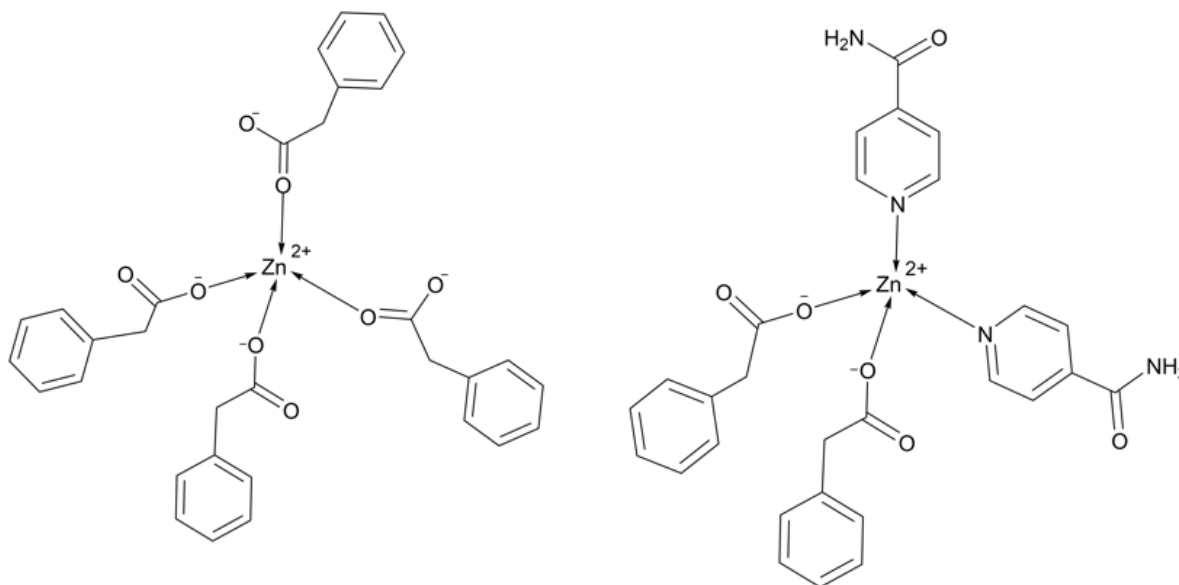
Figure 3-2 Structural representation of Zn-PA (left) and Zn-PA-INAM (right).

The PXRD pattern of the newly-found Zn-PA-INAM is compared with the starting compounds (ZN-PA, and INAM) as well as Phenylacetic acid and the simulated PXRD pattern obtained from SCXRD.
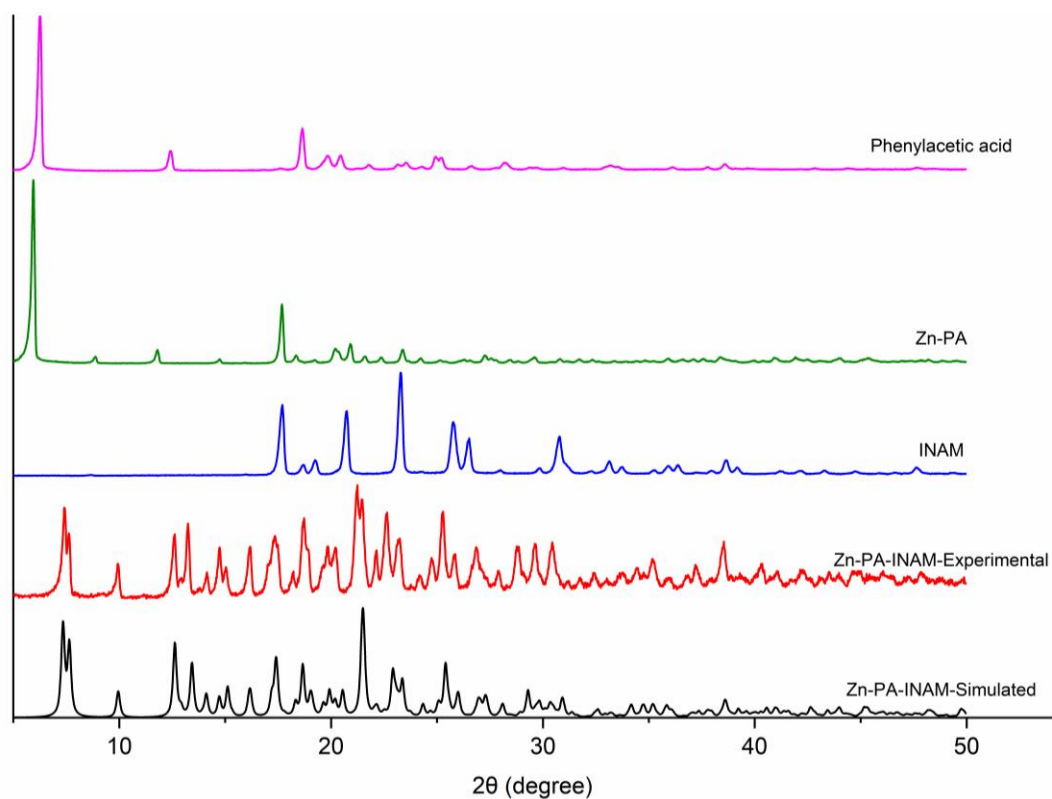
Figure 3-3 PXRD pattern of Zn-PA-INAM and the relevant compounds

Figure 3-3 indicates that the PXRD pattern of experimentally-prepared Zn-PA-INAM does not overlap with any of the starting materials, and in fact, it resembles the simulated pattern we later obtained from the crystal structure solution (SCXRD).

Zn-PA-INAM crystallizes in tetrahydrofuran (THF) solvent with the *P2₁/c* space group with two PA ions, two INAM molecules, and one Zn ion in the asymmetric unit. More detailed information about this crystal structure can be found in Table 3-2.

Table 3-2 Crystal data and structure refinement for Zn-PA-INAM.

Identification code	Zn-PA-INAM
CCDC number	2201551
Empirical formula	$C_{28}H_{26}N_4O_6Zn$
Formula weight	579.90
Temperature	110(2) K
Wavelength	0.71073 Å
Crystal system	Monoclinic
Space group	$P2_1/c$
Unit cell dimensions	a = 8.309(2) Å
	b = 13.172(3) Å
	c = 24.151(5) Å
	α = 90°
	β = 95.338(6)°
	γ = 90°
Volume	2631.7(10) Å^3
Z	4
Density (calculated)	1.464 Mg/m^3
F(000)	1200
Crystal size	0.455 x 0.163 x 0.045 mm^3
Theta range for data collection	1.694 to 26.746°.
Index ranges	-10≤h≤10, -16≤k≤16,
	-30≤l≤30
Reflections collected	81475
Independent reflections	7364 [R(int) = 0.0997]
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	5588 / 0 / 352
Goodness-of-fit on F^2	1.081
Final R indices [I>2sigma(I)]	R1 = 0.0666, wR2 = 0.0859
Largest diff. peak and hole	1.067 and -0.904 e Å^{-3}
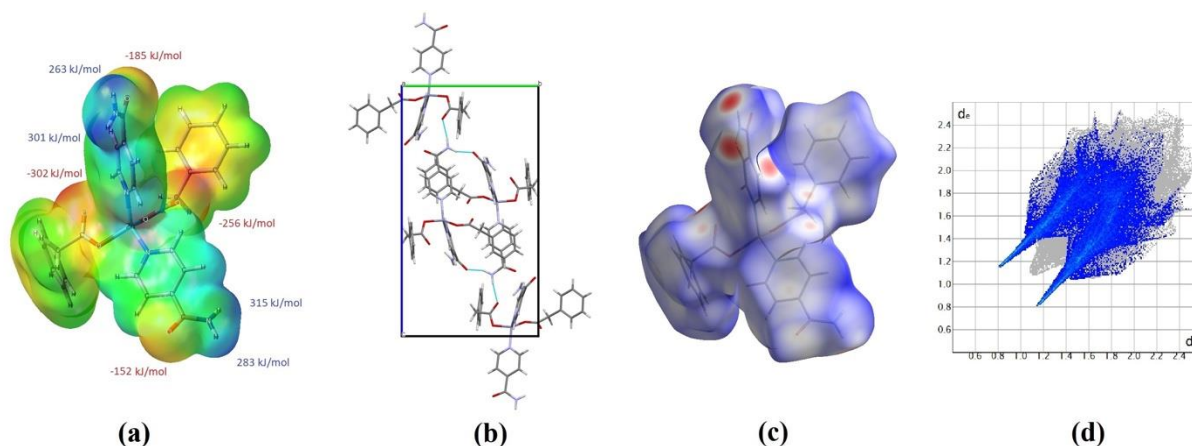
Figure 3-4 Zn-PA-INAM crystal structure analysis and characterization with (a) molecular electrostatic potential map, (b) hydrogen bonding (cyan bonds) in the unit cell (c) Hirshfeld surface, and (d) fingerprint plot with O···H interactions highlighted with blue, obtained from Gaussian 16, Mercury, and CrystalExplorer17.

Figure 3-4.a illustrates the ESP of Zn-PA-INAM with its local minima (red) and maxima (blue) labelled, corresponding to hydrogen bond acceptors and donors, respectively. This crystal has four potential hydrogen bond donors and four acceptors. Red spots in Figure 3-4.c indicates the strongest intermolecular interactions in the solid state which is between hydrogens of NH_2 and various C=O oxygens. Figure 3-4.d confirms that the said O···H hydrogen bonds (blue dots on the fingerprint plot) are the only strong intermolecular interactions present in the crystal structure.
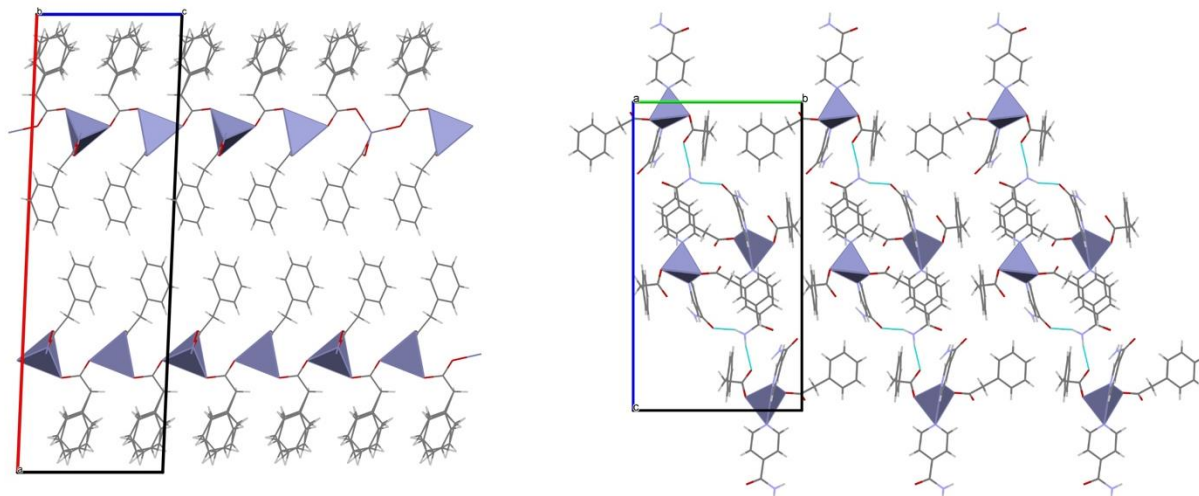
Figure 3-5 Packing of Zn-PA (left) and Zn-PA-INAM (right) crystals, generated in the Mercury program. Zinc ions are shown with polyhedral (demonstrating the tetrahedral nature of their coordination), and everything else is shown with capped stick style. The cyan bonds on Zn-PA-INAM show the hydrogen bondings that are absent in Zn-PA.

With packing analysis, one can see the intermolecular interactions in Zn-PA-INAM are significantly different compared to the starting Zn-PA crystal (Refcode: WEBMEY). Figure 3-5 (left) shows that in Zn-PA crystal packing, Zinc is coordinated with 4 carboxylate oxygens (each contributing half of a negative charge). This coordination mode of carboxylate is referred to as bidentate bridging, connecting 2 Zinc ions (Zeleňák et al., 2007). Figure 3-5 (right) shows Zn-PA-INAM crystal structure. Zinc is still coordinated tetrahedrally, but notably, 2 of the coordinated oxygens are replaced with aromatic nitrogen of INAMs. The resulting binding mode of carboxylate is, therefore, monodentate (Zeleňák et al., 2007). Additionally, the presence of additional hydrogen bond donors/acceptors in the carbamoyl group of INAM is noteworthy. One can see that the cyan bonds, denoting hydrogen bonds, are absent in Zn-PA but form a 3D network in Zn-PA-INAM. This alteration in interactions results in significant structural and physicochemical modifications.

3.3.2 Vibrational Analysis of Zn-PA-INAM

Vibrational spectroscopy is complementary to structural analysis, particularly about the carboxylate modes of binding. Particularly, The Zinc carboxylate bonds can be classified in 4 categories: ionic, monodentate, bidentate chelating, and bidentate bridging (Palacios et al., 2004; Zeleňák et al., 2007).

The Pt-ATR data are illustrated in Figure 3-6. For Zn-PA, the expected symmetric (1419 cm^{-1}) and asymmetric (1527 cm^{-1}) stretching of bidentate bridging of the carboxylate group are labeled. INAM shows two wide peaks in the functional group region that denote symmetric (3176 cm^{-1}) and asymmetric (3362 cm^{-1}) stretching of N-H bond of the carbamoyl group. In the fingerprint region, C=O bond of the carbamoyl group vibrates at 1390 cm^{-1} and 1656 cm^{-1} for its symmetric and asymmetric stretching, respectively. In Zn-PA-INAM, zinc carboxylate bonds are monodentate and they vibrate at 1587 cm^{-1} (asymmetric) and 1419 cm^{-1} (symmetric). It is also shown, that both symmetric (3188 cm^{-1}) and asymmetric (3338 cm^{-1}) stretching vibration of the N-H of the carbamoyl group have shifted in Zn-PA-INAM due to the new hydrogen bonding with the carboxylate group of Zn-PA. The C=O frequency of the carbamoyl group has also shifted in Zn-PA-INAM to 1665 cm^{-1} (asymmetric) and 1382 cm^{-1} (symmetric).

This Pt-ATR data supports the argument made from the structural analysis (illustrated in Figure 3-7) regarding why addition of INAM poses a major alteration in intermolecular interactions in the crystal lattice. Since the powder and single crystals were obtained with different crystallization routes (mechanical grinding with methanol and solvent (THF) evaporation, respectively), the Pt-ATR results are provided to support that we were able to obtain the same product despite using different crystallization routes. Figure 3-8 shows atomic displacement of symmetric/asymmetric stretching of the carboxylates that are visualized by computing force constants and the resulting vibrational frequencies in Gaussian 16.
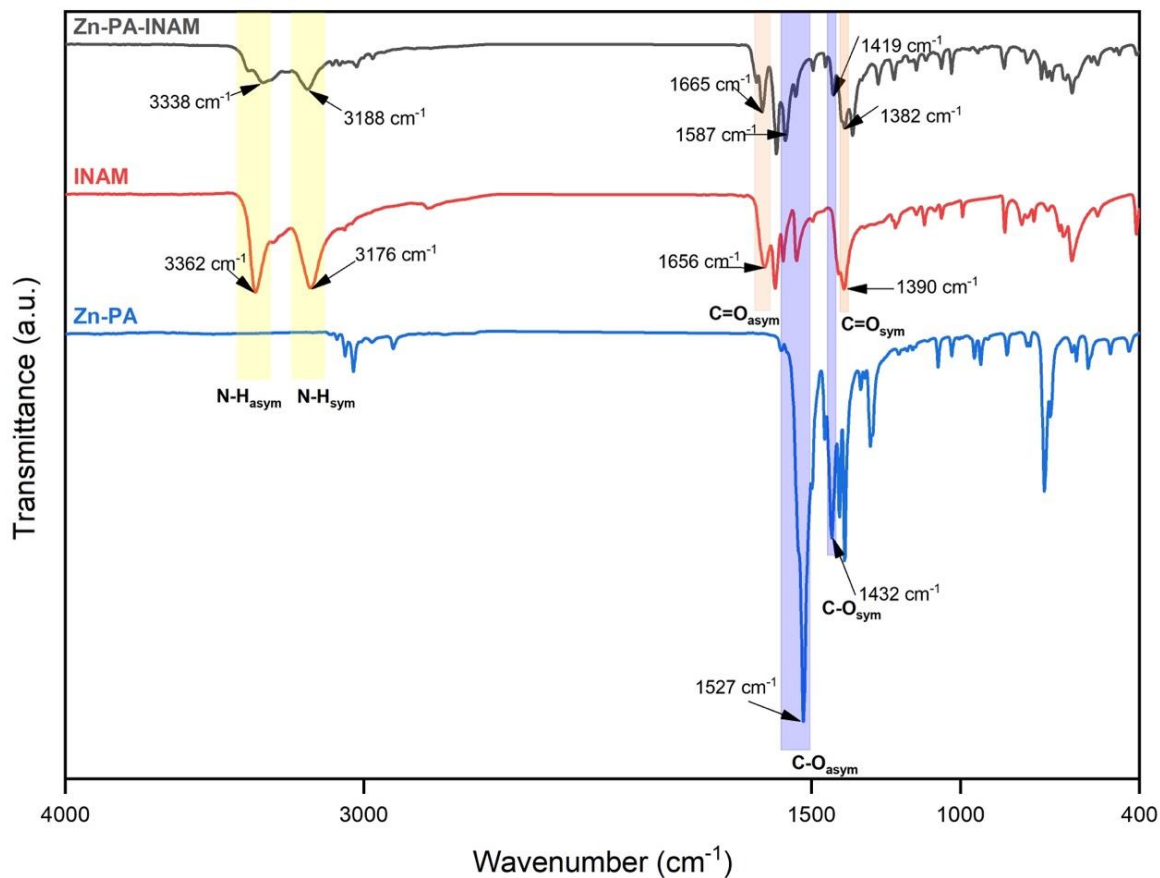
Figure 3-6 Pt-ATR data of Zn-PA-INAM and the starting compounds (Zn-PA and INAM).
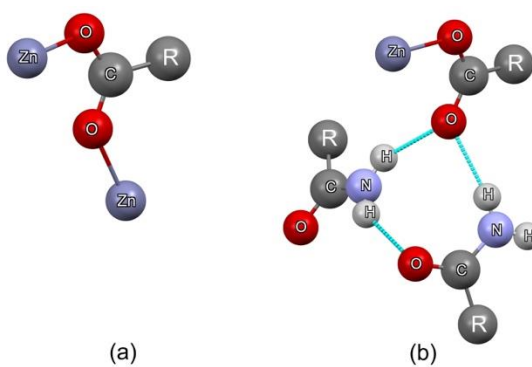
Figure 3-7 The COO⁻ binding modes to the Zn. (a) shows the bidentate bridging with Zn-PA, and (b) shows the monodentate binding in Zn-PA-INAM, leaving one oxygen of COO⁻ free to interact with carbamoyl group via hydrogen bonding. This variation in modes of binding directly translates into vibrational frequencies in Pt-ATR data.
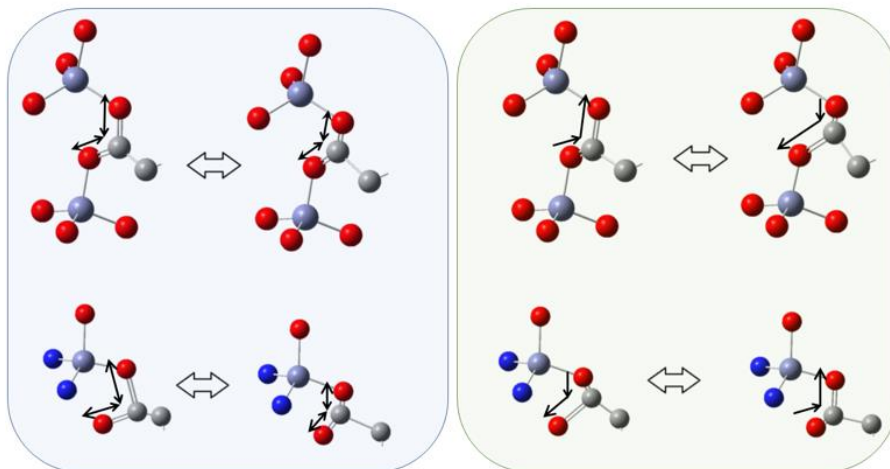
Figure 3-8 Symmetric stretch (left) and asymmetric stretching (right) of Zn-PA (top) and Zn-PA-INAM (bottom).

3.3.3 Contact Angle Comparison

Contact angle determination is a common technique to assess the wettability of a solid surface, such as tablets of pharmaceutical powders (Buckton & Newton, 1985; Karde & Ghoroi, 2014). Wettability is a precursor to dissolution, having significant influence on the dissolution rates and the release characteristics of an API. Work of adhesion (commonly calculated by ignoring the free energy decrease on immersion of the solid in the saturated vapour phase) can also be calculated by the below equation (Karde & Ghoroi, 2014):

$$W_A = \gamma_l^t(1 + \cos\theta)$$

Where, W_A is the work of adhesion, γ_l^t is the surface tension (72.8 mN/m, for water), and θ is the contact angle. Greater work of adhesion signifies more interaction on solid-liquid surface, and therefore, the surface is more wetting (Karde & Ghoroi, 2014).

With sessile drop technique (Figure 3-9), we were able to determine the average contact angle of Zn-PA to be 128.1° and the average contact angle of Zn-PA-INAM to be 27.1° (listed in Table 3-3). Based on the accepted definition (Law, 2014) of hydrophobic (θ>90°) and hydrophilic (θ <90°) surfaces, Zn-PA is hydrophobic but the addition of INAM made the Zn-PA-INAM hydrophilic. This 101° variation in contact angles is strong evidence of the superior

wettability of Zn-PA-INAM. Likewise, the work of adhesion is almost 5 times larger for Zn-PA-INAM compared to Zn-PA (Table 3-3). It is worth noting that the wettability of the tablet might not necessarily be the same as the wettability of individual crystals/particles as the hydraulic press could lead to the preferred orientation of certain facets. However, combined with other analyses presented in this paper, the surface wettability study is strong evidence of major alteration in the hydrophobicity of Zn-PA after cocrystallization with INAM.

Since the tablets were prepared by hydraulic press, comparison between the pre- and post-compression PXRD patterns (reported in Figure A3-6) which signifies the structural integrity of both crystals was maintained after compression. Additionally, the residual solvent in the prepared powder of Zn-PA-INAM could have affected the characterization outcomes. Hence, Thermogravimetric Analysis (TGA), depicted in Figure A3-7 in the supporting information, was performed to confirm the absence of residual solvent in the tablet.
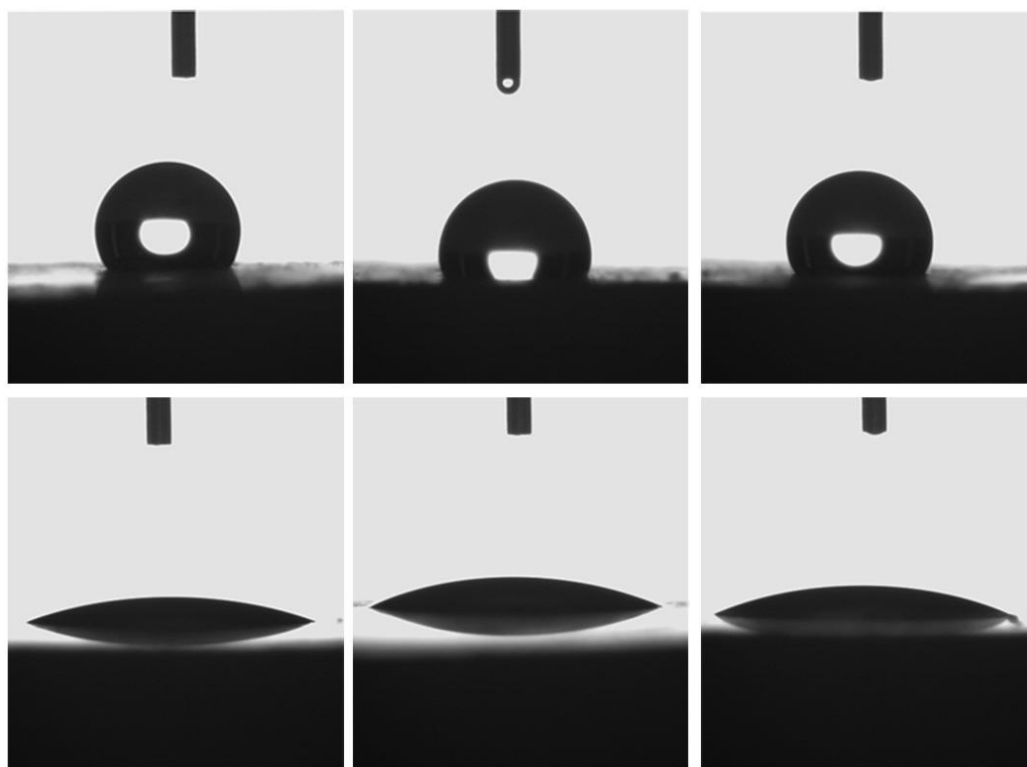
Figure 3-9 Water droplets (10 µl) on three different positions on the surface of tablets of Zn-PA (top row) and Zn-PA-INAM (bottom row).

Table 3-3 Comparison between contact angle and work of adhesion of Zn-PA and Zn-PA-INAM.

Compound	Contact angles (°)				Work of adhesion (mJ/m^2)
	# 1	*# 2*	*# 3*	**Average**	
Zn-PA	*131.4*	*124.0*	*128.9*	**128.1**	**27.9**
Zn-PA-INAM	*25.0*	*27.2*	*29.2*	**27.1**	**137.5**

3.3.4 Thermal Analysis

The DSC analysis is shown in Figure 3-10 and the corresponding onset/peak/endset temperatures, and enthalpy of melting of each compound are summarized in Table A3-1. DSC analysis confirms that the product phase (Zn-PA-INAM) is a pure multicomponent crystal since it has a unique onset of melting (186℃), unlike Zn-PA or the pure components. Absence of a peak related to desorption of moisture or solvents also proves that Zn-PA-INAM does neither contain solvent inclusion nor is a solvate. Note that after the sharp melting peaks, all compounds decompose at a higher temperature with a wide endothermic peak.
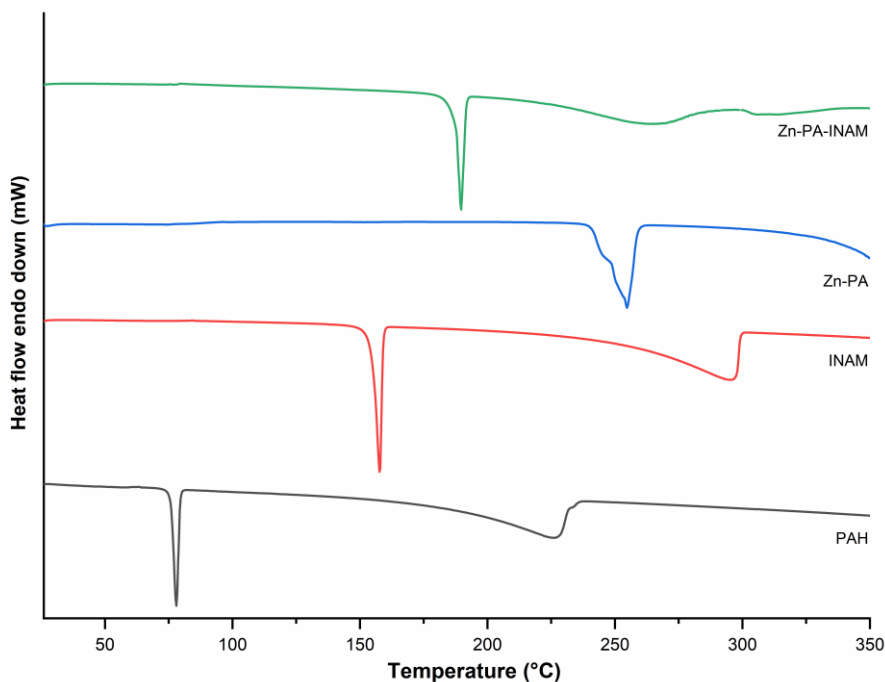
Figure 3-10 Thermographs of Zn-PA-INAM and related compounds, obtained by DSC measurements.

3.3.5 Solubility and Dissolution Rate

HPLC was used to obtain the dissolution rate of Zn-PA-INAM. A list of known concentrations of PAH was first prepared and their HPLC was collected for generating the calibration curve. As shown below (Figure 3-11, left), the retention time was 5.9 minutes. The area under the curve and concentration of PAH are linearly correlated with R^2 of 0.995. This calibration curve was used to measure the Zn-PA and Zn-PA-INAM concertation for dissolution rate analysis (Figure 3-11, middle).
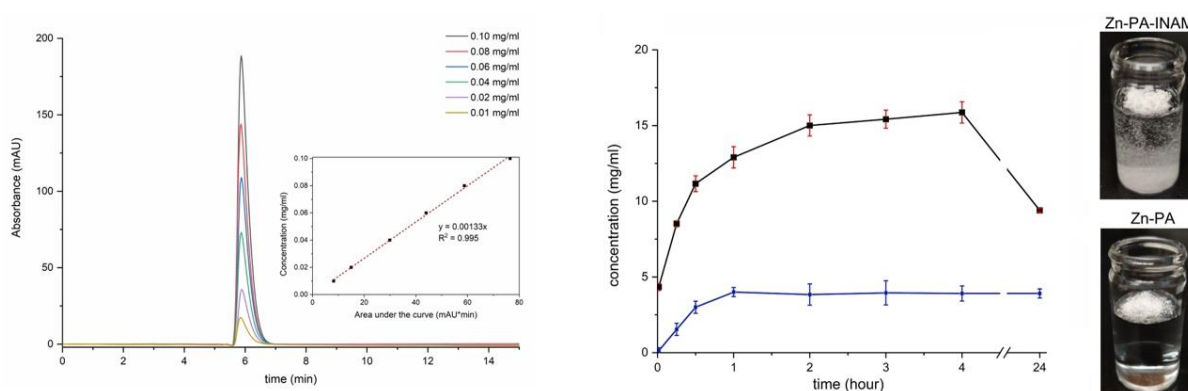
Figure 3-11 HPLC data of a known list of PAH concentrations and the corresponding calibration curve (left). dissolution of Zn-PA-INAM (black) and Zn-PA (blue) over the course of 24 hours (middle). Images of Zn-PA-INAM and Zn-PA in vials showing their interactions with aqueous medium (right).

The hydrophobic Zn-PA remains on the surface while the hydrophilic Zn-PA-INAM immediately sinks and dissolves. Zn-PA-INAM dissolution profile shows that 4.4 mg of Zn-PA-INAM was dissolved immediately while Zn-PA solubility at the beginning is 0. Over the course of 4 hours, the concentration peaked at 15.9 mg/ml of Zn-PA-INAM, and then dropped to 9.4 mg/ml after 24 hours. This drop is due to partial dissociation of Zn-PA-INAM and formation of the PAH. The PXRD of the residue powder after dissolution experiment (Figure A3-5) confirms this observation. Zn-PA solubility reaches equilibrium at 3.9 mg/ml.

3.3.6 Connection between Bonding Motif Modifications and Bulk Properties

CLP-Pixel is used to calculate the energy of interactions in the crystalline state. The program accepts the electron density cube file generated by Gaussian, with ~10^6 datapoints for Zn-PA and Zn-PA-INAM, and condenses them into ~21,000 datapoints called "pixels". Then, the program constructs the crystal models using the symmetry operations of the respective space group and calculates the intermolecular energies by numerical integrals over all pixels. The results are summarized in Table 3-4.

Coulomb-polarization energies are meaningful for polarity; Coulombic energies are important for molecules with hydrogen bonds and have a larger effect (-288.6 kJ/mol) on stabilization of Zn-PA-INAM (N-H···O interactions shown in Figure 3-12.c). Polarization energy is not pairwise-additive and cannot be associated with specific parts of the structure (many-body effect).

Dispersion forces, on the other hand, are most relevant for pi stacking and have a higher contribution in the stabilization of Zn-PA, as illustrated in packing visualization in Figure 3-12.a. Repulsion depends on electron density overlap in the crystal structure, and cell dipole is zero for both crystal structures since their space groups are centrosymmetric.

The expression "like dissolves like" is common knowledge in chemistry (Montes et al., 2003). Water is a polar compound and Zn-PA is mainly non-polar, stabilized by pi-stacking, and lacks any notable polar integrations (e.g. hydrogen bonds). Zn-PA-INAM has significantly higher Coulomb-polarization components in its lattice energy, lowering its hydrophobicity.

Both qualitative and quantitative lattice energy calculations suggest that Zn-PA is hydrophobic and has low water solubility. References (Sanphui et al., 2015; L. Y. Wang et al., 2021; Zhu et al., 2021) also argued the usage of polarity-enhancing formulations to create a crystal with better dissolubility by introducing more polar interactions (coulomb-polarization energies).

Table 3-4 The various components of the lattice energies (kJ/mol) calculated using CLP-Pixel.

Name	Coulombic (E_{col})	Polarization (E_{pol})	Dispersion (E_{disp})	Repulsion (E_{rep})	Cell Dipole (E_{celdip})	Total (E_{tot})
Zn-PA-INAM	-288.6	-119.6	-300.8	283.9	0.0	-425.1
Zn-PA	-113.4	-77.3	-393.5	132.9	0.0	-451.3
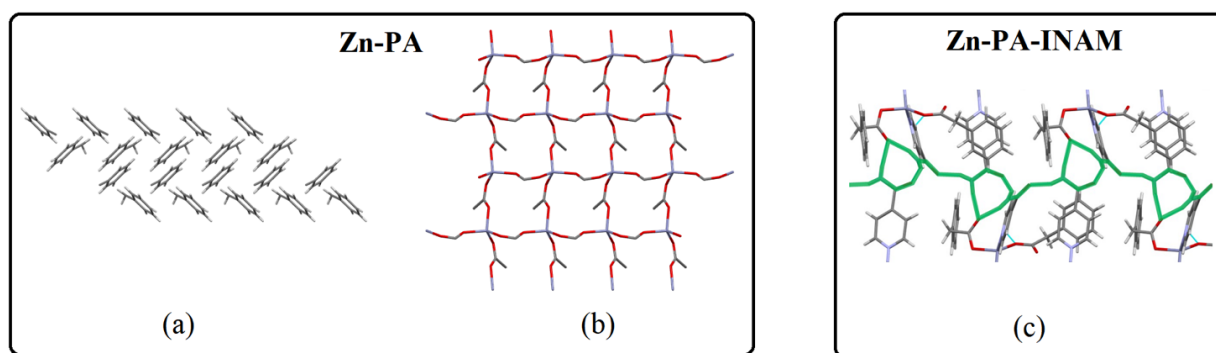
Figure 3-12 Non-polar interactions in Zn-PA crystal structure: (a) strong Pi stacking (dispersion), (b) 2D network of coordination bonds. (c) Polar interactions (green path) in Zn-PA-INAM forming an H-bond network (electrostatic). Note that the view is down the "*a*" crystallographic axis and all other atoms are hidden for visualization purposes.

3.3.7 Crystal Surface Analysis

Surface wettability and powder dissolution are surface properties. Bravais–Friedel–Donnay–Harker (BFDH) model, available as BFDH Morphology module in CCDC Mercury software package (Macrae et al., 2020), predicts the outer shape of a crystal and assigns the miller indices (hkl) of the most relevant crystalline faces, enabling us to analyze the functional groups that are present on each face. As illustrated in Figure 3-13.a, Zn-PA crystal structure results in plate crystals, with {100} family of planes constructing 76.1% of the total facet of the crystals. Zn-PA-INAM (Figure 3-13.b), on the other hand, forms block crystals with 3 morphologically important families of faces: {011}, {002}, and {100} comprising 46.4%, 25.7%, 21.3% of the total facet area, respectively.

As highlighted in Figure 3-13.c, {100} family of faces has no polar function groups available. Figure 3-13.d shows that Zn-PA-INAM has the carbamoyl group on the {100} and {002} faces as well as the carboxylate group on {002}. These polar groups significantly improve the water affinity, reducing the hydrophobicity of the crystal.

Zn-PA				
{hkl}	Multiplicity	d_{hkl}	Distance (mm)	Total facet area %
{100}	2	30.2	3.3	**76.1**
{002}	2	4.8	20.9	7.0
{011}	4	4.4	22.9	6.8
{110}	4	4.8	20.7	6.4
{10$\bar{2}$}	2	4.8	21.0	2.6
{11$\bar{1}$}	4	4.3	23.1	1.0

Zn-PA-INAM				
{hkl}	Multiplicity	d_{hkl}	Distance (mm)	Total facet area %
{011}	4	11.6	8.7	**46.4**
{002}	2	12.0	8.3	**25.7**
{100}	2	8.3	12.1	**21.3**
{10$\bar{2}$}	2	7.1	14.0	3.1
{110}	4	7.0	14.3	2.0
{11$\bar{1}$}	4	6.9	14.5	1.5

(e)
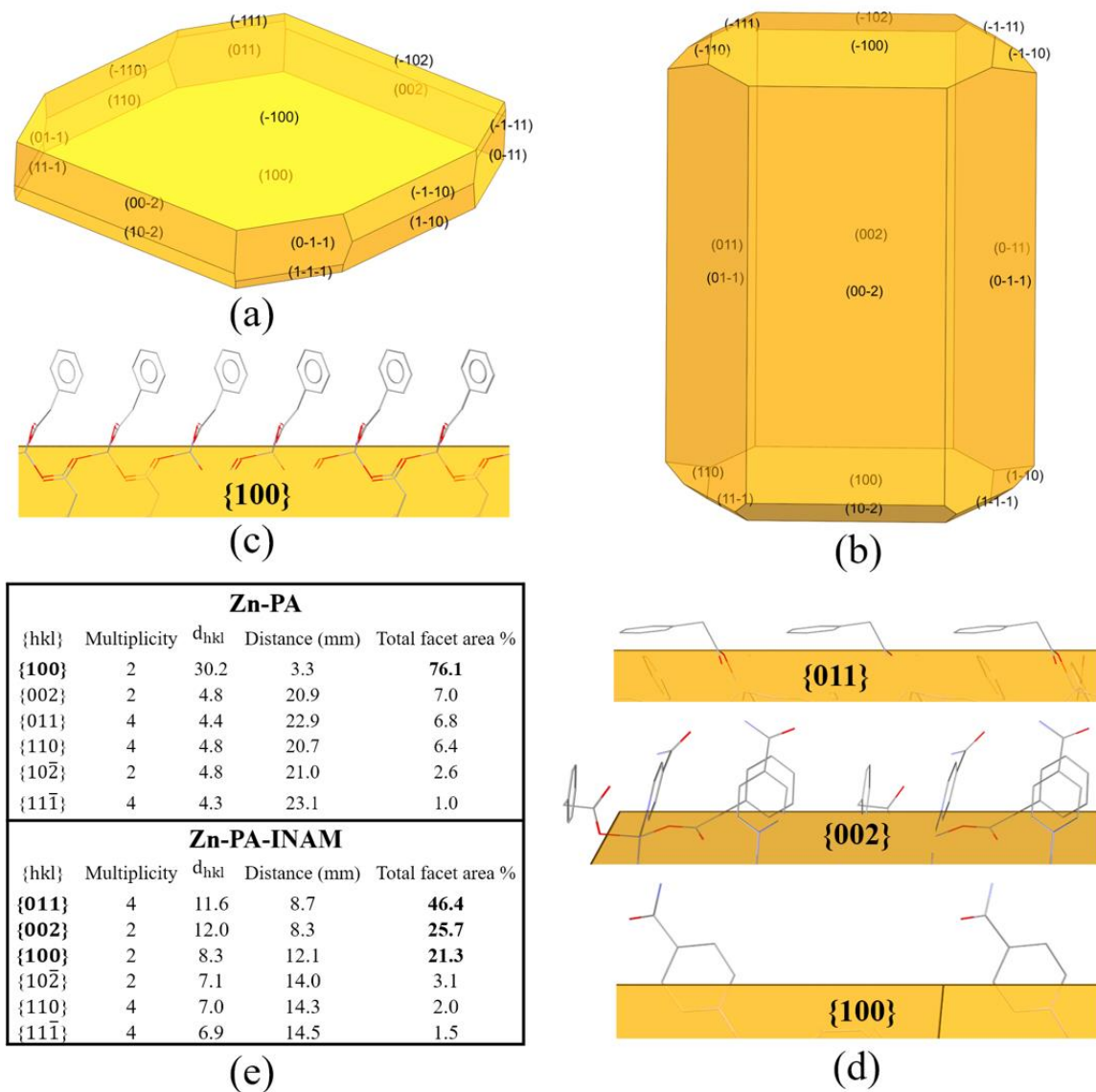
Figure 3-13 BFDH morphology of (a) Zn-PA, and (b) Zn-PA-INAM. Functional groups exposed on the morphologically important faces of (c) Zn-PA, and (d) Zn-PA-INAM. (e) tabular data showing the miller indices, multiplicity, d-spacing, perpendicular distance to the center of the crystal, and the total facet area percentage for morphologically relevant families of faces.

3.4 Conclusion

Co-crystallization of Zn-PA with INAM resulted in a new crystal structure. Upon obtaining the crystal structure of this multicomponent crystal (Zn-PA-INAM), crystal packing, BFDH morphology, *ab initio,* and Hirshfeld surface studies revealed major alteration in solid-state interactions and the emergence of hydrogen bonds that were absent in Zn-PA. Contact angle measurement on hydraulically-pressed tables of Zn-PA and Zn-PA-INAM indicated a marked difference in hydrophobicity between these two compounds, with Zn-PA-INAM being noticeably less hydrophobic. To have a clearer description of this new crystal, vibrational spectroscopy, thermal analysis, and powder dissolution profile are also presented.

3.5 References

Ahmadi, S., Mondal, P. K., Wu, Y., Gong, W., Mirmehrabi, M., & Rohani, S. (2021). Virtual Multicomponent Crystal Screening: Hydrogen Bonding Revisited. *Crystal Growth & Design*, *21*(10), 5862–5872.

Ali, N., Tahir, M. N., Ali, S., Iqbal, M., Munawar, K. S., & Perveen, S. (2014). Synthesis, characterization, crystal structures, enzyme inhibition, DNA binding, and electrochemical studies of zinc(II) complexes. *Journal of Coordination Chemistry*, *67*(7), 1290–1308.

Almarsson, Ö., Peterson, M. L., & Zaworotko, M. (2012). The A to Z of pharmaceutical cocrystals: a decade of fast-moving new science and patents. *Pharmaceutical Patent Analyst*, *1*(3), 313–327.

Almarsson, Ö., & 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*, *17*, 1889–1896.

Batshaw, M. L., Brusilow, S., Waber, L., Blom, W., Brubakk, A. M., Burton, B. K., Cann, H. M., Kerr, D., Mamunes, P., Matalon, R., Myerberg, D., & Schafer, I. A. (1982). Treatment of Inborn Errors of Urea Synthesis — Activation of Alternative Pathways of Waste Nitrogen Synthesis and Excretion. *New England Journal of Medicine*, *306*(23), 1387–1392.

Berge, S. M., Bighley, L. D., & Monkhouse, D. C. (1977). Pharmaceutical Salts. *Journal of Pharmaceutical Sciences*, *66*(1), 1–19.

Berry, D. J., & Steed, J. W. (2017). Pharmaceutical cocrystals, salts and multicomponent systems; intermolecular interactions and property based design. *Advanced Drug Delivery Reviews*, *117*, 3–24.

Bharate, S. S. (2021). Recent developments in pharmaceutical salts: FDA approvals from 2015 to 2019. *Drug Discovery Today*, *26*(2), 384–398.

Bolla, G., & Nangia, A. (2016). Pharmaceutical cocrystals: walking the talk. *Chemical Communications*, *52*(54), 8342–8360.

Bolla, G., Sarma, B., & Nangia, A. K. (2022). Crystal Engineering of Pharmaceutical Cocrystals in the Discovery and Development of Improved Drugs. *Chemical Reviews*, *122*(13), 11514–11603.

Brouillard, A. M., Kraja, A. T., & Rich, M. W. (2019). Trends in Dietary Sodium Intake in the United States and the Impact of USDA Guidelines: NHANES 1999-2016. *The American Journal of Medicine*, *132*(10), 1199–1206.

Bruker. (2012). *APEX2, SAINT*. Bruker AXS Inc., Madison, Wisconsin, USA.

Buckton, G., & Newton, J. M. (1985). Assessment of the Wettability Of Pharmaceutical Powders Using Compressed Discs. *Journal of Pharmacy and Pharmacology*, *37*(Supplement_12), 27P.

Burla, M. C., Caliandro, R., Carrozzini, B., Cascarano, G. L., Cuocci, C., Giacovazzo, C., Mallamo, M., Mazzone, A., & Polidori, G. (2015). Crystal structure determination and refinement via SIR2014. *Journal of Applied Crystallography*, *48*(1), 306–309.

Chasapis, C. T., Loutsidou, A. C., Spiliopoulou, C. A., & Stefanidou, M. E. (2012). Zinc and human health: an update. *Archives of Toxicology*, *86*(4), 521–534.

Cheney, M. L., Weyna, D. R., Shan, N., Hanna, M., Wojtas, L., & Zaworotko, M. J. (2011). Coformer Selection in Pharmaceutical Cocrystal Development: a Case Study of a Meloxicam Aspirin Cocrystal That Exhibits Enhanced Solubility and Pharmacokinetics. *Journal of Pharmaceutical Sciences*, *100*(6), 2172–2181.

Cherukuvada, S., Babu, N. J., & Nangia, A. (2011). Nitrofurantoin–p-aminobenzoic acid cocrystal: Hydration stability and dissolution rate studies. *Journal of Pharmaceutical Sciences*, *100*(8), 3233–3244.

De Las Heras, J., Aldámiz Echevarría, L., Martínez Chantar, M. L., & Delgado, T. C. (2017). An update on the use of benzoate, phenylacetate and phenylbutyrate ammonia scavengers for interrogating and modifying liver nitrogen metabolism and its implications in urea cycle disorders and liver disease. *Expert Opinion on Drug Metabolism & Toxicology*, *13*(4), 439–448.

Delori, A., Galek, P. T. A., Pidcock, E., Patni, M., & Jones, W. (2013). Knowledge-based hydrogen bond prediction and the synthesis of salts and cocrystals of the anti-malarial drug pyrimethamine with various drug and GRAS molecules. *CrystEngComm*, *15*(15), 2916–2928.

Desiraju, G. R. (1995). Supramolecular Synthons in Crystal Engineering—A New Organic Synthesis. *Angewandte Chemie International Edition in English*, *34*(21), 2311–2327.

Desiraju, G. R. (2007). Crystal Engineering: A Holistic View. *Angewandte Chemie International Edition*, *46*(1), 8342–8356.

Desiraju, G. R. (2013). Crystal Engineering: From Molecule to Crystal. *Journal of the American Chemical Society*, *135*(27), 9952–9967.

Devogelaer, J.-J., Meekes, H., Tinnemans, P., Vlieg, E., & de Gelder, R. (2020). Co-crystal Prediction by Artificial Neural Networks. *Angewandte Chemie International Edition*, *59*(48), 21711–21718.

Dolomanov, O. V, Bourhis, L. J., Gildea, R. J., Howard, J. A. K., & Puschmann, H. (2009). OLEX2: a complete structure solution, refinement and analysis program. *Journal of Applied Crystallography*, *42*(2), 339–341.

Frisch, M. J., Head-Gordon, M., & Pople, J. A. (1990). A direct MP2 gradient method. *Chemical Physics Letters*, *166*(3), 275–280.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. a., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. a., Nakatsuji, H., Li, X., Caricato, M., Marenich, a. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., … Fox, D. J. (2016). *Gaussian 16, Revision b.01* (Gaussian 16, Revision b01). Gaussian, Inc., Wallin CT.

Gavezzotti, A. (2005). *Calculation of lattice energies of organic crystals: the PIXEL integration method in comparison with more traditional methods*. *220*(5–6), 499–510.

Head-Gordon, M., Pople, J. A., & Frisch, M. J. (1988). MP2 energy evaluation by direct methods. *Chemical Physics Letters*, *153*(6), 503–506.

JADE 7.0. (2020). Materials Data: Livermore, CA.

Karde, V., & Ghoroi, C. (2014). Influence of surface modification on wettability and surface energy characteristics of pharmaceutical excipient powders. *International Journal of Pharmaceutics*, *475*(1), 351–363.

Karki, S., Friščić, T., Fábián, L., Laity, P. R., Day, G. M., & Jones, W. (2009). Improving Mechanical Properties of Crystalline Solids by Cocrystal Formation: New Compressible Forms of Paracetamol. *Advanced Materials*, *21*(38–39), 3905–3909.

Kavuru, P., Aboarayes, D., Arora, K. K., Clarke, H. D., Kennedy, A., Marshall, L., Ong, T. T., Perman, J., Pujari, T., Wojtas, Ł., & Zaworotko, M. J. (2010). Hierarchy of Supramolecular Synthons: Persistent Hydrogen Bonds Between Carboxylates and Weakly Acidic Hydroxyl Moieties in Cocrystals of Zwitterions. *Crystal Growth & Design*, *10*(8), 3568–3584.

Law, K. Y. (2014). Definitions for Hydrophilicity, Hydrophobicity, and Superhydrophobicity: Getting the Basics Right. *The Journal of Physical Chemistry Letters*, *5*(4), 686–688.

Li, D., Kong, M., Li, J., Deng, Z., & Zhang, H. (2018). Amine–carboxylate supramolecular synthon in pharmaceutical cocrystals. *CrystEngComm*, *20*(35), 5112–5118.

Liu, K., Zhang, Y., Deng, L., Jiao, S., Xiao, Z., Cao, F., & Wang, L. (2018). Synthesis, Structure, and Properties of Coordination Polymers Based on 1,4-Bis((2-methyl-1H-imidazol-1-yl)methyl)benzene and Different Carboxylate Ligands. In *Crystals* (Vol. 8, Issue 7).

Macrae, C. F., Sovago, I., Cottrell, S. J., Galek, P. T. A., McCabe, P., Pidcock, E., Platings, M., Shields, G. P., Stevens, J. S., Towler, M., & Wood, P. A. (2020). Mercury 4.0: from visualization to analysis, design and prediction. *Journal of Applied Crystallography*, *53*(1), 226–235.

Mannava, M. K. C., Gunnam, A., Lodagekar, A., Shastri, N. R., Nangia, A. K., & Solomon, K. A. (2021). Enhanced solubility, permeability, and tabletability of nicorandil by salt and cocrystal formation. *CrystEngComm*, *23*(1), 227–237.

Mittapalli, S., Bolla, G., Perumalla, S., & Nangia, A. (2016). Can we exchange water in a hydrate structure: a case study of etoricoxib. *CrystEngComm*, *18*(16), 2825–2829.

Montes, I., Lai, C., & Sanabria, D. (2003). Like Dissolves Like: A Guided Inquiry Experiment for Organic Chemistry. *Journal of Chemical Education*, *80*(4), 447.

Morrison, A. C., & Ness, R. B. (2011). Sodium Intake and Cardiovascular Disease. *Annual Review of Public Health*, *32*(1), 71–90.

Musumeci, D., Hunter, C. A., Prohens, R., Scuderi, S., & McCabe, J. F. (2011). Virtual cocrystal screening. *Chemical Science*, *2*(5), 883–890.

Nanubolu, J. B., Sridhar, B., Ravikumar, K., Sawant, K. D., Naik, T. A., Patkar, L. N., Cherukuvada, S., & Sreedhar, B. (2013). Polymorphism in metformin embonate salt – recurrence of dimeric and tetrameric guanidinium–carboxylate synthons. *CrystEngComm*, *15*(22), 4448–4464.

Ortega, G., Hernández, J., González, T., Dorta, R., & Briceño, A. (2018). 1,2,4,5-Benzenetetracarboxylic acid: a versatile hydrogen bonding template for controlling the regioselective topochemical synthesis of head-to-tail photodimers from stilbazole derivatives. *Photochemical & Photobiological Sciences*, *17*(5), 670–680.

Palacios, E. G., Juárez-López, G., & Monhemius, A. J. (2004). Infrared spectroscopy of metal carboxylates: II. Analysis of Fe(III), Ni and Zn carboxylate solutions. *Hydrometallurgy*, *72*(1), 139–148.

Praphanproj, V., Boyadjiev, S. A., Waber, L. J., Brusilow, S. W., & Geraghty, M. T. (2000). Three cases of intravenous sodium benzoate and sodium phenylacetate toxicity occurring in the treatment of acute hyperammonaemia. *Journal of Inherited Metabolic Disease*, *23*(2), 129–136.

Saikia, B., Bora, P., Khatioda, R., & Sarma, B. (2015). Hydrogen Bond Synthons in the Interplay of Solubility and Membrane Permeability/Diffusion in Variable Stoichiometry Drug Cocrystals. *Crystal Growth & Design*, *15*(11), 5593–5603.

Sanphui, P., Bolla, G., Nangia, A., & Chernyshev, V. (2014). Acemetacin cocrystals and salts: structure solution from powder X-ray data and form selection of the piperazine salt. *IUCrJ*, *1*(2), 136–150.

Sanphui, P., Devi, V. K., Clara, D., Malviya, N., Ganguly, S., & Desiraju, G. R. (2015). Cocrystals of Hydrochlorothiazide: Solubility and Diffusion/Permeability Enhancements through Drug–Coformer Interactions. *Molecular Pharmaceutics*, *12*(5), 1615–1622.

Sanphui, P., Kumar, S. S., & Nangia, A. (2012). Pharmaceutical Cocrystals of Niclosamide. *Crystal Growth & Design*, *12*(9), 4588–4599.

Sheldrick, G. M. (2015). Crystal structure refinement with SHELXL. *Acta Crystallographica Section C*, *71*(1), 3–8.

Shunnar, A. F., Dhokale, B., Karothu, D. P., Bowskill, D. H., Sugden, I. J., Hernandez, H. H., Naumov, P., & Mohamed, S. (2020). Efficient Screening for Ternary Molecular Ionic Cocrystals Using a Complementary Mechanosynthesis and Computational Structure Prediction Approach. *Chemistry – A European Journal*, *26*(21), 4752–4765.

Smith, A. J., Kavuru, P., Wojtas, L., Zaworotko, M. J., & Shytle, R. D. (2011). Cocrystals of Quercetin with Improved Solubility and Oral Bioavailability. *Molecular Pharmaceutics*, *8*(5), 1867–1876.

Turner, M. J., McKinnon, J. J., Wolff, S. K., Grimwood, D. J., Spackman, P. R., Jayatilaka, D., & Spackman, M. A. (2017). *CrystalExplorer17*. University of Western Australia.

Walker, V. (2009). Ammonia toxicity and its prevention in inherited defects of the urea cycle. *Diabetes, Obesity and Metabolism*, *11*(9), 823–835.

Wang, J., Dai, X. L., Lu, T. B., & Chen, J. M. (2021). Temozolomide–Hesperetin Drug–Drug Cocrystal with Optimized Performance in Stability, Dissolution, and Tabletability. *Crystal Growth & Design*, *21*(2), 838–846.

Wang, L. Y., Zhao, M. Y., Bu, F. Z., Niu, Y. Y., Yu, Y. M., Li, Y. T., Yan, C. W., & Wu, Z. Y. (2021). Cocrystallization of Amantadine Hydrochloride with Resveratrol: The First Drug–Nutraceutical Cocrystal Displaying Synergistic Antiviral Activity. *Crystal Growth & Design*, *21*(5), 2763–2776.

Wang, T., Stevens, J. S., Vetter, T., Whitehead, G. F. S., Vitorica-Yrezabal, I. J., Hao, H., & Cruz-Cabeza, A. J. (2018). Salts, Cocrystals, and Ionic Cocrystals of a "Simple" Tautomeric Compound. *Crystal Growth & Design*, *18*(11), 6973–6983.

Wong, F. (2012). Management of ascites in cirrhosis. *Journal of Gastroenterology and Hepatology*, *27*(1), 11–20.

Zeleňák, V., Vargová, Z., & Györyová, K. (2007). Correlation of infrared spectra of zinc(II) carboxylates with their structures. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *66*(2), 262–272.

Zhu, Q. H., Zhang, G. H., He, L., Qin, S., Tian, J. H., Ma, L., & Tao, G. H. (2021). Enhanced Solubility and Antitumor Activity of Curcumin via Breaking and Rebuilding of the Hydrogen Bond. *ACS Applied Bio Materials*, *4*(1), 918–927.

Chapter 4

Virtual Multicomponent Crystal Screening: Hydrogen Bonding Revisited

A version of this chapter was published as:

4 Virtual Multicomponent Crystal Screening: Hydrogen Bonding Revisited

Abstract

Pharmaceutical cocrystals, salts, and multicomponent crystals, in general, have increasingly come under the spotlight in recent years. A fast and efficient *a priori* theoretical classifier to identify potential coformers is highly sought-after to complement the experimental brute force screening methods. This research examines the qualitative approaches that are based on hydrogen bonding strength. First, molecular electrostatic potential (ESP) maps of 330 coformers were obtained from density functional theory (DFT) simulations, using two geometries: experimentally determined crystal structures, and gas-phase optimization. An in-depth comparison of ESPs revealed the potential pitfalls of these two geometries that are deliberated at length in the manuscript. Next, six active pharmaceutical ingredients (API) and their reported salts/cocrystals on Cambridge Structural Database (CSD) were inversely predicted with ESP analysis. For 2 of these APIs, the prediction showed systematic errors that are resolved with suggestions provided in the manuscript. Subsequently, hydrogen bond energy (HBE) and hydrogen bond propensity (HBP) calculations were put to the test with 2 APIs and 52 organic coformers. Finally, multivariate logistic regression, a linear machine learning (ML) algorithm, showed how a combination of HBE and HBP can be a superior classifier, for which 18 out of 25 positive cases were uninterruptedly identified at the top of the list. Provided that a database of failed attempts of cocrystallization is compiled within the scientific community to supplement the existing positive results (multicomponent crystals on CSD), the combination of chemistry-based parameters and machine learning can be a promising classifier for coformer selection.

4.1 Introduction

The physicochemical properties of the majority of existing or newly discovered active pharmaceutical ingredients (API) limit the drug development process, primarily due to poor solubility and bioavailability (Tiekink & Zukerman-Schpector, 2017). A crystal engineer aims

to design a solid form that is stable under different temperature and humidity conditions, meets the required solubility, dissolution rate and bioavailability criteria, and also shows proper compressibility and flowability (Tiekink & Zukerman-Schpector, 2017).

Multicomponent crystallization is arguably one of the most common and robust design approaches to unfold the possible crystal structure landscape of a target API. Salts, cocrystals, hydrates, solvates, and coamorphous systems all fit under the umbrella term of multicomponent solid forms.

Experimental screening for multicomponent solid forms is an expensive and time-consuming process that is done by a wide range of techniques, including high-throughput solution growth (Morissette et al., 2004; Rodrigues et al., 2020), neat and liquid-assisted grinding (Friščić, 2018; Karki et al., 2007), and slurrying (Zhang et al., 2007). To accelerate the screening process, these brute force screening methods can be accompanied by theoretical tools to predict the multicomponent crystal formation beforehand.

Hydrogen bonding (H-bonding), due to its strong and directional nature, has the reputation of being "the most reliable interaction in the toolkit of a crystal engineer" (Desiraju et al., 2011). Despite countless research being conducted on the role of various molecular features (electrostatic, steric, and geometric effects) on H-bonding, the task to predict which set of intermolecular interactions will be observed in the solid-state remains challenging (Sandhu et al., 2018).

Fábián (2009) utilized the existing structures from Cambridge Structural Database (CSD) to find a set of molecular descriptors that are similar or complementary in molecular pairs of a cocrystal. He concluded that polarity and shape factors and relative molecular dimensions (principle of close packing) are the strongest descriptors.

Price and coworkers (Issa et al., 2009; Karamertzanis et al., 2009) compared the lattice energy of some selected cocrystals with the sum of the lattice energies of their components to check whether the prediction is viable, assuming that a cocrystal has a lower lattice energy and is thermodynamically more stable than its components.

Hunter (Hunter, 2004) argued that repulsion, induction, and dispersion contributions to the enthalpy of intermolecular interactions (i.e. H-bonding) are negligible compared to electrostatic interactions. Consequently, the strength of an H-bonding site of a molecule can be quantified and extracted from the molecular electrostatic potential (ESP) map. More recently, Hunter et al. (Grecu et al., 2014; Musumeci et al., 2011) investigated the applicability of this approach for virtual cocrystal screening.

Hydrogen bond propensity (HBP) (Galek et al., 2009) is a knowledge-based H-bond predictive model that estimates the likelihood of an H-bond formation based on randomly selected related known crystal structures in CSD. The algorithm assigns a propensity value to a pair of H-bond donor-acceptor, considering their respective molecular environments. Although HBP was originally designed for polymorphic assessments, it has attracted considerable attention for multicomponent systems in the recent years (Delori et al., 2013; Majumder et al., 2013; Sandhu et al., 2018; Sarkar et al., 2019, 2020; Sarkar & Aakeröy, 2020).

Although validation and benchmarking of these approaches have been conducted on a few APIs, the field of multicomponent crystal prediction needs to be explored and examined on a larger scale. There are also some uncertainties associated with the validation data set. Besides from definite and accessible crystal structure of multicomponent crystals on CSD, much care must be taken into categorizing the outcome of an experiment into "YES" or "NO" columns.

This research is an attempt to analyze the H-bond energy calculations by molecular electrostatic potential maps (ESP), as well as H-bond propensity assessment for multicomponent crystal screening. We started with producing the ESP of 330 common coformers and studied how the surface maxima and minima (i.e. H-bonding sites) change during the geometry optimization process. Next, we used 6 APIs with a number of reported salts/cocrystals on CSD to check whether these proven positive results are predictable by ESP analysis. The third task we took on was to put H-bond energy and propensity analysis to the test, with an experimental screening of 2 APIs, Rufinamide and Olanzapine, with over 50 organic coformers. Even though each method performed reasonably well, their combination with logistic regression showed significant enhancement in differentiating positive and negative results.

4.2 Experimental Section

4.2.1 Materials

Apotex PharmaChem Inc. (Brantford, ON) donated Rufinamide and Olanzapine that are used as target APIs for screening. Coformers (>97%) were purchased from Alfa Aesar, Sigma-Aldrich, and ACROS Organics and were used as received. The selected list of coformers is shown in Table A4-1.

4.2.2 Experimental Cocrystal Screening

Liquid assisted grinding (LAG) was used to prepare the mixture of the API-coformer, using mortar and pestle grinding and ball milling. For each experiment, a 200 mg mixture of 1:1 molar ratio of API-coformer was prepared. One bulb of MeOH (2 mL) was added at 10 min intervals. The procedure of manual grinding continued for 30 min. In the case of ball milling LAG experiments, the 200 mg mixture and one bulb of MeOH were transported to an air-tight ball mill jar (PTFE SmartSnap Jar – 5 mL) with two 5 mm Zirconia balls. The mixture was mixed for 30 min in a Retsch Mixer Mill MM 200 at 12 Hz frequency and then was dried under vacuum.

Powder X-ray diffraction (PXRD) diffractograms were the primary indicator that whether the result of mixing was re-crystallization (a physical mixture) or co-crystallization (a multi-component crystal). Rigaku-MiniFlex II benchtop diffractometer with a Cu Kα source (1.54059 Å) at 30 kV and 15 mA was used to collect the data, which then was analyzed in JADE 7.0 software (*JADE 7.0*, 2020). In case only the characteristic peaks of the API and coformer were present in the diffractogram of the mixture, it was considered as a "NO". Otherwise, differential scanning calorimetry (DSC) was employed to further identify the outcome of mixing.

Five mg of the sample was put in a 40 μL aluminum crucible and covered with a pinhole lid and characterized with a Mettler Toledo DSC 822e differential scanning calorimeter. The data was collected for the temperature range of 25 to 300°C with a 5°C/min heating rate. When a new phase transformation peak other than the melting point of the pure compounds (or potentially their methanolate solvates) was observed, the result was noted as a "YES" for the

formation of a multi-component crystal. We also tried to confirm the formation of multicomponent crystals by growing single crystals with slow evaporation in different solvents. A detailed characterization of the crystal structures of the successful crystallizations is reported separately (Gong et al., 2021).

4.2.3 Hydrogen Bond Energy (HBE)

Molecular electrostatic potential maps were the primary tool for the HBE analysis. ESP of each molecule was mapped to 0.002 e⁻/Å³ electron density isosurface, computed by Gaussian 16 (Frisch et al., 2016) using B3LYP density functional theory (DFT) method and 6-311++G** basis set. The local maxima (ESP_{max}) and minima (ESP_{min}), in units of kJ/mol, were then used to calculate the H-bond donor parameter (α) and H-bond acceptor parameter (β), using equations 1 and 2, respectively (Musumeci et al., 2011).

$$\alpha = 0.0000162 \, ESP_{max}^2 + 0.00962 \, ESP_{max} \quad (1)$$

$$\beta = 0.000146 \, ESP_{min}^2 - 0.00930 \, ESP_{min} \quad (2)$$

We wrote a Python script in order to find the surface local extrema accurately and rapidly. The total pairing energy in the solid-state was estimated as the sum over a hierarchical listing of a specific number of complementary H-bond donor-acceptor sites ($E = -\sum_{ij} \alpha_i \beta_j$).

Finally, the difference between pairing energies of the API-coformer crystal (E_{C-A} and E_{A-C}) and the individual molecules of API (E_{A-A}) and coformer (E_{C-C}) provided a measure to rank the coformers from the most to the least energetically favourable. ESP of 330 common coformers is listed with a thorough review of relevant literature. The name of these compounds can be found in Table A4-2 in the supporting information.

4.2.4 Hydrogen Bond Propensity (HBP)

Hydrogen bond propensities (HBPs) were calculated using Mercury (Macrae et al., 2020) software. The workflow starts with merging the coordinate files of a target API with a coformer from Table A4-1. This merged file has to be auto-edited in Mercury to specify unidentified bond types (e.g. single, double, triple, aromatic, and delocalized bonds). In the propensity

prediction wizard, the functional groups, and H-bond donors/acceptors are identified. The next step is generating fitting data, in which more than 300 crystal structure for each functional group is randomly selected. The logit model, combined with all seven potential model variables (competition, donor and acceptor atoms of the function group, their steric density, and aromaticity), is used. A full report on the propensities of all H-bond pairs of the donor-acceptor present in the system is then generated with the average receiver operating characteristic (ROC) of 0.82.

The propensity of the highest heteromeric interaction (API-coformer (P_{A-C}) was compared with the highest homomeric interaction (either API-API (P_{A-A}) or coformer-coformer (P_{C-C})). The difference ($\Delta_{HBP} = P_{A-C} - Max(P_{A-A}, P_{C-C})$) was used to predict the outcome of the experiment, with higher values suggesting more favourable cocrystallization.

4.2.5 Estimation of Acid Dissociation Constant (pKa)

Aqueous pKa values of acids and the protonated bases were obtained from the pKa plugin implemented in MarvinSketch 21.3 (*MarvinSketch 21.3*, 2021). Macro mode and static acid/base prefix were used. The temperature was set to 298 K, and the minimum basic pKa and the maximum acidic pKa were set to -2 and 30, respectively. ΔpKa, defined as pKa[protonated base] − pKa[acid], is commonly used for differentiating salts from cocrystals. Cruz-Cabeza (2012) showed, provided that ΔpKa < -1 for cocrystals or ΔpKa > 4 for salts, we can rely on the ΔpKa rule with more than 99% certainty. The probability of salt of cocrystal formation for the in-between region, -1 < ΔpKa < 4, can also be estimated by equations given in the reference.

4.3 Results And Discussion

4.3.1 Comparison between Gas-phase Optimized and Solid-phase Structures

There are two ways to obtain ESP maps: we can either use solid-state coordinates of a molecule from experimentally obtained crystal structure from CSD or use theoretical gas-phase optimization implemented in Gaussian16, in which the molecular geometry is adjusted to find a stationary point on the potential surface. The optimization algorithm is significantly more computationally expensive: for the 330 coformer compounds selected for this study,

optimization, on average, required 17.5 times more CPU time than direct energy calculations. Therefore, it would be ideal to bypass the optimization by using experimental solid-state geometries.

As the first step, we investigated the difference between H-bond donor ($\Delta\alpha$) and acceptor ($\Delta\beta$) values obtained from crystal structures versus gas-phase optimized geometries. The objective was two-fold: first, to check whether the optimization is necessary; and second, to find potential discrepancies where gas-phase optimization is not favored. The summary of this analysis is shown in Figure 4-1.
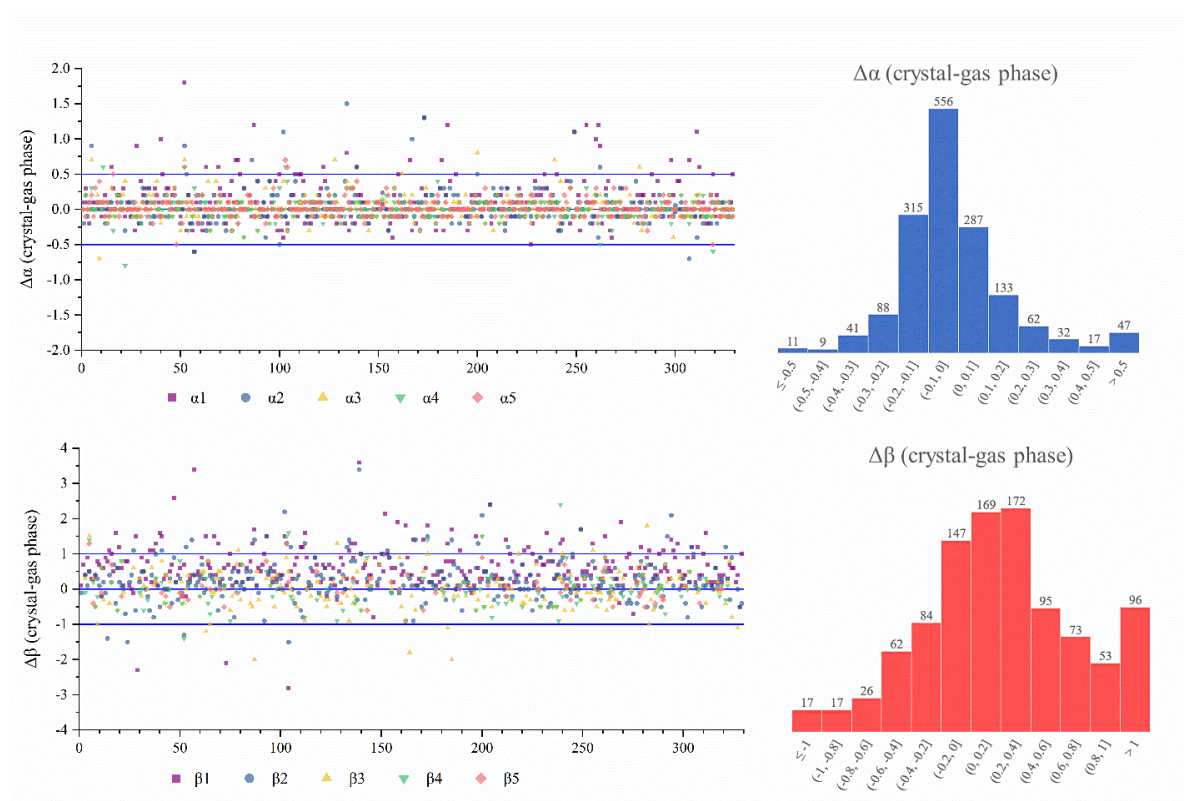
Figure 4-1 Deviation in H-bond donor ($\Delta\alpha$) and acceptor ($\Delta\beta$) values of geometries obtained from crystal structures vs. gas-phase optimization.

Overall, 1598 donors (α) and 1011 acceptors (β) were identified. The vast majority of αs show acceptable agreement between these two methods, with 96.7% being within the ± 0.5 range. Nevertheless, the variations in β values were more pronounced, with 88.8% being within the ± 1.0 range. Upon closer examination of the outliers, some patterns were observed:

First, the bond length of acidic hydrogen of sulfonic acids has significant deviations in crystal structures (ranging from 0.820 to 1.025 Å). In gas-phase optimized geometries, on the other hand, this bond length is consistently and reproducibly at 0.969 Å $\pm$0.001. p-toluenesulfonic acid (Figure 4-2.a), camphorsulfonic acid, 1,5-naphthalenedisulfonic acid, and 3-carboxy-4-hydroxybenzenesulfonic acid are some examples with $\Delta\alpha$ greater than 1.

The most extreme variations in β were related to the carbonyl group (C=O) with 20 structures in the outlier category (Figure 4-2.b). The double bond was relaxed at 1.22$\pm$0.01 in the optimized geometries but had small length or angle variations in crystal structures. Four molecules (malonamide, malonic acid, succinamide, and phthalic acid) had two carbonyl groups in unique orientation: both carbonyl oxygens in crystals shared their electron clouds to form a strong acceptor group on one side of the molecule, whereas, in the gas-phase, these groups are on the opposite side of the molecule (Figure 4-2.c). These two orientations have more than 2 units of difference in their β values. The less extreme, but more common discrepancy (38 cases), was observed for the acidic proton of carboxylic acids (Figure 4-2.d). The elongation of the O-H bond was observed in some crystal structures, resulting in an overestimation of both α and β values.

The third common difference was the shape of NH_2 functional groups. It has an angular geometry in gas-phase but has a planar characteristic in the vicinity of strong acceptors in crystal structures (Figure 4-2.e). Both donor strength of hydrogens and acceptor nature of nitrogen of NH_2 group are underestimated in crystals.

Lastly, intramolecular interactions also alter the α, β values. Glycolic acid (Figure 4-2.f), for instance, forms an intramolecular interaction in the gas-phase, which is replaced by stronger intermolecular interactions in solid-state. Other examples are citric acid, dihydromyricetin, malic acid, salicylamide, salicylic acid, and vanillin. For cases that these intramolecular interactions are consistently absent in the crystal structures, gas-phase optimization entails an underestimation of the strength of H-bond moieties.

Overall, gas-phase optimization, although significantly more time-consuming, is preferred for systematic ESP analysis. Small variations in atomic orientations in different solid-state

environments and uncertainties associated with locating the exact position of H atoms during crystal structure determination, are the main reasons why optimization is superior. Optimization, however, is not flawless; the main pitfall in optimization is with regard to the formation of weak intramolecular interactions that are commonly broken and substituted with stronger intermolecular bonds in solid-state, leading to underestimation of αs and βs. In such cases, specific atomic positions need to be frozen during the optimization process, or crystal structures are taken without optimization.
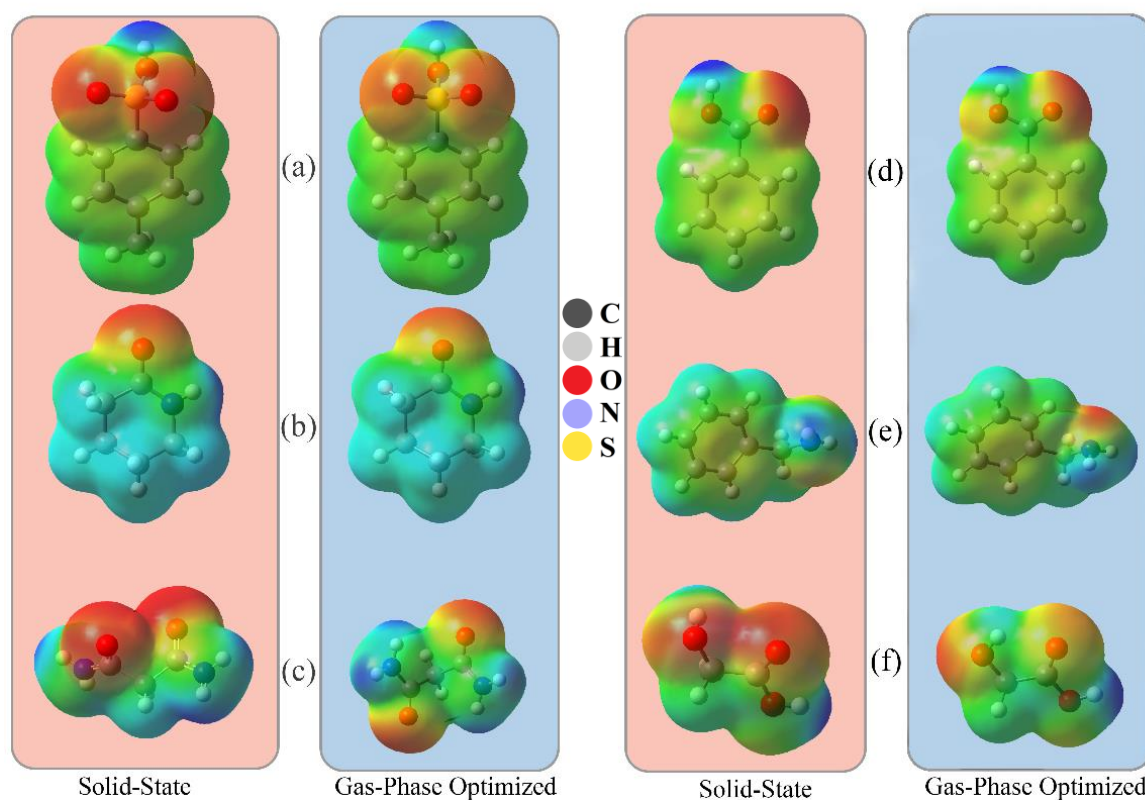
Figure 4-2 Examples of mismatches between solid-state and gas-phase optimized geometries. CSD identifiers/Refcodes of crystal structures are provided in parenthesis. (a) p-toluenesulfonic acid (WUQDUJ), (b) 2-Piperidone (HIQJOJ), (c) malonamide (DADMAZ), (d) benzoic acid (BENZAC12), (e) benzylamine (EVUGIL), (f) glycolic acid (GLICAC01).

4.3.2 H-bonding Energy Analysis of 6 well-studied APIs

To put the HBE analysis to the test, we started with the reverse prediction of some known multicomponent crystals. The most reliable experimental proof of multicomponent crystals is whether their 3D structures are reported on CSD. Six APIs with notable number of salts/cocrystals were selected for this study. Acetaminophen, acetazolamide, carbamazepine, lamotrigine, theophylline, and trimethoprim comprise the list of selected APIs. The multicomponent crystals analyzed in this study are listed in Table A4-3. Although without including the failed cocrystallization attempts the predictive robustness of a classifier cannot be determined, it is an unquestionably necessary element of any predictive tool to explain why reported cocrystals exist.

The primary interaction in the hierarchical mapping of the complementary H-bond donor/acceptor sites makes up the biggest portion of the total net energy. Hence, we started the prediction by considering only the largest α, β of the API and cocrystal, using the below equation:

$$E_{net} = -\left(\alpha_{1,C}\beta_{1,A} + \alpha_{1,A}\beta_{1,C}\right) + \left(\alpha_{1,C}\beta_{1,C} + \alpha_{1,A}\beta_{1,A}\right)$$
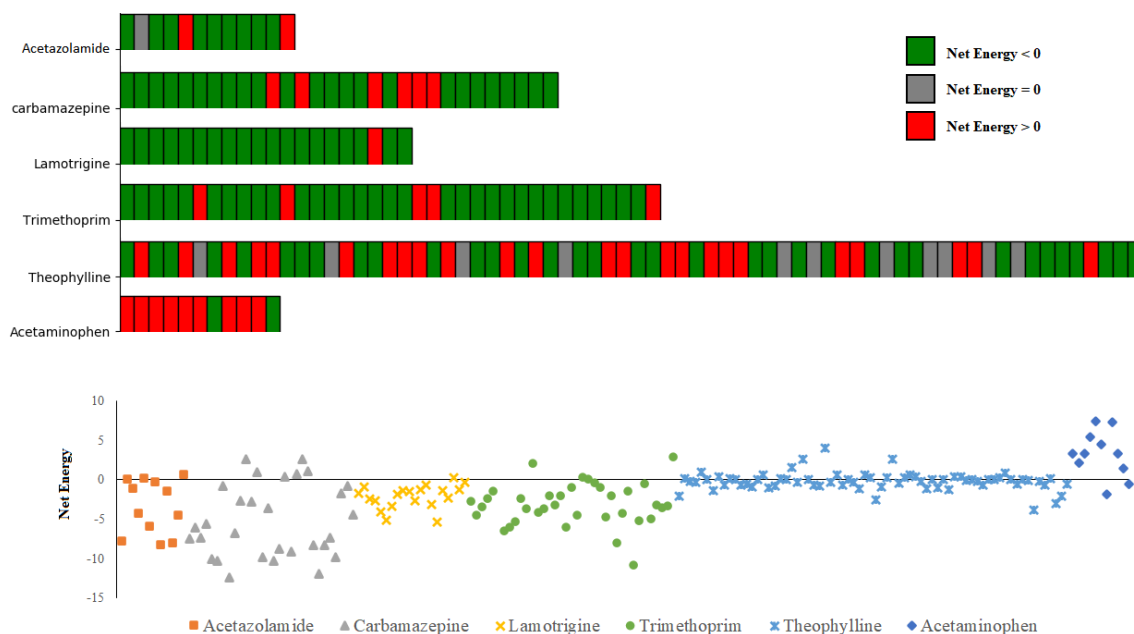
Figure 4-3 Comparing the net energy of the cocrystals/salts of 6 APIs based on the CSD database.

As illustrated in Figure 4-3, the success rate of this prediction was mixed, depending on the API. For the first 4 APIs, 87% of the cocrystals were successfully predicted. For theophylline, 66% of the cocrystals showed an improvement in the net energy, but the points are concentrated around 0, making the prediction outcome statistically insignificant. The results for acetaminophen were surprisingly unexpected, with only 2 out of 11 (18%) cocrystals showing an improvement in the net energy.

At this point, we decided to unravel the cause of the failed predictions of theophylline and acetaminophen case studies. The key logical premise of E_{net} calculation is that the α, β values are paired in descending order. This is true for small molecules, including most of the coformers. However, the competition for pairing H-bond donor/acceptor can be more complex for APIs with multiple functional groups. Such systematic error can lead to overestimation of the API stability (API-API interaction).

Theophylline exists in four polymorphic forms. Theophylline form II is kinetically stable and was considered as the thermodynamically stable form before form IV was discovered (Bobrovs

et al., 2015). As rendered in Figure 4-4.a, the main interactions in theophylline form II are $\alpha_1\beta_3$ and $\alpha_2\beta_2$ with numerical values of 13.5 and 11.7, respectively. Therefore, assuming that $\alpha_1\beta_1$ (24.2) represents the primary interaction of this API, offsets the net energy by 10.7. If we account for this systematic error, 100% of cocrystals can be predicted.

Note that in form IV, the primary interaction is $\alpha_1\beta_2$ (21.7) which is significantly closer to the $\alpha_1\beta_1$ (24.2). However, form IV can only be obtained in specific solvents with a solvent-mediated transformation mechanism (Bobrovs et al., 2015). Therefore, in normal cocrystallization experiments, the competition is between the kinetically stable polymorph (form II) and cocrystals.
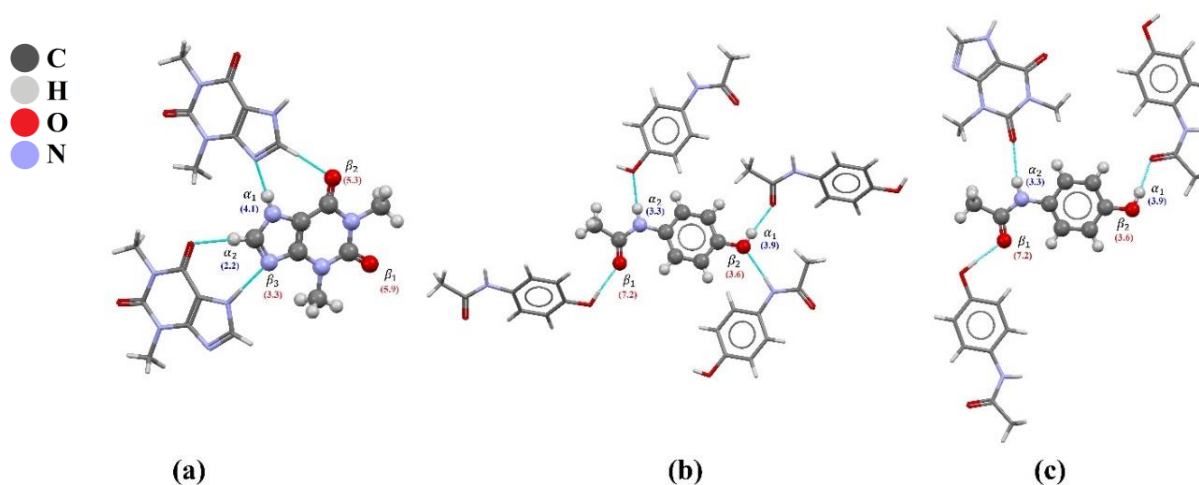
Figure 4-4 Crystal structures of (a) theophylline form II (BAPLOT06), (b) acetaminophen form II (HXACAN46), (c) an acetaminophen cocrystal (KIGLUI01). CSD identifiers are provided in parenthesis.

Acetaminophen was another corner case that defied this predictive approach. It has 3 polymorphs (a stable, a metastable, and an unstable polymorph),(Nichols & Frampton, 1998) with having the crystal structures of the stable and metastable polymorphs reported. Since H-bond pairings in these two polymorphs were identical, for conciseness, only one polymorph is illustrated in Figure 4-4.b.

Intriguingly, $\alpha_1\beta_1$ and $\alpha_2\beta_2$ are both observed. Thus, the problem is not associated with the hierarchical mapping of donors/acceptors. Acetaminophen has 2 comparable strong H-bond

donors (3.9, 3.3) but the acceptors have a meaningful difference in their strength (7.2, 3.6). We observed in many cocrystals of acetaminophen, such as Figure 4-4.c, the cocrystal is stabilized by replacing the second acceptor (3.6) with a stronger one from the coformer, while the strongest API-API interaction remains unchanged.

One way to avoid this corner case of Method A in Figure 4-5, is to change the way cocrystal energies are calculated: instead of using the cross product $\left(\alpha_{1,C}\beta_{1,A} + \alpha_{1,A}\beta_{1,C}\right)$, we can merge the α, β lists of coformer and API, sort them, and then calculate the cocrystal energies with hierarchal mapping (Musumeci et al., 2011). The visual representation of this method (Method B) is shown in Figure 4-5. The logic behind this approach is that the pairing of the donors/acceptors should not be discriminated against based on whether they are a part of the API or the coformer. By looking at the example given in Figure 4-5, one can see that the first pairing energy in cocrystal is, in fact, an API-API interaction (4.2∗7.2), but the second interaction is API-Coformer (3.8∗5.0). Therefore, a cocrystal is still feasible, without breaking the primary API-API interaction.
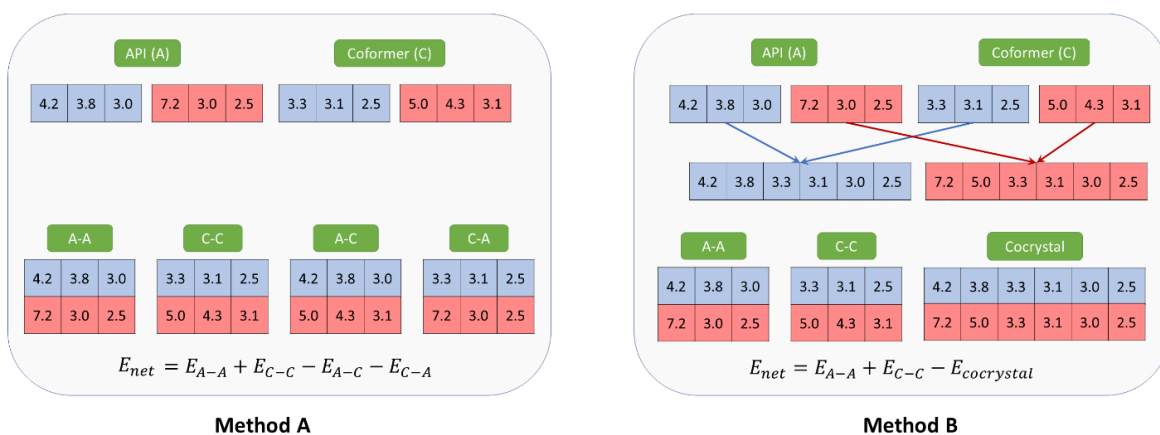
Figure 4-5 Explanation of 2 different ways to compute net H-bond energies for multicomponent crystals.

The way Method B computes the net energy implicitly suggest that the E_{net} can be at most zero. In other words, the formation of cocrystal can only improve the pairing energies with a decrease in E_{net}. Thus, we cannot have a head-to-head comparison for the performance of Method A and B. The main goal of Method B would be to rank the list of coformers for

experimental screening; the top of the list would be enriched with successful multicomponent crystals while the bottom is depleted.

In summary, two potential conditions that challenge the H-bond energy calculations are identified and the solutions to avoid these pitfalls are discussed. First, some APIs defy the hierarchal mapping of donor/acceptor sites, mostly due to the emergence of crystal structure governed by kinetic or steric effects that overrule thermodynamics. If the crystal structure of the target API is known, it is preferred to compute the API-API interactions based on the observed interactions. Second, a different approach in the calculation of pairing energies in cocrystals is suggested (Method B). Examples, where Method B is superior to Method A in cocrystal energy calculations, are discussed. Although without setting a non-zero cut-off, no success rate can be defined for Method B, it serves the purpose of screening by providing a ranked list of coformers.

4.3.3 Applying HBE, HBP, and pKa Rules on Experimental Screening of Two Model APIs

As stated before, the predictive strength of any method cannot be tested without negative results. However, negative results (recrystallization instead of cocrystallization) need to be established carefully with various analytical characterization methods. To prepare a trusted list of experiments, we used liquid-assisted grinding to grind each mixture of API-Coformer for half an hour in the presence of methanol. PXRD, SCXRD, and DSC were employed for the identification of the new phases. The procedure details are available in the Experimental Section.
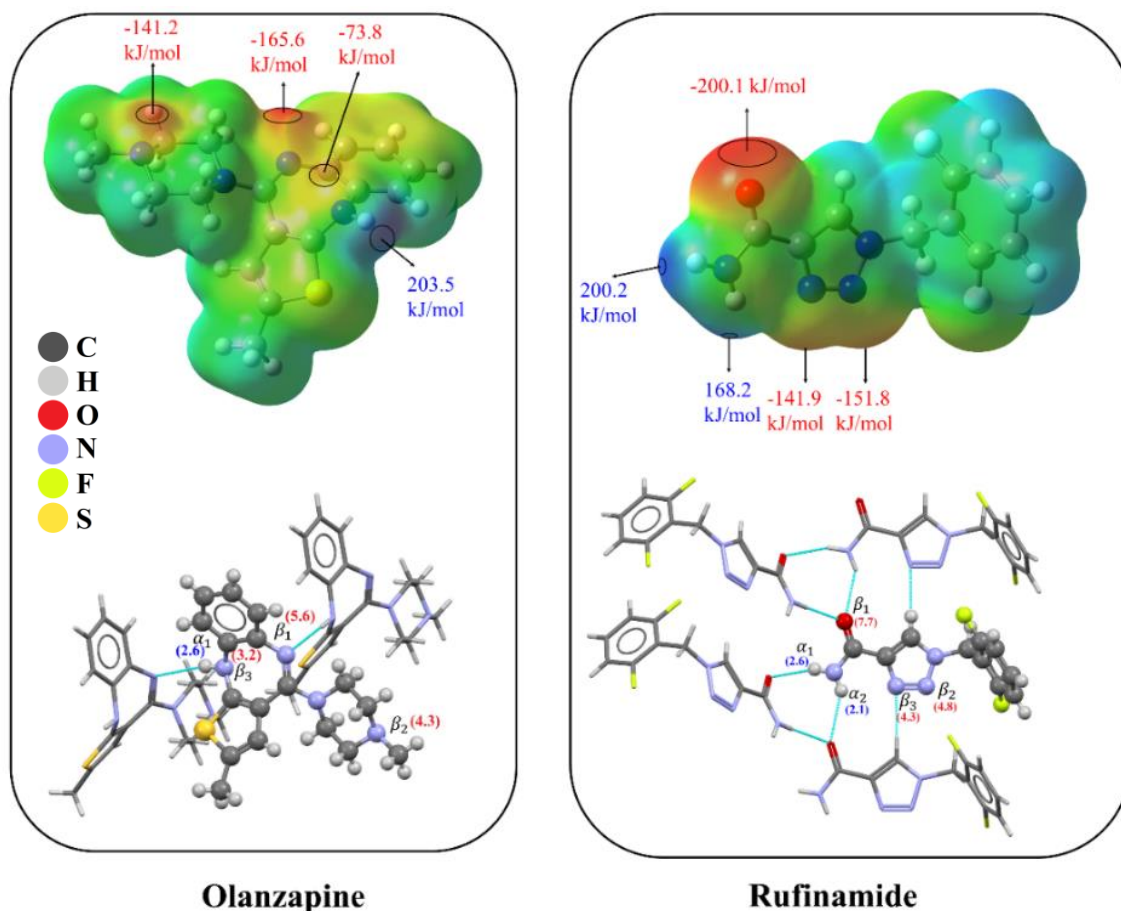
Figure 4-6 top: ESP of the gas-phase optimized geometry of Olanzapine and Rufinamide with local minima/maxima; bottom: interactions in crystal structures of anhydrous Olanzapine (UNOGIN03) and Rufinamide (ZEHZEU). CSD identifiers/Refcodes of crystal structures are provided in parenthesis.

According to the crystal structures depicted in Figure 4-6, the primary interaction in Olanzapine is $\alpha_1\beta_1$, forming a $C(8)$ chain. Olanzapine follows the presumption of hierarchical mapping of donors and acceptors. Rufinamide also forms the $\alpha_1\beta_1$ pair, but the overall network of H-bonds are significantly more intertwined, forming a double ring synthon of $R_2^2(10)R_3^2(8)$ between 3 Rufinamide molecules. By considering the primary interaction only, the API-API interaction will be underestimated. However, the sorted list of coformers would still be applicable for cocrystal screening.

The summary of the experimental screening and ranked list of coformers for Rufinamide and Olanzapine are shown in Table 4-1 and Table 4-2 , respectively. We noticed that even though all methods had acceptable performance in ranking the coformers, they portrayed different images on which coformer is ranked higher. This observation conveyed that a combination of these methods may lead to a better classifier. As an attempt to improve the prediction outcome, Scikit-learn (Pedregosa et al., 2011), a machine learning library for Python, was employed. Using the experimental results as the training dataset for the Logistic Regression enabled us to obtain optimized weights of individual factors of these 3 methods (E_{A-A}, E_{C-C}, E_{A-C}, E_{C-A}, $E_{Cocrystal}$, P_{A-A}, P_{C-C}, P_{A-C}), as well as the number of α_A-β_C and α_C-β_A. The weights are given in Table A4-4. This method is referred to as Multivariate Logistic Regression model in Table 4-1 and Table 4-2.

Table 4-1 Rufinamide screening, with ranked lists of coformers for different predictive approaches. Experiment (PM: physical mixture, CC: cocrystal); ΔpKa (Rufinamide conjugated base – Coformer acid); HBE: hydrogen bond energy; HBP: hydrogen bond propensity; MLR: Multivariate Logistic Regression. Dashed line: the number of positive cases; Bold and green coformers denote the positive experiments.

List of Coformers	Exp.	ΔpKa	HBE Meth. A	HBE Meth. B	HBP	MLR
1,7-Phenanthroline (1,7-Phen)	PM	N/A	TA	TA	SA	**TMA**
'-2'-Bipyridine (bpy)	PM	N/A	3-Cl-ManA	OA	3-Cl-ManA	TA
2-Aminoterephthalic acid (2-ATPA)	PM	-5.0	OA	**TMA**	TU	OA
2-Ethylimidazole (2-EIm)	PM	N/A	SalA	FA	**TMA**	MalicA
3-Aminobenzoic acid (3-ABA)	PM	-14.8	Cat	TPA	PhOH	FA
3-Chloromandelic acid (3-Cl-ManA)	PM	-10.4	FA	Res	DEX	CA
3-Hydroxybenzoic acid (3-HBA)	PM	-5.8	**TMA**	SA	SMT	TPA
4,7-Phenanthroline (4,7-Phen)	PM	-4.2	MalicA	HQ	StA	2-ATPA
4-Aminobenzoic acid (4-ABA)	PM	-4.8	TPA	MalonA	NAM	Res
4-Azabenzimidazole (4-Aza-BZ)	PM	N/A	MalonA	MalicA	Res	HQ
5-Fluorouracil (5-FU)	PM	-5.8	PhOH	TU	AnA	SA
9-Formylphenanthrene (Phen-9-ald)	PM	-11.6	SA	3-HBA	SuA	MalonA
Acesulfame (Ace)	PM	-8.2	3-HBA	CA	TAA	5-FU
Acetaminophen (APAP)	PM	-5.0	Ace	2-ATPA	Phen	Cat
Adenine (A)	PM	-10.5	ManA	SalA	INA	3-HBA
Anthranilic acid (AnA)	PM	-11.3	Res	Cat	4-ABA	3-Cl-ManA
Ascorbic acid (AsA)	PM	-5.9	SAC	PhOH	OA	TU
Benzimidazole (BIM)	PM	-5.2	HQ	AsA	APAP	A
Benzoic acid (BA)	PM	-13.3	TU	3-Cl-ManA	Im	SalA
Catechol (Cat)	PM	-5.1	CA	5-FU	CA	SuA
Citric acid (CA)	PM	-10.3	2-ATPA	SuA	MEL	AsA
Dextrose (DEX)	PM	-4.1	INA	Ace	MENTH	PhOH
Fumaric acid (FA)	PM	-12.3	5-FU	DEX	MalonA	SAC
Hydroquinone (HQ)	PM	-4.4	TAA	SAC	BIM	TAA
Imidazole (Im)	PM	-10.7	AnA	TAA	TPA	AnA
Isonicotinamide (INAM)	PM	-14.4	A	APAP	TA	Ace
Isonicotinic acid (INA)	PM	-14.7	BA	AnA	HQ	3-ABA
Maleic acid (MaleicA)	PM	-4.6	StA	9,10-PQ	VAN	ManA
Malic acid (MalicA)	PM	-3.9	SuA	Phen	SalA	INA
Malonic acid (MalonA)	PM	-4.2	AsA	ManA	UREA	MaleicA
Mandelic acid (ManA)	PM	-3.4	3-ABA	UREA	INAM	4-ABA
Melamine (MEL)	PM	-4.8	INAM	1,7-Phen	ManA	INAM
Menthol (MENTH)	PM	-16.7	NAM	INAM	BA	APAP
Nicotinamide (NAM)	PM	-20.6	MaleicA	SMT	AsA	SMT
Oxalic acid (OA)	PM	-14.4	4-ABA	INA	Phen-9-ald	BA
Phenanthrenequinone (9,10-PQ)	PM	-2.4	VAN	4-Aza-BZ	Cat	NAM
Phenanthridine (Phen)	PM	N/A	APAP	Phen-9-ald	9,10-PQ	StA
Phenol (PhOH)	PM	N/A	9,10-PQ	4-ABA	SAC	DEX
Resorcinol (Res)	PM	-11.0	MENTH	BA	MalicA	bpy
Saccharin (SAC)	PM	-10.3	SMT	NAM	A	MEL
Salicylic acid (SalA)	PM	-2.9	BIM	A	MaleicA	MENTH
Stearic acid (StA)	PM	-3.8	MEL	4,7-Phen	2-ATPA	VAN
Succinic acid (SA)	PM	-6.0	2-EIm	StA	bpy	4-Aza-BZ
Sulfamethazine (SMT)	PM	-4.6	UREA	3-ABA	4-Aza-BZ	BIM
Sulfanilic acid (SuA)	PM	-8.0	Im	MaleicA	5-FU	UREA
Tartaric acid (TA)	PM	2.4	DEX	MEL	1,7-Phen	Im
Terephthalic acid (TPA)	PM	-3.7	Phen-9-ald	VAN	4,7-Phen	2-ethylIm
Thioacetamide (TAA)	PM	-4.3	1,7-Phen	MENTH	FA	4,7-Phen
Thiourea (TU)	PM	-14.4	4,7-Phen	bpy	3-HBA	1,7-Phen
Trimesic acid (TMA)	CC	-16.2	Phen	2-EIm	2-EIm	Phen-9-ald
Urea (UREA)	PM	-4.1	4-Aza-BZ	BIM	Ace	Phen
Vanillin (VAN)	PM	-16.7	bpy	Im	3-ABA	9,10-PQ
Ture Positive Rate in %			0%	0%	0%	100%
Number of misplaced coformers			1	1	1	0

Table 4-2 Olanzapine screening, with ranked lists of coformers based on different predictive approaches. Experiment (PM: physical mixture, CC: cocrystal, S: salt, A: amorphous); The second column shows the ΔpKa (Olanzapine conjugated base – Coformer best acid); HBE: hydrogen bond energy; HBP: hydrogen bond propensity; MLR: Multivariate Logistic Regression; Dashed line: the number of positive cases; Bold and green coformers denote the positive experiments.

List of Coformers	Exp.	ΔpKa	HBE Meth. A	HBE Meth. B	HBP	MLR
1,7-Phenanthroline (1,7-Phen)	PM	N/A	**TA**	**TA**	StA	**OA**
'-2'-Bipyridine (bpy)	PM	N/A	**SalA**	**OA**	5-FU	**FA**
2-Aminoterephthalic acid (2-ATPA)	S	3.21	**TMA**	**TMA**	OA	**Res**
2-Ethylimidazole (2-EIm)	PM	-6.56	**Cat**	**FA**	**SalA**	**HQ**
3-Aminobenzoic acid (3-ABA)	PM	2.43	**FA**	**TPA**	Cat	**TPA**
3-Chloromandelic acid (3-Cl-ManA)	A	4.09	**OA**	4-Aza-BZ	**HQ**	**TMA**
3-Hydroxybenzoic acid (3-HBA)	S	3.4	**PhOH**	**Res**	**PhOH**	**SA**
4,7-Phenanthroline (4,7-Phen)	PM	N/A	9,10-PQ	**HQ**	**Res**	**MalonA**
4-Aminobenzoic acid (4-ABA)	S	2.47	**TPA**	**SA**	**ManA**	**Cat**
4-Azabenzimidazole (4-Aza-BZ)	PM	-3.33	**3-Cl-ManA**	**MalonA**	**3-HBA**	**3-HBA**
5-Fluorouracil (5-FU)	PM	0.06	**Res**	**CA**	**3-Cl-ManA**	**3-Cl-ManA**
9-Formylphenanthrene (Phen-9-ald)	PM	N/A	**SA**	TU	**FA**	**TA**
Acesulfame (Ace)	PM	3.27	**HQ**	**MalicA**	DEX	**CA**
Acetaminophen (APAP)	PM	-2.22	Ace	**3-HBA**	**CA**	**SalA**
Adenine (A)	PM	-3.05	**3-HBA**	9,10-PQ	VAN	**MalicA**
Anthranilic acid (AnA)	S	2.35	**MalonA**	**2-ATPA**	bpy	**PhOH**
Ascorbic acid (AsA)	PM	3.08	**SAC**	AsA	9,10-PQ	**AnA**
Benzimidazole (BIM)	PM	-5.01	TU	**SalA**	BIM	**SAC**
Benzoic acid (BA)	S	3.16	**2-ATPA**	**PhOH**	**MaleicA**	TAA
Catechol (Cat)	CC	-2.1	**AnA**	**SuA**	4-Aza-BZ	**2-ATPA**
Citric acid (CA)	A	4.19	TAA	5-FU	**MalonA**	Ace
Dextrose (DEX)	PM	-4.06	**CA**	APAP	4,7-Phen	5-FU
Fumaric acid (FA)	S	3.89	**MalicA**	**Cat**	INA	**ManA**
Hydroquinone (HQ)	CC	-2.44	**ManA**	UREA	**SA**	**MaleicA**
Imidazole (Im)	PM	-6.16	**BA**	TAA	**4-ABA**	**SuA**
Isonicotinamide (INAM)	PM	-6.47	Phen	**AnA**	MENTH	INA
Isonicotinic acid (INA)	PM	3.6	A	**3-Cl-ManA**	1,7-Phen	INAM
Maleic acid (MaleicA)	S	4.39	INA	Ace	**TA**	A
Malic acid (MalicA)	S	4.04	StA	**4-ABA**	**MalicA**	TU
Malonic acid (MalonA)	S	4.81	1,7-Phen	**SAC**	**2-ATPA**	APAP
Mandelic acid (ManA)	A	3.49	**3-ABA**	NAM	TU	AsA
Melamine (MEL)	PM	-8.48	Phen-9-ald	**ManA**	**BA**	**BA**
Menthol (MENTH)	PM	-12.31	MENTH	**3-ABA**	**TMA**	NAM
Nicotinamide (NAM)	CC	-6.15	4,7-Phen	bpy	**AnA**	StA
Oxalic acid (OA)	S	5.88	INAM	**BA**	**SAC**	**3-ABA**
Phenanthrenequinone (9,10-PQ)	PM	N/A	MEL	INA	Phen-9-ald	MENTH
Phenanthridine (Phen)	PM	N/A	**4-ABA**	Phen	**SuA**	4-Aza-BZ
Phenol (PhOH)	CC	-2.78	NAM	1,7-Phen	Phen	bpy
Resorcinol (Res)	CC	-2.02	VAN	A	**TPA**	SMT
Saccharin (SAC)	A	5.3	5-FU	Phen-9-ald	APAP	**4-ABA**
Salicylic acid (SalA)	S	4.45	SMT	StA	AsA	VAN
Stearic acid (StA)	PM	2.29	**SuA**	2-EIm	**3-ABA**	BIM
Succinic acid (SA)	S	3.69	AsA	4,7-Phen	UREA	DEX
Sulfamethazine (SMT)	PM	0.25	APAP	BIM	MEL	Im
Sulfanilic acid (SuA)	A	10.63	UREA	**MaleicA**	Ace	2-EIm
Tartaric acid (TA)	S	4.52	bpy	Im	A	MEL
Terephthalic acid (TPA)	S	3.92	BIM	INAM	INAM	4,7-Phen
Thioacetamide (TAA)	PM	-6.12	2-EIm	DEX	Im	1,7-Phen
Thiourea (TU)	PM	-7.98	Im	MEL	NAM	Phen-9-ald
Trimesic acid (TMA)	A	4.1	DEX	MENTH	2-EIm	UREA
Urea (UREA)	PM	-8.49	**MaleicA**	SMT	TAA	Phen
Vanillin (VAN)	PM	-0.57	4-Aza-BZ	VAN	SMT	9,10-PQ
Ture Positive Rate in %			**84%**	**68%**	**60%**	**88%**
Number of misplaced coformers			**4**	**8**	**10**	**3**

As shown in Table 4-1 , Rufinamide proved to be a challenging API for cocrystallization, with only one positive result out of 52 experiments. Figure A4-1 shows the experimental evidence for Rufinamide-Trimesic acid cocrystal. Since the ΔpKa is -16.3, reflecting the fact that Rufinamide does not have any ionizable group, Rufinamide-Trimesic acid multicomponent crystal can be labelled as a cocrystal. Trimesic acid was ranked close to the top of the list of individual H-bond methods (7th, 3rd, and 4th). However, the multivariate logistic regression model was able to correctly rank this coformer at the top of the list.

Table 4-2 reveals the summary of Olanzapine screening that had 25 positive results out of 52 experiments and previous works (Andrusenko et al., 2020; Chakrabarti et al., 1992; da Costa et al., 2019, 2020; Gong et al., 2021; Hickey & Remenar, 2004; Keltjens, 2005; Koźluk, 2007; Nanubolu & Ravikumar, 2017; Surampudi et al., 2020; Thakuria & Nangia, 2013). Figure A4-2 shows the experimental evidence for Olanzapine multicomponent solid forms. To qualitatively compare these predictive approaches, the receiver operating characteristic (ROC) plot and the respective area under the curve (AUC) for each method are illustrated in Figure 4-7. ROC, defined as the true positive rate (true positives/all positive cases) against the false positive rate (false positives/all negative cases), illustrates the ability of a binary classifier to distinguish between positive and negative instances. The dashed line is known as the no-discrimination line, while a perfect classifier follows the left x-axis and the top y-axis (AUC = 1). AUC measures the entire area of the ROC curve and can be interpreted as the probability that a classifier ranks a random positive instance higher than a negative sample.
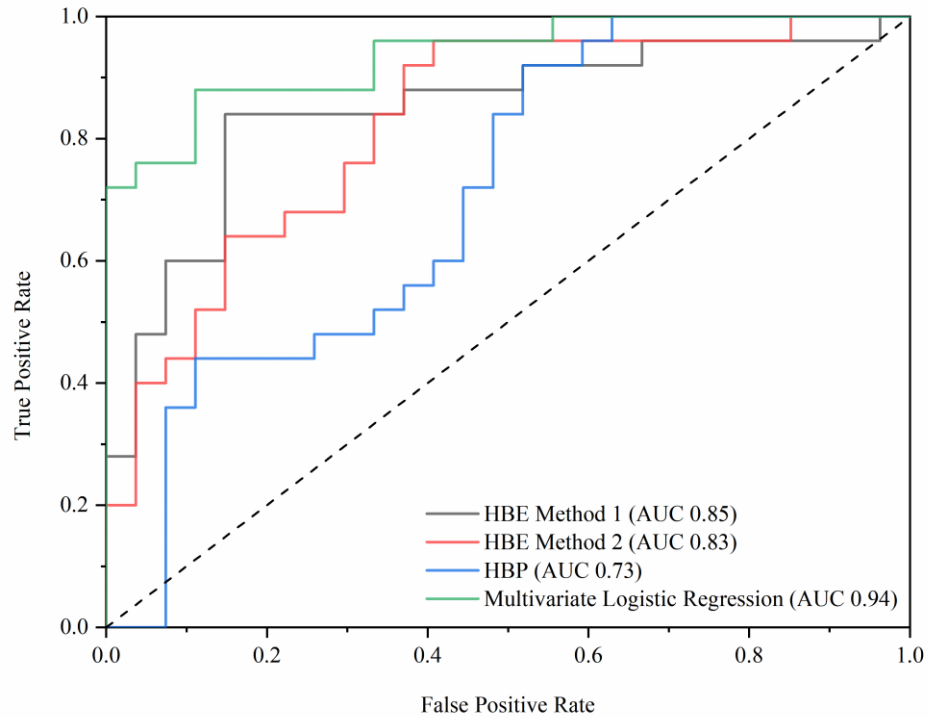
Figure 4-7 Receiver operating characteristic (ROC) plot of 4 prediction approaches of Olanzapine screening.

AUCs of both HBE methods were comparable (0.85 and 0.83), HBP showed a lower AUC (0.73), while the multivariate logistic regression model showed excellent classification, with 18 out of 25 positive cases (72%) being uninterruptedly ranked at the top of the list. The summary of the advantages and limitations of different methods discussed in this paper is provided in Table 4-3.

Table 4-3 Comparison of the advantages and limitations of HBE Methods, HBP, and multivariate logistic regression.

Approach	HBE Method A	HBE Method B	HBP	multivariate logistic regression
Advantages	▪ can be fully automated ▪ the α, β library of coformers can be reused for every API	▪ correctly predicts the cocrystal stability when the first homogenous interaction is not replaced ▪ can be fully automated ▪ the α, β library of coformers can be reused for every API	▪ relies on existing experimental data only ▪ considers up to seven model variables: competition, donor and acceptor atoms of the function group, their steric density, and aromaticity	▪ improves with more data (characteristics of machine learning) ▪ performs the best among these four methods ▪ reduces the impact of corner cases by combining different approaches
Limitations	▪ under-represent the cocrystal stability when the first homogenous interaction is not replaced. e.g., acetaminophen ▪ requires knowledge of quantum calculations and access to the Gaussian program	▪ requires knowledge of quantum calculations and access to the Gaussian program	▪ molecules with functional groups without sufficient reported structures. e.g., acesulfame, benzimidazole, saccharin, thiourea, urea ▪ time-consuming: the procedure needs to be done for each pair of API-coformer ▪ available only in the fully licensed version of CSD Mercury	▪ requires raw data for three methods (ESP and the CSD database)

All of the methods discussed in this paper predict whether a multicomponent solid form is favorable over its individual components, without labeling which subcategory it belongs to. In particular, salts and cocrystals need to be differentiated. For each identified multicomponent solid form with deposited crystal structure on CSD, ΔpKa is used to separate salts from cocrystals. All of the cases were successfully categorized as salts and cocrystals with the ΔpKa rule. Table A4-5 summarizes the probability of salt versus cocrystal based on equations given in reference (Cruz-Cabeza, 2012).

4.4 Conclusion

In the quest to find pharmaceutical multicomponent solid forms, a theoretical classifier to identify potential coformers rapidly is highly in demand. We examined the coformer classification robustness of simulation-based hydrogen bond energy (HBE) and knowledge-based hydrogen bond propensity (HBP) calculations. After an in-depth study of molecular electrostatic potential (ESP) maps of 330 coformers, pitfalls of using crystal structures geometries and gas-phase optimized geometries are discussed. Next, reverse prediction of salts/cocrystals of six active pharmaceutical ingredients (API) on Cambridge Structural Database (CSD) were performed with ESP analysis. The prediction showed systematic errors for two of the APIs, for which suggestions are provided to avoid those situations. Thereafter, HBE and HBP calculations were put to the test against the experimental screening of 2 APIs (Olanzapine and Rufinamide) with 52 organic coformers. Even though the performance of HBE and HBP were comparable, the ranked lists of coformers were significantly different. Therefore, multivariate logistic regression was employed to combine the identifiers of HBE and HBP and suggest a superior classifier that outperformed the individual methods. In the case of Olanzapine, 18 out of 25 positive cases were uninterruptedly identified by the multivariate logistic regression at the top of the list. In conclusion, the integration of chemistry-based identifiers and machine learning algorithms can be the ultimate classifier for cocrystal screening.

4.5 References

Andrusenko, I., Potticary, J., Hall, S. R., & Gemmi, M. (2020). A new olanzapine cocrystal obtained from volatile deep eutectic solvents and determined by 3D electron diffraction. *Acta Crystallographica Section B, Structural Science, Crystal Engineering and Materials*, *76*(Pt 6), 1036–1044.

Bobrovs, R., Seton, L., & Dempster, N. (2015). The reluctant polymorph: investigation into the effect of self-association on the solvent mediated phase transformation and nucleation of theophylline. *CrystEngComm*, *17*(28), 5237–5251.

Chakrabarti, J. K., Hotten, T. M., & Tupper, D. E. (1992). *2-Methyl-Thieno-Benzodiazepine* (Patent No. US005229332A). United States Patent.

Cruz-Cabeza, A. J. (2012). Acid–base crystalline complexes and the pKa rule. *CrystEngComm*, *14*(20), 6362–6365.

da Costa, N. F., Fernandes, A. I., & Pinto, J. F. (2020). Measurement of the amorphous fraction of olanzapine incorporated in a co-amorphous formulation. *International Journal of Pharmaceutics*, *588*, 119716.

da Costa, N. F., Pinto, J. F., & Fernandes, A. I. (2019). Co-amorphization of olanzapine for solubility enhancement. *Annals of Medicine*, *51*(sup1), 87.

Delori, A., Galek, P. T. A., Pidcock, E., Patni, M., & Jones, W. (2013). Knowledge-based hydrogen bond prediction and the synthesis of salts and cocrystals of the anti-malarial drug pyrimethamine with various drug and GRAS molecules. *CrystEngComm*, *15*(15), 2916–2928.

Desiraju, G. R., Vittal, J. J., & Ramanan, A. (2011). *Crystal Engineering: A Textbook*. World Scientific Publishing Co. Pte. Ltd.

Fábián, L. (2009). Cambridge Structural Database Analysis of Molecular Complementarity in Cocrystals. *Crystal Growth & Design*, *9*(3), 1436–1443.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. a., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. a., Nakatsuji, H., Li, X., Caricato, M., Marenich, a. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., … Fox, D. J. (2016). *Gaussian 16, Revision b.01* (Gaussian 16, Revision b01). Gaussian, Inc., Wallin CT.

Friščić, T. (2018). Mechanochemistry in Co-crystal Synthesis. In *Co-crystals: Preparation, Characterization and Applications* (pp. 147–193). The Royal Society of Chemistry.

Galek, P. T. A., Allen, F. H., Fábián, L., & Feeder, N. (2009). Knowledge-based H-bond prediction to aid experimental polymorph screening. *CrystEngComm*, *11*(12), 2634–2639.

Gong, W., Mondal, P. K., Ahmadi, S., Wu, Y., & Rohani, S. (2021). Cocrystals, Salts, and Salt-Solvates of Olanzapine; selection of coformers and improved solubility. *International Journal of Pharmaceutics*, 608, 121063.

Grecu, T., Hunter, C. A., Gardiner, E. J., & McCabe, J. F. (2014). Validation of a Computational Cocrystal Prediction Tool: Comparison of Virtual and Experimental Cocrystal Screening Results. *Crystal Growth & Design*, *14*(1), 165–171.

Hickey, M. B., & Remenar, J. (2004). *Novel olanzapine forms and related methods of treatment* (Patent No. US 2006/0223794 A1). United States Patent Application Publication.

Hunter, C. A. (2004). Quantifying Intermolecular Interactions: Guidelines for the Molecular Recognition Toolbox. *Angewandte Chemie International Edition*, *43*(40), 5310–5324.

Issa, N., Karamertzanis, P. G., Welch, G. W. A., & Price, S. L. (2009). Can the Formation of Pharmaceutical Cocrystals Be Computationally Predicted? I. Comparison of Lattice Energies. *Crystal Growth & Design*, *9*(1), 442–453.

JADE 7.0. (2020). Materials Data: Livermore, CA.

Karamertzanis, P. G., Kazantsev, A. V, Issa, N., Welch, G. W. A., Adjiman, C. S., Pantelides, C. C., & Price, S. L. (2009). Can the Formation of Pharmaceutical Cocrystals Be Computationally Predicted? 2. Crystal Structure Prediction. *Journal of Chemical Theory and Computation*, *5*(5), 1432–1448.

Karki, S., Friščić, T., Jones, W., & Motherwell, W. D. S. (2007). Screening for Pharmaceutical Cocrystal Hydrates via Neat and Liquid-Assisted Grinding. *Molecular Pharmaceutics*, *4*(3), 347–354.

Keltjens, R. (2005). *Stable salts of olanzapine* (Patent No. WO2005070938A1). World Intellectual Property Organization.

Koźluk, T. (2007). *New salts of olanzapine and method of their preparation* (Patent No. WO2007032695A1). World Intellectual Property Organization.

Macrae, C. F., Sovago, I., Cottrell, S. J., Galek, P. T. A., McCabe, P., Pidcock, E., Platings, M., Shields, G. P., Stevens, J. S., Towler, M., & Wood, P. A. (2020). Mercury 4.0: from visualization to analysis, design and prediction. *Journal of Applied Crystallography*, *53*(1), 226–235.

Majumder, M., Buckton, G., Rawlinson-Malone, C. F., Williams, A. C., Spillman, M. J., Pidcock, E., & Shankland, K. (2013). Application of hydrogen-bond propensity calculations to an indomethacin–nicotinamide (1 : 1) co-crystal. *CrystEngComm*, *15*(20), 4041–4044.

MarvinSketch 21.3. (2021). ChemAxon (chemaxon.com).

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*(3), 275–300.

Musumeci, D., Hunter, C. A., Prohens, R., Scuderi, S., & McCabe, J. F. (2011). Virtual cocrystal screening. *Chemical Science*, *2*(5), 883–890.

Nanubolu, J. B., & Ravikumar, K. (2017). Designing a new cocrystal of olanzapine drug and observation of concomitant polymorphism in a ternary cocrystal system. *CrystEngComm*, *19*(2), 355–366.

Nichols, G., & Frampton, C. S. (1998). Physicochemical Characterization of the Orthorhombic Polymorph of Paracetamol Crystallized from Solution. *Journal of Pharmaceutical Sciences*, *87*(6), 684–693.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Rodrigues, M., Lopes, J., Guedes, A., Sarraguça, J., & Sarraguça, M. (2020). Considerations on high-throughput cocrystals screening by ultrasound assisted cocrystallization and

vibrational spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *229*, 117876.

Sandhu, B., McLean, A., Sinha, A. S., Desper, J., Sarjeant, A. A., Vyas, S., Reutzel-Edens, S. M., & Aakeröy, C. B. (2018). Evaluating Competing Intermolecular Interactions through Molecular Electrostatic Potentials and Hydrogen-Bond Propensities. *Crystal Growth & Design*, *18*(1), 466–478.

Sarkar, N., & Aakeröy, C. B. (2020). Evaluating hydrogen-bond propensity, hydrogen-bond coordination and hydrogen-bond energy as tools for predicting the outcome of attempted co-crystallisations. *Supramolecular Chemistry*, *32*(2), 81–90.

Sarkar, N., Gonnella, N. C., Krawiec, M., Xin, D., & Aakeröy, C. B. (2020). Evaluating the Predictive Abilities of Protocols Based on Hydrogen-Bond Propensity, Molecular Complementarity, and Hydrogen-Bond Energy for Cocrystal Screening. *Crystal Growth & Design*, *20*(11), 7320–7327.

Sarkar, N., Sinha, A. S., & Aakeröy, C. B. (2019). Systematic investigation of hydrogen-bond propensities for informing co-crystal design and assembly. *CrystEngComm*, *21*(40), 6048–6055.

Surampudi, A. V. S. D., Rajendrakumar, S., Nanubolu, J. B., Balasubramanian, S., Surov, A. O., Voronin, A. P., & Perlovich, G. L. (2020). Influence of crystal packing on the thermal properties of cocrystals and cocrystal solvates of olanzapine: insights from computations. *CrystEngComm*, *22*(39), 6536–6558.

Thakuria, R., & Nangia, A. (2013). Olanzapinium Salts, Isostructural Solvates, and Their Physicochemical Properties. *Crystal Growth & Design*, *13*(8), 3672–3680.

Tiekink, E., & Zukerman-Schpector, J. (2017). *Multi-Component Crystals*. De Gruyter.

Zhang, G. G. Z., Henry, R. F., Borchardt, T. B., & Lou, X. (2007). Efficient Co-crystal Screening Using Solution-Mediated Phase Transformation. *Journal of Pharmaceutical Sciences*, *96*(5), 990–995.

Chapter 5

Machine Learning-guided Prediction of Cocrystals from DFT-derived Point Clouds

A manuscript utilizing the content of this chapter has been prepared for submission and is currently under review.

5 Machine Learning-guided Prediction of Cocrystals from DFT-derived Point Clouds

Abstract

The design and synthesis of cocrystals have emerged as a promising crystal engineering strategy for enhancing the physicochemical properties of a diverse range of target molecules. A prediction strategy to identify whether a pair of target and auxiliary molecules would form a cocrystal or not can greatly accelerate the process of cocrystal discovery. In this study, we compiled and performed DFT calculations for 12,776 molecules (6,388 cocrystals). All entries in the database were obtained from successful and unsuccessful experimental attempts reported in the literature. Electrostatic potential (ESP) surfaces were then extracted from the DFT results and used for the development of four machine learning models (PointNet, ANN, RF, Ensemble). The Ensemble model, leveraging the complementary strengths of the PointNet, ANN, and RF models, demonstrated superior discriminatory performance with a BACC (0.942) and an AUC (0.986) on the unseen test data subset. To assess the performance of the models on individual molecules, we separated the cocrystals of caffeine, fumaric acid, and salicylic acid from the overall database. The Ensemble model exhibited remarkable robustness, classifying the 312 cocrystals in this subset into their respective classes with an average BACC of 98%.

5.1 Introduction

Cocrystals have gained significant attention in recent years due to their versatile applications. They are commonly used in the pharmaceutical industry to improve the solubility and bioavailability of poorly soluble drugs (Duggirala et al., 2016; Kavanagh et al., 2019; Xiouras et al., 2022). Additionally, organic charge-transfer cocrystals have led to performance improvements in solar cells (Goh et al., 2016), sensors (Ture et al., 2022), and transistors (Huang et al., 2019; S. Li et al., 2022). Energetic cocrystals have also given rise to high-energy and low-sensitivity materials such as explosives and propellants (Bennion & Matzger, 2021; Qiao et al., 2022; Sultan et al., 2022; Zhou et al., 2021). Cocrystals also have shown potential

in the pigment industry for altering colouristic properties of pigments by rearranging pigment components or incorporating alternative chromophores into the lattice (Z. Li et al., 2022; Yang et al., 2022). Therefore, cocrystals have the potential to revolutionize various industries by providing new materials with enhanced properties and applications.

When seeking new multicomponent molecular solids for a target compound, such as cocrystals, the selection of an auxiliary compound from a vast list of potential molecules can be a daunting task. In order to refine our search and focus on the most promising candidates, accurate theoretical tools are necessary to complement brute-force experimental screening methods.

Such a theoretical tool must be specifically designed to target non-covalent intermolecular interactions that drive the periodic assembly of molecules into a molecular crystal. It should consider the diverse range of interactions, such as strong and directional hydrogen bonds (also known as the master key of molecular recognition), halogen bonds, π-interactions, and numerous weak isotropic interactions (C···C, C···H and H···H). Furthermore, such a theoretical tool should incorporate the principle of close packing to maximize space utilization within the crystalline lattice.

At present, integrating all these factors into a chemistry-based theoretical framework that is both fast and efficient, yet still reliable enough to accurately predict cocrystals across a broad range of target and auxiliary compounds, remains a challenge.

Machine learning has gained significant popularity in recent years as a means to accelerate the discovery of new materials (including cocrystals) (Kumar & Nanda, 2021), enabling researchers to rapidly screen and narrow down the search from a large pool of promising candidates (Butler et al., 2018; Gao et al., 2022; Juan et al., 2021; Spellings & Glotzer, 2018; Xiouras et al., 2022).

Graph convolutional network (GCN) is by far the most common machine learning algorithm used for cocrystal prediction, in which molecules are represented as a set of matrices to capture atom connectivity and features (Devogelaer et al., 2020; Hao et al., 2022; Jiang et al., 2021; Xiao et al., 2022). Alternatively, various molecular descriptors have also been extracted and

inputted in machine learning models such as multivariate adaptive regression splines (Przybyłek et al., 2019), multivariable logistic regression (Ahmadi et al., 2021), random forest (RF) (Wang et al., 2020), artificial neural network (ANN) (Mswahili et al., 2021), support vector machine (Hao et al., 2022), and extreme gradient boosting (Yang et al., 2022).

We propose to incorporate spatial data into the prediction of a multicomponent crystal formation. The outer shape of individual molecules can be estimated by truncating the DFT-obtained electron density at a cut-off (0.002 e^-/bohr^3). The resulting 3D surface gives a physically-reasonable representation of the molecular surfaces available for intermolecular interactions. To determine how molecules interact with one another, the three-dimensional charge distributions of molecules are essential. Electrostatic potential (ESP) maps allow us to account for variably charged regions of a molecule.

The combination of ESP values and a three-dimensional molecular surface generates a four-dimensional representation of a molecule (x, y, z, ESP), which serves as feature-rich input data for deep learning algorithms. This representation provides a comprehensive understanding of the charge distributions and molecular shape, allowing for a more accurate prediction of cocrystals.

In this work, we used DFT-derived point cloud data (ESP maps) to predict the likelihood of cocrystal formation between molecular pairs. Two different approaches were employed: 1) PointNet (Qi et al., 2017) as a deep learning model to automatically extract important features, and 2) artificial neural network (ANN) and random forest (RF) models on manually extracted relevant features from ESP maps, such as hydrogen bond donors and acceptors (α, β) (Hunter, 2004; Musumeci et al., 2011), volume, surface area, and sphericity. Finally, to improve the robustness of the model, we employed Ensemble learning, which combined the results from the previous models (PointNet, ANN, and RF). The Ensemble model demonstrated superior discriminatory performance on the test data subset, surpassing other models. This approach holds significant promise as a predictive tool for identifying potential cocrystals during the screening process.

5.2 Result and Discussion

5.2.1 Data collection

The positive cocrystal database was obtained from the Cambridge Structural Database (CSD), while the negative cases were obtained from the literature. The detailed information can be found in the supporting information. The list of positive cases includes CSD Refcodes, while the negative cases are accompanied by the corresponding article in which the data was published.

We automated the process of obtaining ESP maps for 12,776 molecules, comprising 6,388 pairs. Of these, 5,394 were successful cocrystallization cases (positive), while 994 were unsuccessful experiments (negative). To collect the positive samples from the Cambridge Structural Database (CSD) (Groom et al., 2016), we utilized the CSD Python API module of the Mercury software and developed a Python script to search the database for the 5,394 reported cocrystals and download their CIF files. Subsequently, the coordinates of the target and auxiliary molecules (x,y,z) were saved into separate XYZ files. For negative samples, the coordinate files of the target and auxiliary molecules were obtained by the PubChem (Kim et al., 2023) Compound Identification (CID) code.

Geometry optimization and energy calculation were performed to relax the molecular structure, followed by the computation of volumetric data for electron density and ESP using DFT calculations at the B3LYP/6-311++G** level of theory in Gaussian 16 (Frisch et al., 2016). Finally, the isosurface of the electron density, labeled with their respective ESP values, was transformed into a set of points in 3D space, creating a point cloud data for the PointNet model.

5.2.2 PointNet

We adopted the PointNet (Qi et al., 2017) architecture to create a new deep-learning network for cocrystal prediction by processing ESP map data of both the target and auxiliary molecules. The network consists of two main parts: feature extraction and classification (see Figure 5-1). Two distinct sets of point cloud data, consisting of (x, y, z, ESP) information for both the target and auxiliary molecules, are utilized as the input for the algorithm. The feature extraction part

consists of four shared multilayer perceptron (MLP) with 1D convolutions applied to each point of the molecule's point cloud in order to capture the local spatial information of ESP maps. To avoid redundancy in the model's parameters, the feature extraction parameters are kept identical for both the target and auxiliary molecules. Max pooling reduces the dimensionality of features by taking the maximum value of a local pool to represent the whole pool. In our model, max pooling reduces the dimensionality of the features from (320, 3118) to (320, 1). The two resulting feature vectors from the target and auxiliary molecules are concatenated and inputted into the classification component of the network. The classification stage comprises four fully connected neural network layers with ReLU activation function, except for the last layer which has a sigmoid activation function to confine the output to a cocrystallization probability in the range of 0 to 1. The final binary prediction of cocrystallization is obtained by applying a threshold that maximizes the sum of the model's sensitivity and specificity over the test data.

The local maxima (ESP_{max}) and minima (ESP_{min}) of the ESP surface were accurately and rapidly determined using a Python script. These points were then assigned to the closest nucleus within the molecular structure. Thereafter, these values were used to calculate the H-bond donor parameter (α) and H-bond acceptor parameter (β) based on the equations illustrated in Figure 5-1. To account for steric and spatial features, the volume (V), area (A), and sphericity (Ψ) of the ESP maps were also calculated and considered as input features.

The lists of α and β for the target and auxiliary molecules within a potential cocrystal vary in length depending on the molecular structure. To standardize the length for ML purposes, the padding technique was utilized. The padded lists of α and β, along with geometric features (V, A, Ψ), were subsequently concatenated and inputted into the selected classification algorithms (artificial neural network or random forest). This process is visually represented in Figure 5-1.

5.2.3 Data Pre-processing

The dataset of 6,388 pairs of target and auxiliary molecules is randomly split into train and test sets using a 75:25 ratio. To ensure rotational invariance of the model, data augmentation is used by applying a random rotation matrix to the point cloud representation of the ESP maps of each molecule in the training set. Additionally, point features are normalized to the range of [0, 1] to account for differences in the scales of the point features.

To maintain uniform dimensionality of the input matrix for all molecules, we randomly sampled 3118 points (the number of points present in the smallest molecule) from each molecule's point cloud.
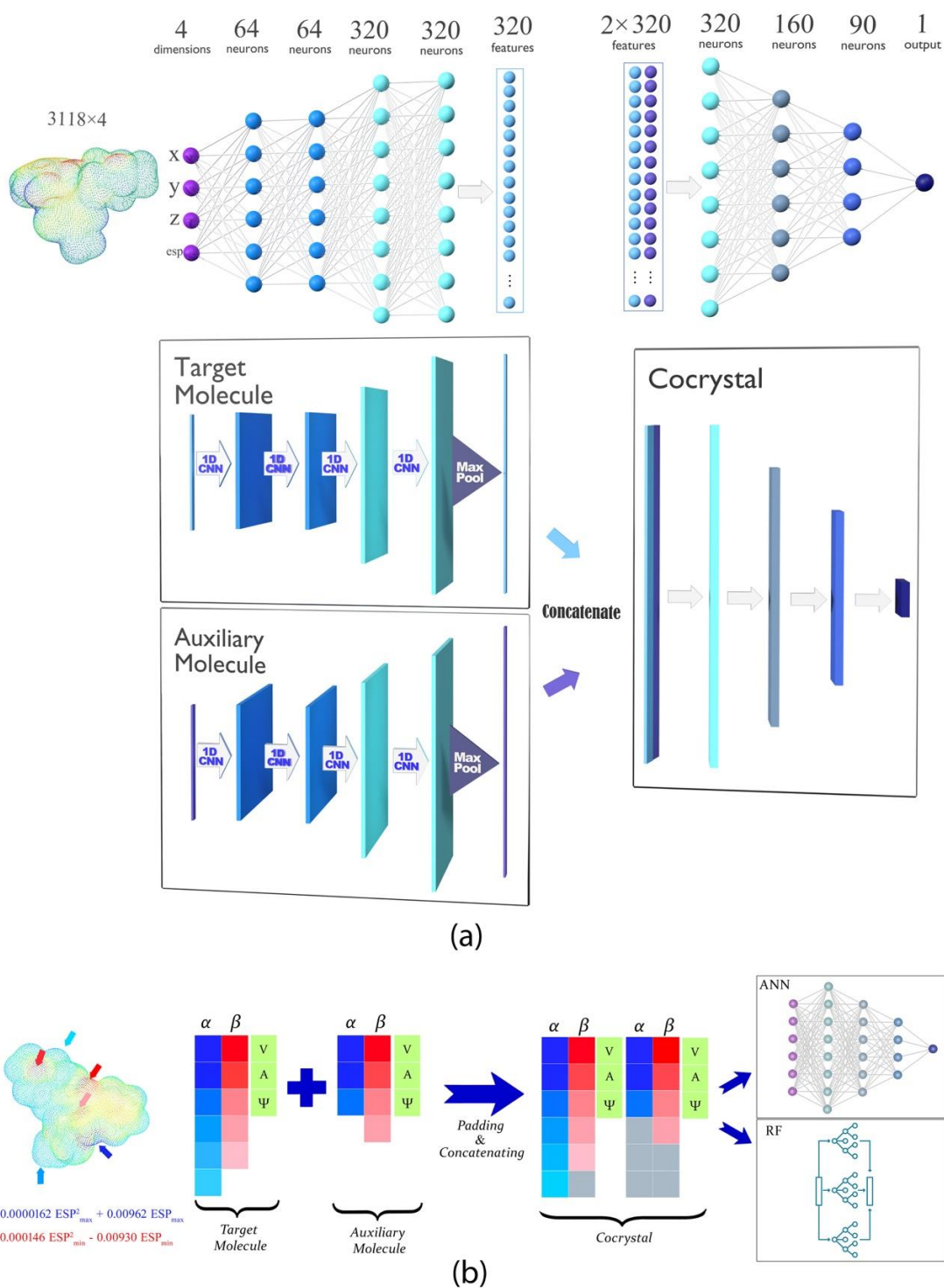
Figure 5-1 The schematic diagram of the (a) PointNet deep neural network, and (b) manual feature extraction and subsequent classification algorithms. The equations used to calculate α and β are also shown in (b).

5.2.4 Model Performance Analysis

Four models (PointNet, ANN, RF, Ensemble) were trained on our cocrystal dataset. Bayesian optimization was used to search for the optimal hyper-parameters (Wu et al., 2019) of each model (summarized in Table A5-1). The test dataset, consisting of the previously unseen 25% portion of the dataset, served as the basis for evaluating the performance of the four developed models. Key performance metrics, including True Positive Rate (TPR), True Negative Rate (TNR), Balanced Accuracy (BACC), and Area Under the ROC Curve (AUC), were employed in this evaluation.

The results, summarized in Figure 5-2, demonstrate the effectiveness of the trained PointNet and feature-driven models (ANN, RF) in achieving reliable performance for both positive and negative samples, as indicated by their high TPR and TNR values. Furthermore, the high values of BACC indicate a well-balanced performance of the models, accounting for the imbalanced nature of the dataset. The last key metric in our analysis is AUC, which is threshold-independent and robust to imbalanced data. AUC provides an assessment of the models' discriminative performance by indicating the probability of a classifier accurately ranking a randomly chosen positive sample above a randomly chosen negative sample.

Among all the models, RF and PoinNet exhibited the best TPR (0.949) and TNR (0.920), respectively. By leveraging the complementary strengths of each model, the Ensemble model achieved the highest BACC of 0.942 and an impressive AUC of 0.986.
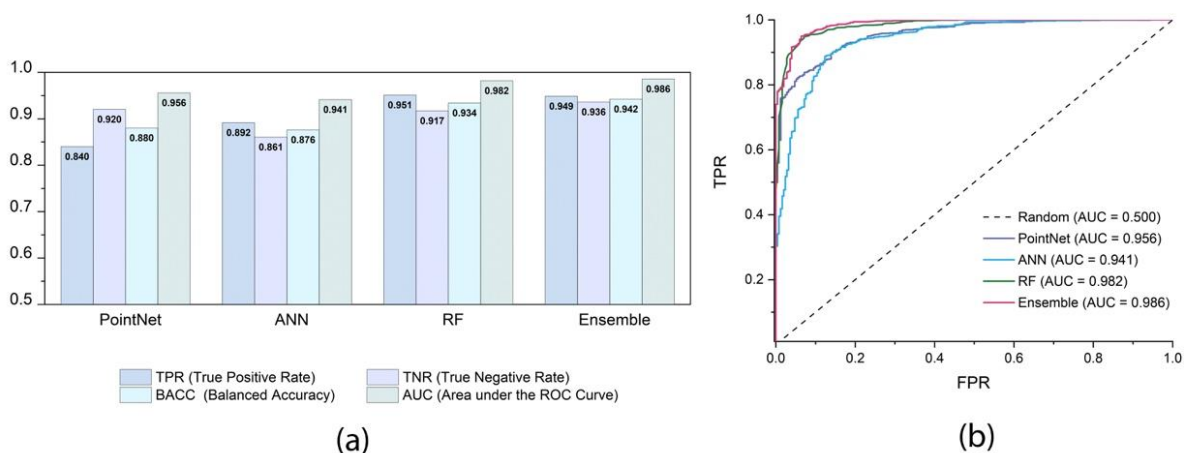
Figure 5-2 Comparison of (a) performance metrics (BACC, TPR, TNR, and AUC), and (b) ROC curves for the four developed ML models.

5.2.5 Visualizing the Classification Strength of PointNet and ANN Models

Since ML models deal with high dimensional data, it is necessary to use a non-linear dimensionality reduction algorithm to visualize the underlying parameter and features. We employed t-Distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten & Hinton, 2008) as the preferred non-linear dimensionality reduction algorithm to gain deeper insights into the performance of the trained models. t-SNE helps to visualize higher-dimensional data in a lower-dimensional space while preserving the original clustering in the high-dimensional space. This enables us to find and illustrate trends and patterns that are not apparent in the original high-dimensional data.

Figure 5-3 illustrates the two-dimensional t-SNE graphs that compare the raw data and the penultimate layer (last hidden layer) for PointNet and ANN models. In Figure 5-3.a, the raw data for the PointNet model displays complete randomness in the points, demonstrating minimal clustering or discernible patterns. However, after training the model (Figure 5-3.b), the feature clusters become more pronounced and distinct, signifying that the model has successfully learned to discriminate positive and negative samples, resulting in the formation of two distinct clusters. Similarly, Figure 5-3.c portrays the random distribution of the extracted features in the input data, whereas meaningful clustering is observed after the ANN model training (Figure 5-3.d).
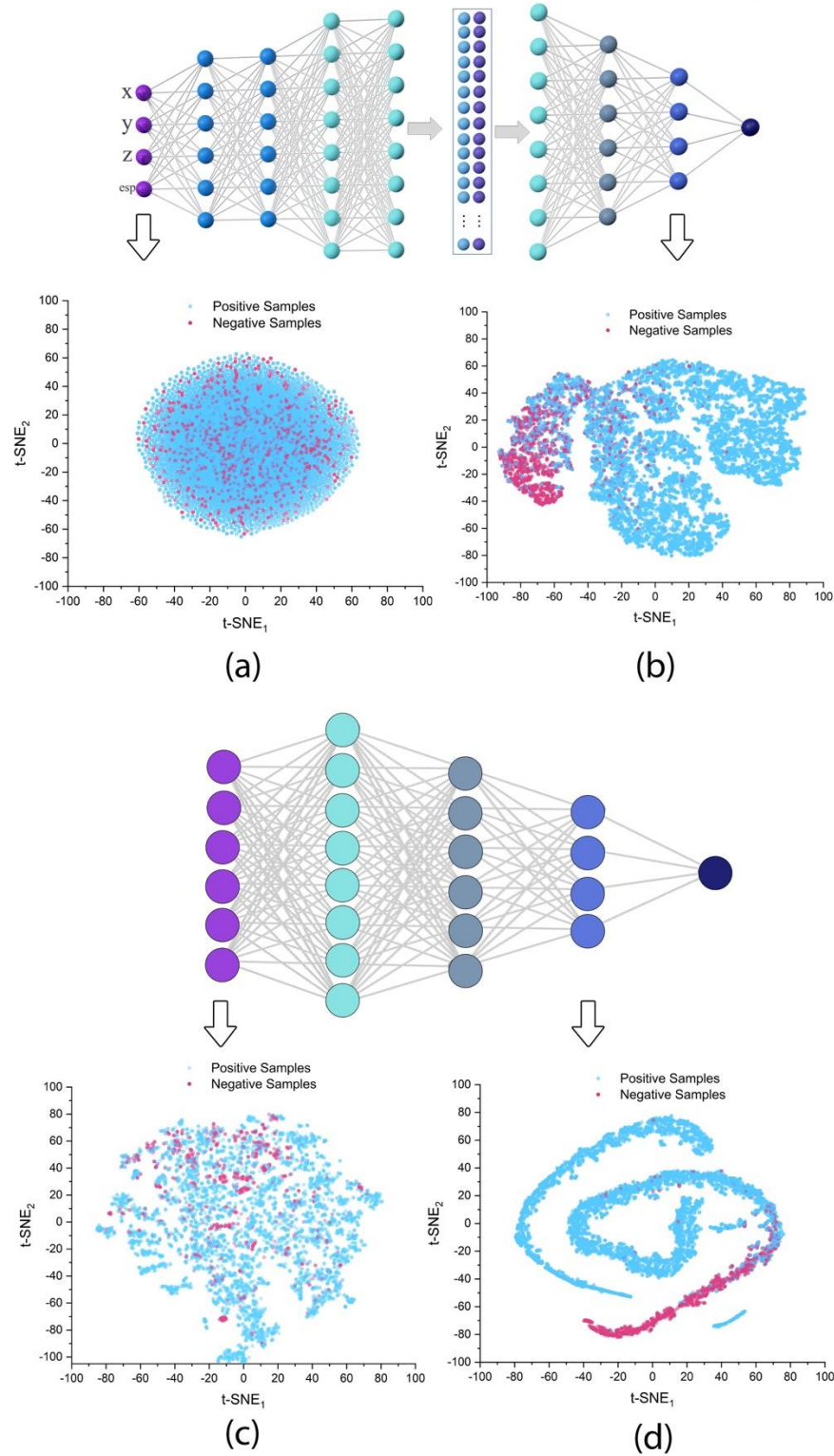
Figure 5-3 Two-dimensional t-SNE analysis for (a) raw point cloud data, (b) penultimate layer of PointNet, (c) manually extracted features $(\alpha, \beta, V, A, \Psi)$ at the input layer of ANN, and (d) penultimate layer of ANN model.

5.2.6 Performance Evaluation on Selected Molecules

To evaluate the performance of the developed models on individual molecules, we used Python scripting to identify and select three molecules with a substantial number of reported positive and negative cocrystals: caffeine, fumaric acid, and salicylic acid.

Caffeine is a stimulant employed in analgesics, alertness enhancers, and the treatment of pulmonary complications associated with premature birth. Fumaric acid is an acidulant with FDA GRAS (Generally Recognized as Safe) status and is commonly used in pharmaceutical cocrystals. Salicylic acid finds many applications mainly in skincare and topical treatments, and it is also widely used as an auxiliary compound for pharmaceutical cocrystals.

The breakdown of positive and negative cases for three model compounds is presented in Figure 5-4.a. A total of 312 cocrystals (214 positive and 98 negative) were included in the analysis. Figure 5-4.b-d displays the balanced accuracy (BACC), true positive rate (TPR), and true negative rate (TNR) of the four developed models for these case studies. The key finding is that the Ensemble model demonstrates remarkable robustness and accurately, classifying the coformers into their respective classes with an average balanced accuracy of 98%.
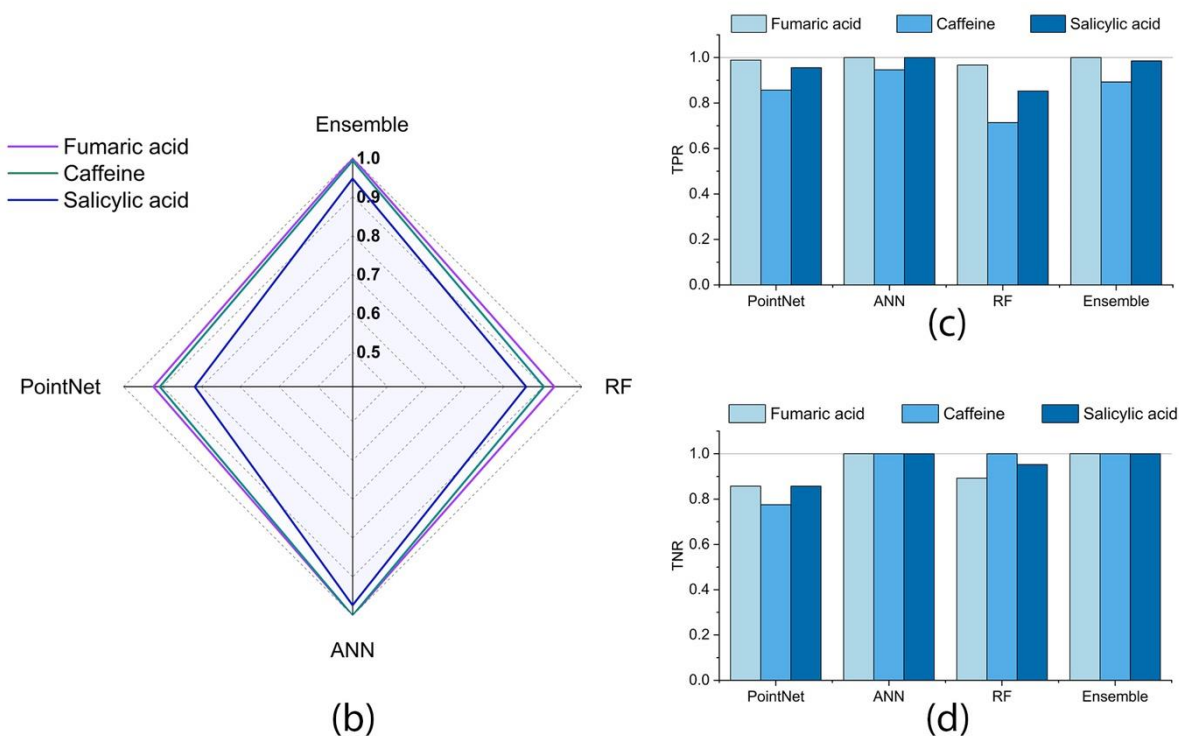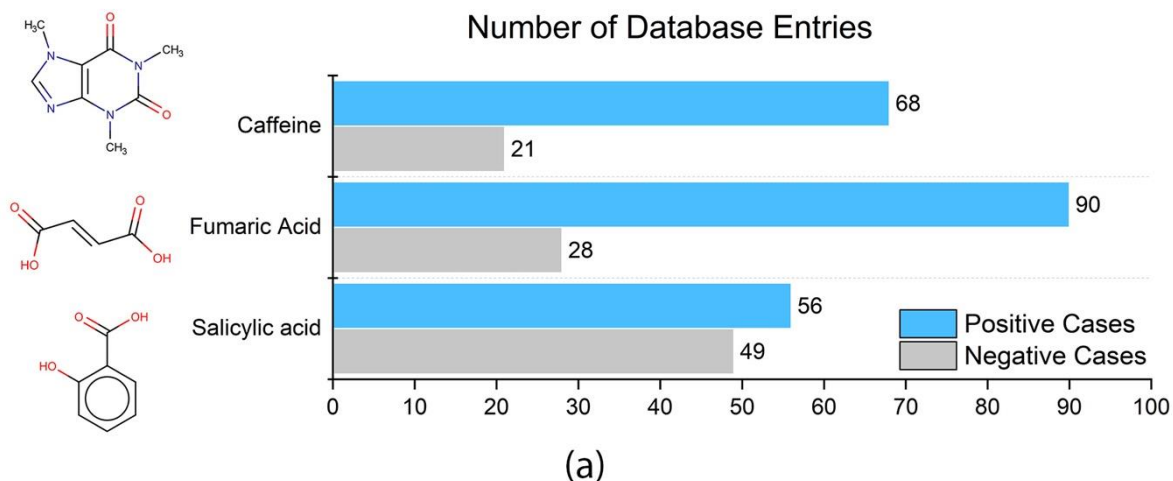
Figure 5-4 (a) The molecular structure and the number of database positive and negative entries of caffeine, fumaric acid, and salicylic acid. The performance of the four ML models over key metrics of (b) BACC, (c) TPR, and (d) TNR.

5.3 Conclusion

In this study, we developed two innovative approaches for cocrystal prediction utilizing ESP maps data obtained from DFT calculations. The first approach involved the use of PointNet, a deep learning model, to automatically extract important features and predict cocrystal formation. In the second approach, we extracted important features from ESP maps $(\alpha, \beta, V, A, \Psi)$ and utilized them to train ANN and RF models for cocrystal prediction. Lastly, by leveraging the complementary strengths of the abovementioned models, the Ensemble model was developed which demonstrated superior performance on the test data with BACC (0.942) and AUC (0.986). A robust dimensionality reduction method (t-SNE) helped us to visualize the random distribution of the input features and observe their successful clustering into positive and negative clusters following their processing through the networks (PointNet and ANN). Finally, we conducted case studies for three compounds in the database, namely caffeine, fumaric acid, and salicylic acid, which had a notable number of positive and negative entries. We observed that the Ensemble model exhibits remarkable robustness, accurately classifying the coformers of these 3 case studies into their respective classes with an average balanced accuracy (BACC) of 98%.

5.4 References

Ahmadi, S., Mondal, P. K., Wu, Y., Gong, W., Mirmehrabi, M., & Rohani, S. (2021). Virtual Multicomponent Crystal Screening: Hydrogen Bonding Revisited. *Crystal Growth & Design*, *21*(10), 5862–5872.

Bennion, J. C., & Matzger, A. J. (2021). Development and Evolution of Energetic Cocrystals. *Accounts of Chemical Research*, *54*(7), 1699–1710.

Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, *559*(7715), 547–555.

Devogelaer, J.-J., Meekes, H., Tinnemans, P., Vlieg, E., & de Gelder, R. (2020). Co-crystal Prediction by Artificial Neural Networks. *Angewandte Chemie International Edition*, *59*(48), 21711–21718.

Duggirala, N. K., Perry, M. L., Almarsson, Ö., & Zaworotko, M. J. (2016). Pharmaceutical cocrystals: along the path to improved medicines. *Chemical Communications*, *52*(4), 640–655.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. a., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. a., Nakatsuji, H., Li, X., Caricato, M., Marenich, a. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., … Fox, D. J. (2016). *Gaussian 16, Revision b.01* (Gaussian 16, Revision b01). Gaussian, Inc., Wallin CT.

Gao, C., Min, X., Fang, M., Tao, T., Zheng, X., Liu, Y., Wu, X., & Huang, Z. (2022). Innovative Materials Science via Machine Learning. *Advanced Functional Materials*, *32*(1), 2108044.

Goh, T., Huang, J.-S., Yager, K. G., Sfeir, M. Y., Nam, C.-Y., Tong, X., Guard, L. M., Melvin, P. R., Antonio, F., Bartolome, B. G., Lee, M. L., Hazari, N., & Taylor, A. D. (2016). Quaternary Organic Solar Cells Enhanced by Cocrystalline Squaraines with Power Conversion Efficiencies >10%. *Advanced Energy Materials*, *6*(21), 1600660.

Groom, C. R., Bruno, I. J., Lightfoot, M. P., & Ward, S. C. (2016). The Cambridge Structural Database. *Acta Crystallographica Section B*, *72*(2), 171–179.

Hao, Y., Hung, Y. C., & Shimoyama, Y. (2022). Investigating Spatial Charge Descriptors for Prediction of Cocrystal Formation Using Machine Learning Algorithms. *Crystal Growth & Design*, *22*(11), 6608–6615.

Huang, Y., Wang, Z., Chen, Z., & Zhang, Q. (2019). Organic Cocrystals: Beyond Electrical Conductivities and Field-Effect Transistors (FETs). *Angewandte Chemie International Edition*, *58*(29), 9696–9711.

Hunter, C. A. (2004). Quantifying Intermolecular Interactions: Guidelines for the Molecular Recognition Toolbox. *Angewandte Chemie International Edition*, *43*(40), 5310–5324.

Jiang, Y., Yang, Z., Guo, J., Li, H., Liu, Y., Guo, Y., Li, M., & Pu, X. (2021). Coupling complementary strategy to flexible graph neural network for quick discovery of coformer in diverse co-crystal materials. *Nature Communications*, *12*(1), 5950.

Juan, Y., Dai, Y., Yang, Y., & Zhang, J. (2021). Accelerating materials discovery using machine learning. *Journal of Materials Science & Technology*, *79*, 178–190.

Kavanagh, O. N., Croker, D. M., Walker, G. M., & Zaworotko, M. J. (2019). Pharmaceutical cocrystals: from serendipity to design to application. *Drug Discovery Today*, *24*(3), 796–804.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2023). PubChem 2023 update. *Nucleic Acids Research*, *51*(D1), D1373–D1380.

Kumar, A., & Nanda, A. (2021). In-silico methods of cocrystal screening: A review on tools for rational design of pharmaceutical cocrystals. *Journal of Drug Delivery Science and Technology*, *63*, 102527.

Li, S., Zheng, L., Chan, Y., Li, B., Sun, Y., Sun, L., Zhen, C., Zhang, X., & Hu, W. (2022). An organic cocrystal based on phthalocyanine with ideal packing mode towards high-performance ambipolar property. *Journal of Materials Chemistry C*, *10*(25), 9596–9601.

Li, Z., Zhou, J., Zhang, K., Zhang, Y., Wu, S., & Gong, J. (2022). Playing with Isostructurality from Binary Cocrystals to Ternary Cocrystal Solvates of Quercetin: Tuning Colors of Pigment. *Crystal Growth & Design*, *22*(9), 5322–5334.

Mswahili, M. E., Lee, M.-J., Martin, G. L., Kim, J., Kim, P., Choi, G. J., & Jeong, Y.-S. (2021). Cocrystal Prediction Using Machine Learning Models and Descriptors. In *Applied Sciences* (Vol. 11, Issue 3).

Musumeci, D., Hunter, C. A., Prohens, R., Scuderi, S., & McCabe, J. F. (2011). Virtual cocrystal screening. *Chemical Science*, *2*(5), 883–890.

Przybyłek, M., Jeliński, T., Słabuszewska, J., Ziółkowska, D., Mroczyńska, K., & Cysewski, P. (2019). Application of Multivariate Adaptive Regression Splines (MARSplines) Methodology for Screening of Dicarboxylic Acid Cocrystal Using 1D and 2D Molecular Descriptors. *Crystal Growth & Design*, *19*(7), 3876–3887.

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 652–660.

Qiao, S., Wang, J., Yu, Y., Liu, Y., Yang, Z., & Li, H. (2022). Two novel TNB energetic cocrystals with low melting point: a potential strategy to construct melt cast explosive carriers. *CrystEngComm*, *24*(16), 2948–2953.

Spellings, M., & Glotzer, S. C. (2018). Machine learning for crystal identification and discovery. *AIChE Journal*, *64*(6), 2198–2206.

Sultan, M., Wu, J., Haq, I. U., Imran, M., Yang, L., Wu, J., Lu, J., & Chen, L. (2022). Recent Progress on Synthesis, Characterization, and Performance of Energetic Cocrystals: A Review. In *Molecules* (Vol. 27, Issue 15).

Ture, S. A., Pattathil, S. D., Patil, V. B., Yelamaggad, C. V, Martínez-Máñez, R., & Abbaraju, V. (2022). Synthesis and fluorescence sensing of energetic materials using benzenesulfonic acid-doped polyaniline. *Journal of Materials Science: Materials in Electronics*, *33*(11), 8551–8565.

Van der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*(11), 2579–2605.

Wang, D., Yang, Z., Zhu, B., Mei, X., & Luo, X. (2020). Machine-Learning-Guided Cocrystal Prediction Based on Large Data Base. *Crystal Growth & Design*, *20*(10), 6610–6621.

Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., & Deng, S.-H. (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb. *Journal of Electronic Science and Technology*, *17*(1), 26–40.

Xiao, F., Cheng, Y., Wang, J.-R., Wang, D., Zhang, Y., Chen, K., Mei, X., & Luo, X. (2022). Cocrystal Prediction of Bexarotene by Graph Convolution Network and Bioavailability Improvement. In *Pharmaceutics* (Vol. 14, Issue 10).

Xiouras, C., Cameli, F., Quilló, G. L., Kavousanakis, M. E., Vlachos, D. G., & Stefanidis, G. D. (2022). Applications of Artificial Intelligence and Machine Learning Algorithms to Crystallization. *Chemical Reviews*, *122*(15), 13006–13042.

Yang, D., Wang, L., Yuan, P., An, Q., Su, B., Yu, M., Chen, T., Hu, K., Zhang, L., Lu, Y., & Du, G. (2022). Cocrystal virtual screening based on the XGBoost machine learning model. *Chinese Chemical Letters*, 107964.

Zhou, S., Wu, F., Tang, G., Wang, Y., & Pang, A. (2021). Effects of 2CL-20/HMX cocrystals on the thermal decomposition behavior and combustion properties of polyether solid propellants. *Energetic Materials Frontiers*, *2*(2), 96–104.

Chapter 6

From Crystal Structures to Macroscopic Morphologies of Rufinamide Polymorphs

6 From Crystal Structures to Macroscopic Morphologies of Rufinamide Polymorphs

Abstract

Crystal morphology plays a pivotal role in the production process of pharmaceutical compounds and significantly influences the properties of the final product. Achieving control over crystal morphology, especially for thin and thread-like crystals, presents significant challenges to the industry. In this study, we investigate the crystal morphology of Rufinamide (RUF) as a model compound with notorious thread-like morphology. Through temperature cycling, solvent screening, and additive selection, we explore different strategies for modulating crystal morphologies of RUF. Among these, temperature cycling shows promising results, while solvent and additive screening yields limited success. Furthermore, we elucidate the crystal structures of RUF polymorphs A, B, and C, and utilize BFDH and experimental face indexing to connect the microscopic crystal structure information to observed macroscopic morphologies. Finally, we propose the use of amide- and acid-based organic molecules as habit-modifying additives to further enhance the crystal morphology of RUF based on the BFDH analysis. The insights gained from this study provide a deeper understanding of crystal morphology control and present opportunities for future research in enhancing the crystal morphology of RUF and addressing similar challenges in other pharmaceutical compounds.

6.1 Introduction

Shape and size are two critical characteristics of crystals that significantly influence various physicochemical properties of an active pharmaceutical ingredient (API) during the production and delivery stages (Coombes et al., 2005; Winn & Doherty, 2000). They have a direct impact on filtration, washing, and drying, which in turn impose limitations on material production rates. In addition, crystal shape plays a role in properties related to storage, handling, flowability, and tabletability. During the delivery stage of pharmaceuticals, morphology becomes crucial as dissolution rates of different crystal faces alter drug release and absorption

(Romero et al., 1991). Furthermore, some physical characteristics of solid particles, including bulk density, wettability, and mechanical strength are also impacted by crystal shape (Winn & Doherty, 2000). Enhancing the crystal morphology of an API, therefore, is of great industrial importance, as evidenced by Upjohn Company's patent on methods to improve the crystal shape of ibuprofen (Gordon & Amin, 1984).

Crystal morphology is a product of the intricate relationship between solid-state characteristics and external variables. Please note in this chapter, the terms crystal morphology, form, habit, and shape are used interchangeably, although they refer to subtly different crystallographic concepts (Wells, 1946). Internal factors that affect morphology include the chemical composition, crystal structure, and polymorphism of the crystal. External factors, such as the nature of the solvent and the presence of additives/impurities, also influence crystal morphology. In order to improve the design process of solid particles, it is imperative to incorporate crystal shape optimization and prediction in the overall design procedure. Engineering crystal shapes is usually achieved through three strategies: 1) screening solvents, 2) introducing growth-inhibiting additives, and 3) manipulating cycles of crystal growth and dissolution through temperature cycling (Dandekar et al., 2013). These approaches enable researchers to control and tailor crystal morphologies as desired. A detailed exploration of the various parameters that influence crystal morphology is provided in this chapter.

6.1.1 Intrinsic Factors Influencing Crystal Morphology

To uncover the influences of internal factors on crystal morphology, an in-depth understanding of crystallography is required, as there exists a fundamental relationship between the unit cell parameters and the macroscopic crystal faces of a crystal. Crystallographic planes are commonly denoted by Miller indices (hkl), which, in combination with the face normals of the unit cell (a^*, b^*, c^*), enable the determination of the normal vector for any crystallographic face ($H = ha^* + kb^* + lc^*$). Consequently, the faces of a macroscopic crystal can be identified and labeled with unique Miller indices, a process known as face indexing (Girolami, 2016). This interplay between crystallography and crystal morphology governs the internal influences on crystal morphology. However, even with the crystal structure readily available, predicting the specific faces that shape a macroscopic crystal, as well as the contribution of

each face to the total surface area (habit), remains a nontrivial task. Below, two models that aim to predict the crystal shape solely based on internal factors are introduced. These methods provide valuable knowledge toward morphology prediction and contribute to the advancement of solid particle design.

In the mid-1800s, Bravais observed that certain crystal faces consistently appear for a given molecule and he attributed this phenomenon to the notion of surface architectures (motifs). He argued that faces with motifs that have high molecular densities are most energetically stable compared to faces with lower molecular densities (Bravais, 1866). Following the advancement of crystallography techniques, Bravais's findings were further refined by Friedel, Donnay, and Harker (Donnay & Harker, 1937; Friedel, 1907), which finally led to the development of the Bravais-Friedel-Donnay-Harker (BFDH) model. The BFDH model prediction is based on the knowledge of unit cell dimensions and the position of molecules. In essence, crystal faces with the highest density demonstrate the largest spacing between adjacent layers, making them the most energetically stable. These faces are the slowest growing faces and hold the highest levels of morphological importance (MI). For more detailed information on the BFDH model, interested readers are encouraged to refer to the article by Prywer (2004).

The second approach for predicting crystal morphology was introduced in a series of three articles (Hartman & Perdok, 1955c, 1955b, 1955a). They proposed that crystal morphology predictions could be improved by considering the number and magnitude of intermolecular interactions in the lattice rather than pure geometric features. In other words, quantifying the intermolecular interactions offers a more precise measure of crystal face stability and growth rate. Therefore, the concept of attachment energy (AE) was introduced as *"the bond energy released when one building unit is attached to the surface of a crystal face"* (Hartman & Perdok, 1955a).

Since their inception, both BFDH and AE models have been implemented in numerous simulation packages and have shown good agreement with experimental results of vapour-grown (sublimation) crystals. Studies on sucrose (Saska & Myerson, 1983), benzamide, benzoic acid, and α-glycine (Berkovitch-Yellin, 1985) are some of the pioneering computer implementations of morphology predictions. More recent studies on crystal morphologies,

such as lovastatin (Turner et al., 2019), venlafaxine hydrochloride (Liang et al., 2022), and β-HMX (Wang et al., 2022) have also discussed the application of these models in predicting vacuum crystal morphologies.

6.1.2 Crystal Morphology, Process Parameters and Growth Environment

6.1.2.1 Temperature Cycling

The crystallization process could be manipulated to control the size and morphology of the crystal. Among various techniques, temperature cycling is a highly effective method in enhancing crystal quality. The concept behind temperature cycling is straightforward: cooling increases the degree of supersaturation, leading to crystallization, while heating promotes dissolution of the crystals. However, it is important to note that the heating-cooling process is not reversible due to the occurrence of Ostwald ripening (Ostwald, 1897). This phenomenon causes smaller crystals to dissolve more readily upon heating, while larger crystals grow more steadily during cooling. More recently, a more in-depth perspective on this phenomenon has been put forward as the "dissolution–ripening–regrowth–relaxation" mechanism (van Westen & Groot, 2018).

Importantly, this effect extends to different crystal faces as well. By implementing successive heating-cooling cycles, the dissolution of longer crystal axes and the recrystallization of smaller axes may lead to improved crystal morphology. In a recent study on cocrystal crystallization in ethanol, the application of temperature cycling resulted in a reduction of the mean aspect ratio of the final product from 10 to 3.3 (Civati et al., 2021). Temperature cycling is commonly implemented in a jacketed batch crystallizer attached to a water bath with a programmed temperature profile (Wu et al., 2016).

6.1.2.2 Solvent Effects

The choice of solvents can significantly influence crystal morphology, and there are two perspectives regarding the effect of solvent molecules on crystal morphology. Firstly, solvent molecules can bind temporarily and selectively to specific crystalline faces, reduce their growth rates, and change the overall shape of the crystal. Secondly, attractive solute-solvent interactions on certain crystal faces decrease the interfacial tension. This reduction in

interfacial tension leads to the transition of the crystal-solvent interface from a smooth surface to a rough one, ultimately, accelerating the surface growth rate (Lahav & Leiserowitz, 2001).

6.1.2.3 The Effect of Additives (Habit Modifiers)

Among a diverse range of foreign molecules that act as habit modifiers, tailor-made additives, and non-size-matched additives are two of the most common subclasses. Tailor-made additives structurally resemble the host molecule but with a modified moiety. This modified moiety is designed to selectively adsorb onto specific crystallographic faces. By occupying these surface sites, the adsorbed additive hinders the regular deposition of subsequent molecular layers of the host molecules, thereby inhibiting the growth of the crystal on that particular face (Graham et al., 2013; Weissbuch et al., 1991). Non-size-matched additives such as water-soluble polymers, celluloses, surfactants, and biologically active macromolecules are also excellent candidates for controlling and altering the crystal shape. These high molecular weight additives are less likely to be incorporated into the lattice of the host molecule due to their size and steric hindrance (Hatcher et al., 2020; Joshi, 2012). Since many polymers and surfactants are commonly used as excipients in final drug formulations, their presence during the crystallization stage raises fewer regulatory concerns (Hatcher et al., 2020).

6.1.2.4 Incorporating External Factors in Morphology Prediction

BFDH and AE models account for internal forces in the absence of an external force. This assumption is valid for sublimation crystallization but for an accurate prediction of crystal morphology obtained from solution crystallization or in the presence of impurities, it is essential to incorporate external factors in the prediction.

One of the pioneering works in incorporating solvent effect in morphology prediction was to use the solvent-accessible surface areas of the molecules on crystal faces as a measure of solvent binding (Myerson & Saska, 1990). Molecular Dynamics (MD) simulation has also been extensively explored (Walker & Roberts, 1993; Wang et al., 2022; Zhou et al., 2020). In MD simulations, first, AE is used to identify the dominant crystal faces. Then, two layers are modeled: the solvent layer, and the crystal face layer which exposes the previously determined

dominant faces from the AE model. After running the MD simulation, the interaction energy at the solvent-crystal interface is calculated and used to modify the AE model predictions.

Regarding the prediction of habit-modifier additives on crystal morphology, large molecular sizes and low concentrations make MD simulations impractical for studying their effects in a statistically meaningful way. The alternative suggestion has been to use the atom-atom potential energy calculations at various sites on different crystal faces to study the nature of host-additive interactions and estimate their contributions to the crystal habit (Berkovitch-Yellin et al., 1985). Such methodologies also rely on the BFDH and AE models to identify morphologically important faces and cannot be done in isolation.

6.1.3 Rufinamide as a Model Compound

Rufinamide (RUF) is a triazole derivative commonly prescribed for the treatment of seizures associated with the Lennox-Gastaut Syndrome (Brodie et al., 2009). RUF is marketed under the brand names Banzel®, Xilep®, and Inovelon® (Asconapé, 2010; Hakimian et al., 2007). Among the four reported polymorphs of RUF, polymorph A is considered the thermodynamically stable form (Martin et al., 2014; Portmann et al., 2004). As of today, the crystal structure of polymorph B of RUF is the only one among its polymorphs that is available in the Cambridge Structural Database (CSD) (Salunke et al., 2018).

RUF crystallizes in extremely thin thread-like (filiform) morphology, which leads to issues such as agglomeration, low apparent density, and poor flow properties (Razzetti et al., 2011). In this chapter, we determine the crystal structures of RUF polymorphs A, B, and C, study their packing in the lattice, perform morphology studies, and explore temperature cycling, as well as solvent and additive screening to transform the filiform (thread-like) habit of polymorph A into a more isometric crystal.

6.2 Methodology

6.2.1 Materials

Apotex PharmaChem Inc. (Brantford, ON) donated RUF (polymorph A) and it was used as the starting compound for this study. Solvents, with analytical reagent purity, were purchased from

Fisher Chemical (Ontario, Canada) and were used as received. Additives (>97% purity) were purchased from Alfa Aesar, Sigma Aldrich, and ACROS Organics and were used as received.

6.2.2 Crystal Structure Determination

To obtain single crystals of RUF polymorphs, cooling, slow evaporation, vapour diffusion, and sublimation setups were employed. Acetone, acetonitrile, chloroform, cyclohexane, cyclopentane, dichloromethane (DCM), dioxane, dimethylformamide (DMF), dimethyl sulfoxide (DMSO) ethanol, hexane, isopropanol, methanol, methyl formate, methanol, tetrahydrofuran (THF), toluene, and water were used in this study. The antisolvent crystallization setup was prepared based on the suggested solvent-antisolvent combinations (Spingler et al., 2012). Sublimation crystallization was done by heating the RUF polymorph A to 250°C in a sealed 5 ml round bottom flask. Cooling crystallization was performed by dissolving 5 mg of RUF in 5 ml of solvents at below boiling point temperatures of each solvent and then slowly cooling to room temperature using a linear cooling profile.

Successful single crystals for RUF polymorph A were obtained by dissolving 4 mg of RUF in 10 ml in pH 2 aqueous solution of HCl at 80°C and slow cooling to room temperature. RUF polymorph B single crystals were prepared by slow evaporation of a saturated solution of methanol. RUF polymorph C was obtained by sublimation crystallization, as well as DCM-cyclopentane and THF-cyclohexane vapour diffusion setups.

Rigaku OD Synergy-HyPix diffractometer equipped with Cu Kα radiation source (λ = 1.54184 Å) and Oxford Instruments Cryojet XL was used for data collection at 160(1) K. Pre-experiment, data collection, data reduction, and analytical absorption correction were performed with CrysAlisPro software (*CrysAlisPro*, 2021). Olex2 1.3 (Dolomanov et al., 2009) was used as a GUI for solving the structure with the SHELXT (Sheldrick, 2015b) program and refining with the SHELXL program package (Sheldrick, 2015a) by full-matrix least-squares minimization. All the non-hydrogen atoms were refined anisotropically while hydrogen atoms were placed in the calculated positions.

Powder X-ray diffraction (PXRD) diffractograms were collected on a Rigaku-MiniFlex II benchtop diffractometer with a Cu Kα source (1.54059 Å) at 30 kV and 15 mA. JADE software was used for the analysis of the PXRD measurements.

6.2.3 Cooling Crystallization Setup

Cooling crystallization was conducted in a 2-liter batch crystallizer. A propeller agitator at 200 RPM was used for mixing. A saturated solution was prepared by heating the RUF suspension using the Julabo FP50-HP refrigerated/heating circulator, followed by a controlled cooling stage to initiate crystallization. Cooling profiles were programmed in the EasyTEMP control software. Focused beam reflectance measurement (FBRM) was used to track the crystal count to zero to achieve a clear solution at the beginning of the experiment. FBRM was also employed to determine the crystallization onset and track the average particle size throughout the experiment.

6.2.4 Morphology Analysis

Structural and packing analyses were performed on the CCDC Mercury program (Macrae et al., 2020). BFDH Morphology module in the CCDC Mercury software package was used to predict the outer shape of polymorphs of RUF and assign the miller indices (hkl) of the most relevant crystalline faces. Experimental face indexing was performed using the CrysAlisPro software, immediately after crystal structure determination.

6.3 Result and Discussion

6.3.1 Temperature Cycling vs. Linear Cooling

Cooling crystallization of RUF from an aqueous solution yields polymorph A which among the known polymorphs of RUF has the most undesired thread-like morphology. We designed an experiment in order to study the morphology enhancement capability of temperature cycling compared to standard linear cooling. A 640-mg sample of RUF was added to 1.6 liter of water in the batch crystallizer. Using a Julabo FP50-HP refrigerated/heating circulator, the temperature was increased to 75°C until all of the RUF crystals were fully dissolved.

For the temperature cycling experiment, each cycle consisted of a 5°C cooling segment followed by a 3°C heating segment. The temperature was cycled from 75°C to 20°C with a heating rate of 1°C.min^{-1} and a cooling rate of 0.5°C.min^{-1}. Also, to assist the breakage of the needles into smaller isometric crystals, wet milling was conducted at 30°C. The experiment lasted for 335 minutes. To compare the results with the standard linear cooling profile, a cooling profile from 75°C to 20°C was programmed for the same duration of 335 minutes. The results are shown in Figure 6-1, which showed a significant improvement in the morphology of RUF form A. PXRD analysis was performed to ensure no polymorph transition occurred by changing the cooling profile.
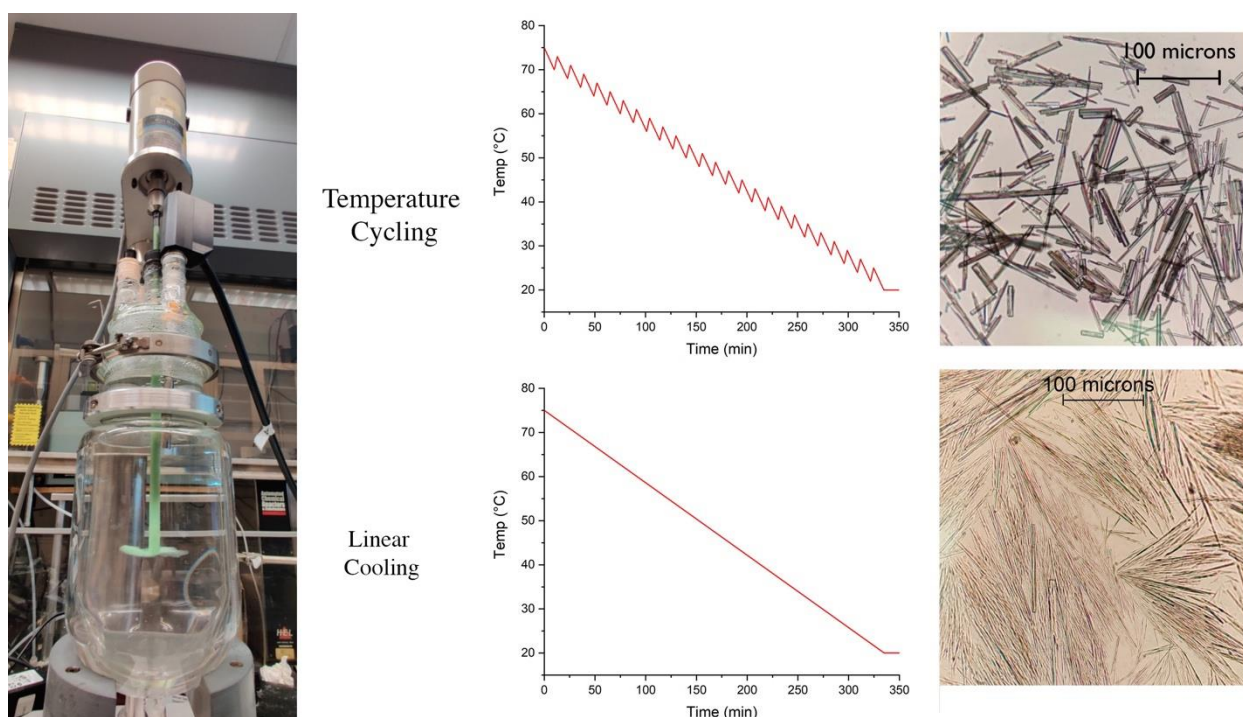
Figure 6-1 Comparison between the morphology of RUF polymorph A crystallization in a batch crystallizer with linear cooling (bottom) and temperature cycling (top).

6.3.2 Solvent Screening and Habit-modifying Additives

To qualitatively get the solubility of RUF in various solvents, 10 mg of RUF was added to 10 ml of the below solvents. For DMSO and DMF, the solubility was significantly high and 10 mg of RUF was readily dissolved in 2 ml of solvent. The solubilities were then ranked as:

DMSO > DMF > Acetone > THF > Methanol > Acetonitrile > Chloroform > Ethanol >
Isopropanol > Toluene > Ethyl acetate > Hexane

The vials were then slightly heated until a clear solution was obtained. The vials were then placed in a refrigerator (4°C) until crystals formed. Using an optical microscope with 5X magnification, the crystal morphologies were examined in search of a solvent that potentially improved the morphology of RUF polymorph A. The findings of this study are summarized in Figure 6-2.
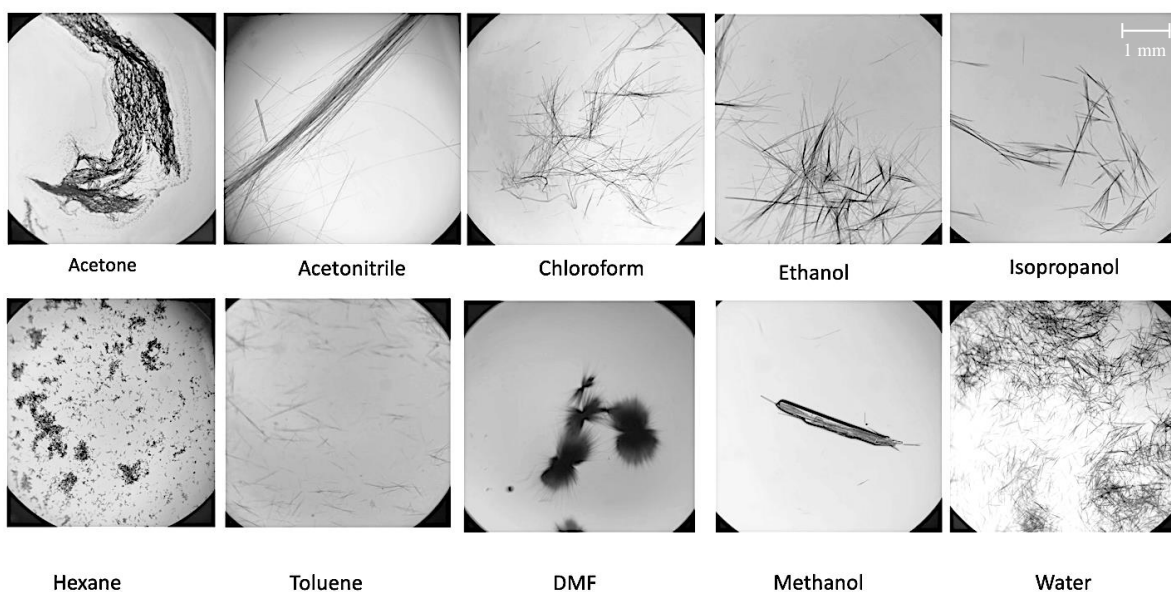
Figure 6-2 The effect of solvents on the morphology of RUF crystals at 5X magnification.

Overall, most solvents had either a negative or neutral impact on the crystal morphology compared to the water, which served as the reference solvent. However, methanol was an exception, as it yielded a large single crystal with significantly improved morphology. To confirm the crystal structure, a scale-up experiment was conducted to generate ~100 mg of RUF crystals from methanol. The resulting crystals were characterized as polymorph B through PXRD analysis. Since the objective was to enhance the morphology of polymorph A, further investigation of methanol crystallization was not pursued.

Next, the influence of habit-modifying additives on the morphology of RUF crystals was investigated. Thirteen non-size-matched additives from three categories of water-soluble

polymers, surfactants, and celluloses were selected for screening purposes. In a 10 ml vial, 4 mg of RUF was added, and dilute aqueous solutions of the additives were prepared and introduced to the vial to attain the desired concentrations. Through some trial and error, we determined that a concentration of 25 ppm was suitable for polymers and surfactants, while a concentration of 4000 ppm was required for cellulose compounds to observe the morphology impact on RUF crystals. The prepared vials were then heated to ensure a clear solution was achieved and then placed in a refrigerator (4°C) until crystals formed. The findings of the screening, observed under 50X magnification, are summarized in Figure 6-3.
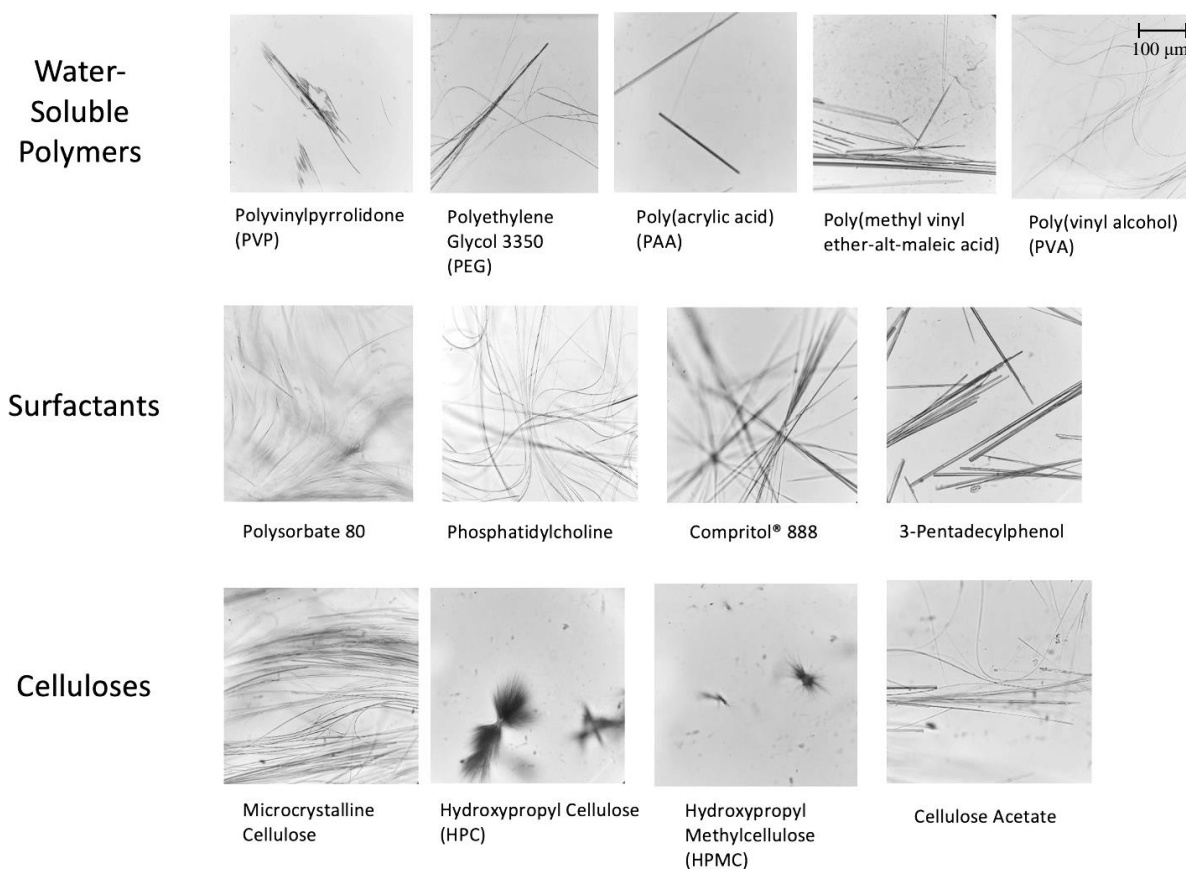
Figure 6-3 The effect of additives on morphology of RUF crystals at 50X magnification.

Similar to the solvent screening, the majority of additives had a negative impact on the morphology. There were two exceptions: PVP promoted a more two-dimensional growth (plate-like crystals) and 3-pentadecylphenol slightly improved the crystal thickness.

To further investigate the effects of these two additives, a larger-scale crystallization was conducted using controlled cooling with stirring. In these experiments, 4 mg of the additive was added to 200 ml of water along with 80 mg of RUF, resulting in a 5% mass ratio of additive to RUF. After a linear cooling crystallization from 75 to 4ºC, no major improvement in morphology was observed. Additionally, the presence of additives had a notable impact on the nucleation temperature and induction time of crystallization. In the case of PVP, nucleation was detected at 10ºC using FBRM, whereas in the standard aqueous solution, nucleation typically occurred around 40ºC. This finding suggests that the crystal yield would be compromised in the presence of PVP. Therefore, further investigation of these two additives was not pursued.

6.3.3 Crystal Structure Analysis of RUF Polymorphs

To adopt a more informed approach in additive selection or the development of tailor-made additives, a comprehensive understanding of the crystallography of RUF polymorphs, their molecular packing, and the identification of Miller indices of the growth faces that need to be hindered is crucial. As established in the introduction, the crystal structure of polymorph A is not available in CSD, which necessitates the determination of its crystal structure. The task of growing single crystals with sufficient thickness in all dimensions is an extremely challenging task for thread-like crystals. However, we managed to successfully grow large-enough crystals of RUF polymorphs A, B, and C, and obtain the crystal structure. The crystal data and structure refinement of these structures are reported in Table 6-1. This detailed information serves as a foundation for designing effective strategies to modify crystal morphology and achieve the desired crystal habit.

Table 6-1 Crystal data and structure refinement for RUF polymorphs A, B, and C.

Identification code	RUF polymorph A	RUF polymorph B	RUF polymorph C
Empirical formula	$C_{10}H_8F_2N_4O$	$C_{10}H_8F_2N_4O$	$C_{10}H_8F_2N_4O$
Formula weight	238.20	238.20	238.20
Temperature	160(1) K	160(1) K	160(1) K
Wavelength	1.54184 Å	1.54184 Å	1.54184 Å
Crystal system	orthorhombic	triclinic	monoclinic
Space group	$Pna2_1$	$P\bar{1}$	$P2_1/c$
	a = 24.6006(14) Å	a = 5.30360(10) Å	a = 11.0136(6) Å
	b = 22.9868(14) Å	b = 11.9470(3) Å	b = 5.3440(3) Å
Unit cell dimensions	c = 5.3672(3) Å	c = 17.1986(4) Å	c = 17.2627(11) Å
	$\alpha = 90°$	$\alpha = 107.324(2)°$	$\alpha = 90°$
	$\beta = 90°$	$\beta = 92.221(2)°$	$\beta = 91.404(5)°$
	$\gamma = 90°$	$\gamma = 102.702(2)°$	$\gamma = 90°$
Volume	3035.1(3) Å^3	1008.47(4) Å^3	1015.72(10) Å^3
Z	12	4	4
Density (calculated)	1.564 Mg/m^3	1.569 Mg/m^3	1.558 Mg/m^3
Absorption coefficient	1.145 mm^{-1}	1.149 mm^{-1}	1.141 mm^{-1}
F(000)	1464	488	488
Crystal size	0.15x0.02x0.01 mm^3	0.16x0.04x0.02 mm^3	0.16x0.03x0.01 mm^3
Theta range	2.631 to 47.645°	2.708° to 80.118°	4.015° to 80.238°
	$-23 \leq h \leq 18$	$-6 \leq h \leq 6$	$-9 \leq h \leq 14$
Index ranges	$-18 \leq k \leq 22$	$-15 \leq k \leq 14$	$-6 \leq k \leq 6$
	$-4 \leq l \leq 5$	$-20 \leq l \leq 21$	$-21 \leq l \leq 21$
Reflections collected	2508	22068	7647
Refinement method	Full-matrix least-squares on F^2	Full-matrix least-squares on F^2	Full-matrix least-squares on F^2
Data / restraints / parameters	2508 / 1 / 462	4359 / 0 / 307	1988 / 0 / 155
Goodness-of-fit on F^2	1.095	1.094	1.060
Final R indices	R1=0.0759,wR2=0.2230	R1=0.0331,wR2=0.0889	R1=0.0550,wR2=0.1701
Largest diff. peak and hole	0.005 and -0.004 e Å^{-3}	0.227 and -0.225 e Å^{-3}	0.320 and -0.278 e Å^{-3}

<div style="text-align:center">(a)   (b)   (c)</div>
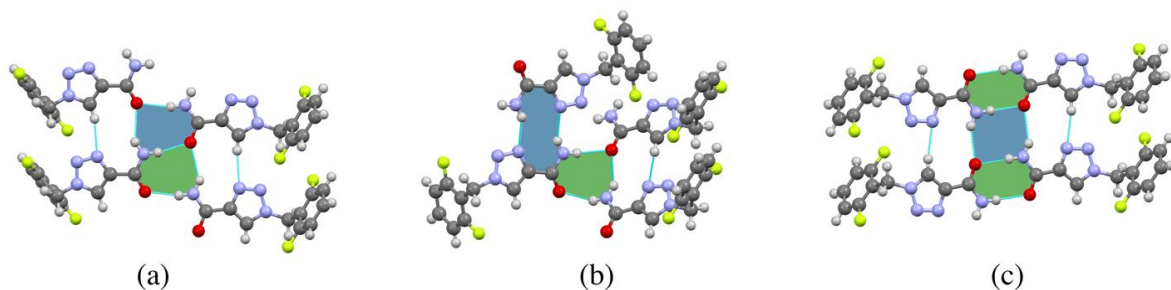
Figure 6-4 Hydrogen bonding patterns in three polymorphs of RUF (A, B, C).

To distinguish between various patterns of intermolecular interactions, graph set approach is a valuable tool. It categorizes the interactions into four patterns of chain (C), ring (R), intramolecular (S), and other finite patterns (D). Each pattern is then further defined by a subscript for the number of hydrogen-bond donors (d), and a superscript indicating the acceptors (a). Additionally, the number of atoms in the pattern is specified in parenthesis. The final graph set descriptor follows the format of $G_d^a(n)$ where G represents one of the four designators (Bernstein et al., 1995).

As shown in Figure 6-4.a, polymorph A features a network of N-H⋯O hydrogen bonds with the graph set descriptor of $R_3^2(8)$. Polymorph B exhibits the same $R_3^2(8)$ pattern along with an additional $R_2^2(10)$ ring of N-H⋯N bonds (Figure 6-4.c). Lastly, polymorph C (Figure 6-4.c) displays a dimeric $R_2^2(8)$ arrangement of N-H⋯O bonds as well as a $R_4^2(8)$ configuration of N-H⋯O bonds.

To determine the crystal growth direction and exposed functional groups, experimental face indexing and BFHD morphology analysis were conducted and are summarized in Figure 6-5 and Table 6-2. There is a good agreement between the predicted and experimental face indexing of polymorph B. However, for polymorphs A and C, some similarities and differences are observed. These discrepancies may be attributed to the small size of the crystals used for measurement, which made it challenging to accurately identify the crystal faces in the camera. In contrast, the selected crystal for the measurement of polymorph B was larger and easier to index. The elongated nature of polymorph A was captured by the BFDH model, but estimated

aspect ratio is not accurate. Nonetheless, the BFDH models provide a valuable starting point for strategically selecting additives in the next stage of this project.
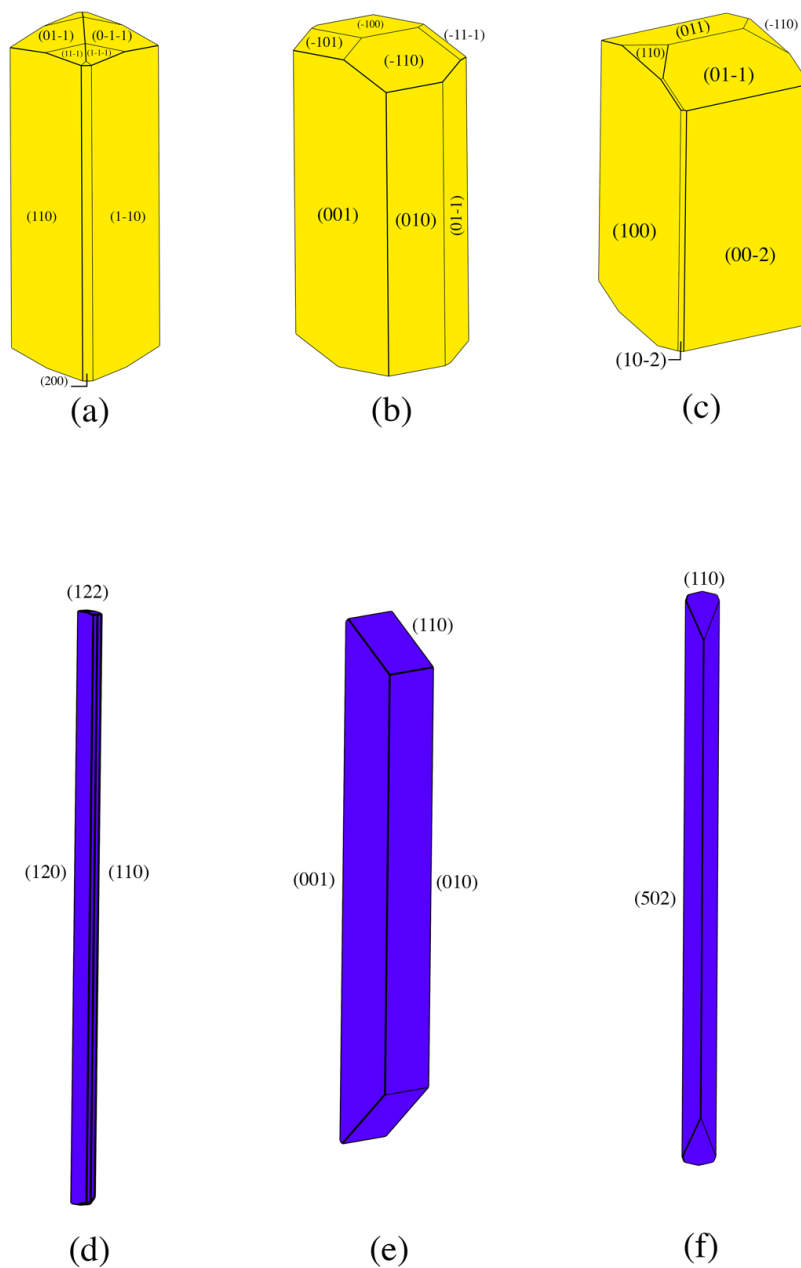
Figure 6-5 BFDH morphologies (a-c) and experimentally indexed faces (d-f) of RUF polymorphs A, B, and C, respectively.

Table 6-2 Detailed information on the morphologically important crystalline faces of RUF polymorphs. Family of plane normals, number of planes in each family (multiplicity), interplanar spacing (d_{hkl}), the perpendicular distance from the face to the center of the crystal, and total facet area are reported.

{hkl}	Multiplicity	d_{hkl}	Perpendicular distance (mm)	Total facet area %
RUF polymorph A				
{**110**}	4	16.80	5.95	81.4
{**200**}	2	12.30	8.13	4.1
{**011**}	2	5.23	19.13	5.2
{**01$\bar{1}$**}	2	5.23	19.13	5.2
RUF polymorph B				
{**001**}	2	16.32	6.13	45.0
{**010**}	2	11.06	9.04	19.9
{**01$\bar{1}$**}	2	10.88	9.19	17.6
{**$\bar{1}$10**}	2	5.17	19.36	7.9
{**100**}	2	5.14	19.45	5.9
RUF polymorph C				
{**100**}	2	11.01	9.08	43.8
{**002**}	2	8.63	11.59	31.5
{**011**}	4	5.10	19.59	20.7

RUF polymorph A is characterized by the presence of two families of planes, namely {011} and {01$\bar{1}$}, which form the dominant top crystalline face. In the case of RUF polymorph B, the largest faces on the elongated axis are {100} and {$\bar{1}$10}. On the other hand, RUF polymorph C is predominantly composed of {011} planes on its top face. To illustrate the functional groups exposed on each face, the BFDH are filled with the RUF molecules, as shown in Figure 6-6.

(a)   (b)   (c)
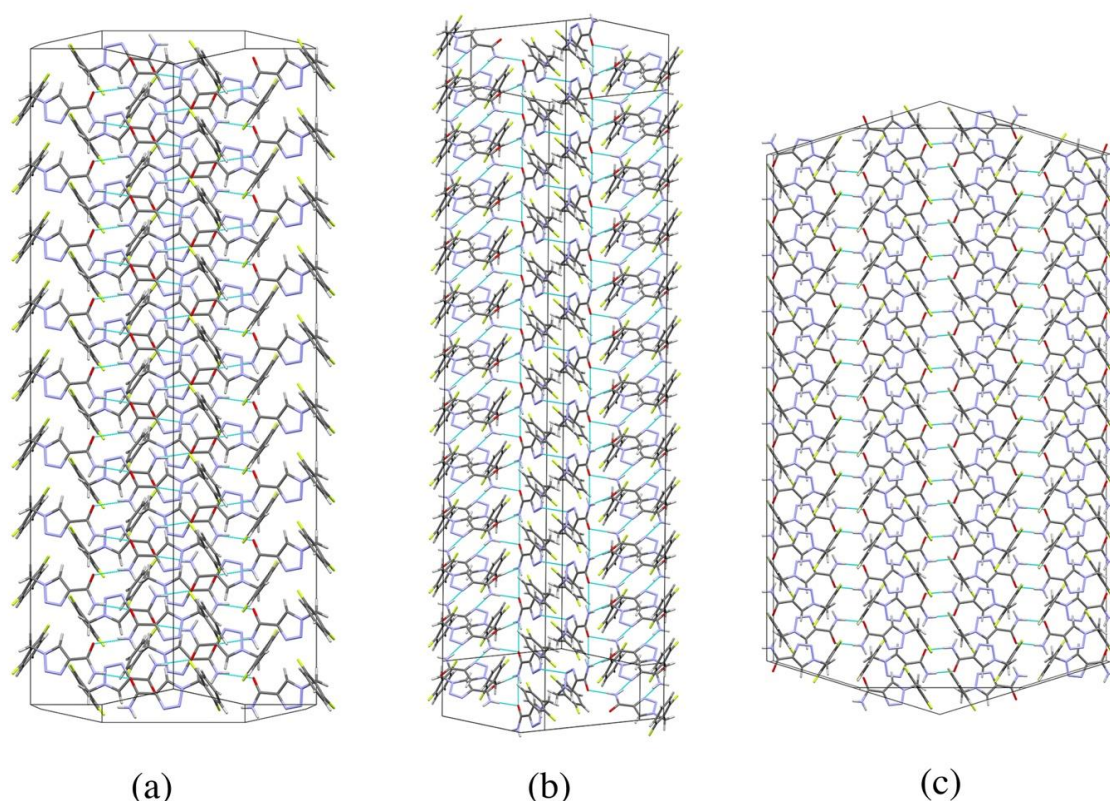
Figure 6-6 BFDH morphologies of RUF polymorphs A, B, and C filled with the RUF molecules.

Considering that polymorph A is the thermodynamically stable polymorph and presents the most challenging morphology, we focus on Figure 6-6.a. One can see that on the side faces, the aromatic ring of RUF is exposed while the atoms of the amide group are exposed on the top faces. Based on this observation, we propose exploring additives that can form temporary hydrogen bond synthons with the amide group, such as the acid⋯amide dimer heterosynthon (Saha & Desiraju, 2018) and amide⋯amide homosynthon (Adalder et al., 2012). Furthermore, in our previous study on RUF, cocrystallization attempts with 52 coformers resulted in the formation of only one new crystalline phase with trimesic acid. Thus, small water-soluble organic molecules containing amide and carboxylic acid groups should be first considered for morphology control, such as acetamide, urea, glutaramic acid as amide candidates; and formic acid, acetic acid, succinic acid, and fumaric acid as potential carboxylic acid-based molecules, for morphology control experiments.

6.4 Conclusion

This study highlighted the importance of crystal morphology control, particularly due to the challenges associated with the production and physical properties of thread-like and needle-like crystals. We explored the intrinsic relationship between the microscopic crystal structure and the macroscopic morphologies, elaborating on the BFDH and AE models used to predict crystal morphologies. Additionally, we elucidated the crystal structure of RUF polymorphs A, B, and C, and reported their respective BFDH morphologies. Among the three strategies employed for crystal morphologies (temperature cycling, solvent screening, and additives), temperature cycling showed promising results, while solvent and additive screening did not yield favorable results. Finally, based on the crystal structure and BFDH analysis, we propose the selection of amide- and acid-based organic molecules as habit-modifying additives for RUF. These findings open new avenues for further research on RUF morphology enhancement and offer valuable insights that can be applied to address similar challenges in other pharmaceutical compounds.

6.5 References

Adalder, T. K., Sankolli, R., & Dastidar, P. (2012). Homo- or Heterosynthon? A Crystallographic Study on a Series of New Cocrystals Derived from Pyrazinecarboxamide and Various Carboxylic Acids Equipped with Additional Hydrogen Bonding Sites. *Crystal Growth & Design*, *12*(5), 2533–2542.

Asconapé, J. J. (2010). The Selection of Antiepileptic Drugs for the Treatment of Epilepsy in Children and Adults. *Neurologic Clinics*, *28*(4), 843–852.

Berkovitch-Yellin, Z. (1985). Toward an ab initio derivation of crystal morphology. *Journal of the American Chemical Society*, *107*(26), 8239–8253.

Berkovitch-Yellin, Z., Van Mil, J., Addadi, L., Idelson, M., Lahav, M., & Leiserowitz, L. (1985). Crystal morphology engineering by "tailor-made" inhibitors; a new probe to fine intermolecular interactions. *Journal of the American Chemical Society*, *107*(11), 3111–3122.

Bernstein, J., Davis, R. E., Shimoni, L., & Chang, N.-L. (1995). Patterns in Hydrogen Bonding: Functionality and Graph Set Analysis in Crystals. *Angewandte Chemie International Edition in English*, *34*(15), 1555–1573.

Bravais, A. (1866). *Études cristallographiques*. Gauthier-Villars, Paris.

Brodie, M. J., Rosenfeld, W. E., Vazquez, B., Sachdeo, R., Perdomo, C., Mann, A., & Arroyo, S. (2009). Rufinamide for the adjunctive treatment of partial seizures in adults and adolescents: A randomized placebo-controlled trial. *Epilepsia*, *50*(8), 1899–1909.

Civati, F., Svoboda, V., Urwin, S. J., McArdle, P., Erxleben, A., Croker, D., & ter Horst, J. (2021). Manipulating Cocrystal Size and Morphology using a Combination of Temperature Cycling and Additives. *Crystal Growth & Design*, *21*(3), 1496–1506.

Coombes, D. S., Catlow, C. R. A., Gale, J. D., Rohl, A. L., & Price, S. L. (2005). Calculation of Attachment Energies and Relative Volume Growth Rates As an Aid to Polymorph Prediction. *Crystal Growth & Design*, *5*(3), 879–885.

CrysAlisPro (1.171.40.122a). (2021). Rigaku Oxford Diffraction.

Dandekar, P., Kuvadia, Z. B., & Doherty, M. F. (2013). Engineering Crystal Morphology. *Annual Review of Materials Research*, *43*(1), 359–386.

Dolomanov, O. V, Bourhis, L. J., Gildea, R. J., Howard, J. A. K., & Puschmann, H. (2009). OLEX2: a complete structure solution, refinement and analysis program. *Journal of Applied Crystallography*, *42*(2), 339–341.

Donnay, J. D. H., & Harker, D. (1937). A new law of crystal morphology extending the law of Bravais. *American Mineralogist*, *22*, 446–467.

Friedel, G. (1907). Études sur la loi de Bravais. *Bulletin de La Société Française de Minéralogie*, *30*(9), 326–455.

Girolami, G. S. (2016). *X-ray Crystallography*. University Science Books.

Gordon, R. E., & Amin, S. I. (1984). *Crystallization of ibuprofen*. United States Patent.

Graham, K. R., Stalder, R., Wieruszewski, P. M., Patel, D. G., Salazar, D. H., & Reynolds, J. R. (2013). Tailor-Made Additives for Morphology Control in Molecular Bulk-Heterojunction Photovoltaics. *ACS Applied Materials & Interfaces*, *5*(1), 63–71.

Hakimian, S., Cheng-Hakimian, A., Anderson, G. D., & Miller, J. W. (2007). Rufinamide: a new anti-epileptic medication. *Expert Opinion on Pharmacotherapy*, *8*(12), 1931–1940.

Hartman, P., & Perdok, W. G. (1955a). On the relations between structure and morphology of crystals. I. *Acta Crystallographica*, *8*(1), 49–52.

Hartman, P., & Perdok, W. G. (1955b). On the relations between structure and morphology of crystals. II. *Acta Crystallographica*, *8*(9), 521–524.

Hartman, P., & Perdok, W. G. (1955c). On the relations between structure and morphology of crystals. III. *Acta Crystallographica*, *8*(9), 525–529.

Hatcher, L. E., Li, W., Payne, P., Benyahia, B., Rielly, C. D., & Wilson, C. C. (2020). Tuning Morphology in Active Pharmaceutical Ingredients: Controlling the Crystal Habit of Lovastatin through Solvent Choice and Non-Size-Matched Polymer Additives. *Crystal Growth & Design*, *20*(9), 5854–5862.

Joshi, S. S. (2012). Crystal habit modification using habit modifiers. In *Modern Aspects of Bulk Crystal and Thin Film Preparation* (pp. 413–436). IntechOpen.

Lahav, M., & Leiserowitz, L. (2001). The effect of solvent on crystal growth and morphology. *Chemical Engineering Science*, *56*(7), 2245–2253.

Liang, C., Zhuang, J., Zhuang, C., Zhang, Z., Lv, G., & Zhang, G. (2022). Crystal morphology prediction and experimental verification of venlafaxine hydrochloride. *Powder Diffraction*, *37*(3), 133–142.

Macrae, C. F., Sovago, I., Cottrell, S. J., Galek, P. T. A., McCabe, P., Pidcock, E., Platings, M., Shields, G. P., Stevens, J. S., Towler, M., & Wood, P. A. (2020). Mercury 4.0: from

visualization to analysis, design and prediction. *Journal of Applied Crystallography*, *53*(1), 226–235.

Martin, A. A. D. L., Bellmunt, J. B., Clotet, J. H., Carandell, L. S., PASCUAL, G. F., Bertran, J. C., & Barjoan, P. D. (2014). *Process for preparing rufinamide intermediate*. Google Patents.

Myerson, A. S., & Saska, M. (1990). Calculation of Crystal Habit and Solvent-Accessible Areas of Sucrose and Adipic Acid Crystals. In A. S. Myerson & K. Toyokura (Eds.), *Crystallization as a Separations Process* (pp. 55–71). ACS, Washington.

Ostwald, W. (1897). Studien über die Bildung und Umwandlung fester Körper. *Zeitschrift Für Physikalische Chemie*, *22U*, 289–330.

Portmann, R., Hofmeier, U. C., Burkhard, A., Scherrer, W., & Szelagiewicz, M. (2004). *Crystal modification of 1-(2, 6-difluorobenzyl)-1H-1, 2, 3-triazole-4-carboxamide and its use as antiepileptic*. Google Patents.

Prywer, J. (2004). Explanation of some peculiarities of crystal morphology deduced from the BFDH law. *Journal of Crystal Growth*, *270*(3–4), 699–710.

Razzetti, G., Vladiskovic, C., & Allegrini, P. (2011). *Crystalline forms of rufinamide*. Google Patents.

Romero, A. J., Lukas, G., & Rhodes, C. T. (1991). Influence of different sources on the processing and biopharmaceutical properties of high-dose ibuprofen formulations. *Pharmaceutica Acta Helvetiae*, *66*(2), 34–43.

Saha, S., & Desiraju, G. R. (2018). Acid···Amide Supramolecular Synthon in Cocrystals: From Spectroscopic Detection to Property Engineering. *Journal of the American Chemical Society*, *140*(20), 6361–6373.

Salunke, N., Thipparaboina, R., Chavan, R. B., Lodagekar, A., Mittapalli, S., Nangia, A., & Shastri, N. R. (2018). Rufinamide: Crystal structure elucidation and solid state characterization. *Journal of Pharmaceutical and Biomedical Analysis*, *149*, 185–192.

Saska, M., & Myerson, A. S. (1983). The theoretical shape of sucrose crystals from energy calculations. *Journal of Crystal Growth*, *61*(3), 546–555.

Sheldrick, G. M. (2015a). Crystal structure refinement with SHELXL. *Acta Crystallographica Section C*, *71*(1), 3–8.

Sheldrick, G. M. (2015b). SHELXT – Integrated space-group and crystal-structure determination. *Acta Crystallographica Section A Foundations and Advances*, *71*(1), 3–8.

Spingler, B., Schnidrig, S., Todorova, T., & Wild, F. (2012). Some thoughts about the single crystal growth of small molecules. *CrystEngComm*, *14*(3), 751–757.

Turner, T. D., Hatcher, L. E., Wilson, C. C., & Roberts, K. J. (2019). Habit Modification of the Active Pharmaceutical Ingredient Lovastatin Through a Predictive Solvent Selection Approach. *Journal of Pharmaceutical Sciences*, *108*(5), 1779–1787.

van Westen, T., & Groot, R. D. (2018). Effect of Temperature Cycling on Ostwald Ripening. *Crystal Growth & Design*, *18*(9), 4952–4962.

Walker, E., & Roberts, K. J. (1993). Molecular Simulations Application Notes, CGR 10. *Biosym/Molecular Simulations Inc*.

Wang, K., Huang, S., & Zhu, W. (2022). Solvent effects on the crystal morphology of β-HMX studied by modified attachment energy and occupancy models. *Journal of Crystal Growth*, *585*, 126605.

Weissbuch, I., Addadi, L., Lahav, M., & Leiserowitz, L. (1991). Molecular Recognition at Crystal Interfaces. *Science*, *253*(5020), 637–645.

Wells, A. F. (1946). XXI. Crystal habit and internal structure -I. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *37*(266), 184–199.

Winn, D., & Doherty, M. F. (2000). Modeling crystal shapes of organic materials grown from solution. *AIChE Journal*, *46*(7), 1348–1367.

Wu, Z., Yang, S., & Wu, W. (2016). Application of temperature cycling for crystal quality control during crystallization. *CrystEngComm*, *18*(13), 2222–2238.

Zhou, T., Chen, F., Li, J., He, L., Ren, Y., Wang, X., Cao, D., & Wang, J. (2020). Morphology prediction of 5,5′-bistetrazole-1,1′-diolate (BTO) crystal in solvents with different models using molecular dynamics simulation. *Journal of Crystal Growth*, *548*, 125843.

Chapter 7

Conclusion and Recommendation

7   Conclusion and Recommendation

The final chapter of this thesis summarizes the achievements and offers four additional suggestions for future research endeavors.

7.1 Summary and Research Findings

In this thesis, we investigated the multifaceted field of pharmaceutical multicomponent crystals with a particular emphasis on cocrystals. The aim was to enhance the understanding of cocrystal formation, explore their physicochemical properties, and develop strategies for cocrystal design. Through a comprehensive exploration, topics such as high-throughput virtual screening, synthesis, characterization, and potential applications of cocrystals were explored. Additionally, the control of crystal morphology as an additional topic of interest is studied. The major conclusions from this research are as follows:

i. Conducting a relatively small-scale brute-force experimental screening of combinations of the API with 10-50 coformers may or may not result in the discovery of new cocrystals. The success rate of this screening approach greatly depends on the nature of the API. For example, the notable difference in success rates between Olanzapine (50%) and Rufinamide (2%) cocrystallization attempts serves as evidence that the challenge of discovering new cocrystals is highly dependent on the specific API being studied.

ii. There are many reasons to attempt to supplement the brute-force experimental screening with a virtual screening technique. Having limited quantities of API available in the early stages of development, encountering failure in initial experimental screenings, and the desire to gain a competitive edge and secure all potential patents for an API, are some of the arguments why there is a critical need for a guiding tool in the comprehensive search for cocrystals. Such a predictive tool can be obtained through knowledge-based approaches, such as the hydrogen bond propensity (HBP) available in the CSD Mercury program, or theoretical-based approaches, such as molecular electrostatic potential (ESP) maps obtained through DFT calculations.

iii. In this thesis, we have emphasized that ESP maps, in particular, provide a feature-rich representation of molecules, making them a promising approach for cocrystal predictions. The simplest way to utilize ESP maps for cocrystal prediction is by extracting hydrogen bond donor and acceptor parameters (α, β) and subsequently calculating pairing energies. Further improvement in the predictive power of ESP maps for cocrystal prediction can be achieved by applying machine learning models. Either through the deep learning model of PointNet, which takes the whole ESP map as an input and automatically extracts its features, or through manual feature extractions (α, β, A, V, Ψ) combined with artificial neural network (ANN) or random forest (RF) machine learning models, the predictive performance of ESP maps are significantly improved. In addition, a combination of these three models (PointNet, ANN, RF) within an Ensemble learning model yields the highest balanced accuracy (94%) for unseen cocrystals in the test dataset.

iv. In a case study aimed at overcoming the hydrophobicity of an API (Zn-PA), we successfully cocrystallized the Zn-PA with isonicotinamide (INAM), which resulted in a significant reduction in hydrophobicity, as evidenced by a reduction in the droplet contact angle from 128.1° to 27.1°. To understand the underlying reasons for this substantial change, we extensively characterized the crystal using various techniques and determined its crystal structure. The crystal structure analysis revealed a significant modification in the intermolecular interactions between Zn-PA and INAM compared to Zn-PA alone. The dispersion-based pi-stacking interaction in Zn-PA was replaced by the coulomb-polarization effect of hydrogen bonds in Zn-PA-INAM. Furthermore, morphology analysis demonstrated that unlike Zn-PA, Zn-PA-INAM exhibited exposed polar groups on its prominent crystalline faces, contributing to the reduction in hydrophobicity.

v. In another case study, we encountered a transient solvate of Dasatinib with methanol (DAS-MeOH). The desolvation of this crystal occurs at room temperature after it is removed from the mother liquor. Hot-stage microscopy revealed that the crystal remains stable only at temperatures lower than -27°C outside of the mother liquor. To understand this unusual behavior, we conducted various analyses, including obtaining the crystal structure and examining the position of solvent molecules within the lattice. Due to the

rapid desolvation process, the crystals would fracture and transform into microcrystals of an anhydrous polymorph of DAS, distinct from the original material. Through particle size distribution analysis, we proposed that this desolvation process could be utilized for particle size reduction.

vi. Some molecules, such as Rufinamide, tend to crystallize in an extremely fine and thread-like morphology. This leads to poor crystal quality and various challenges in the production and application of these molecules. In this case study, we discovered that temperature cycling effectively reduces the aspect ratio of these thread-like crystals. Additives and solvents also have a significant impact on the morphology, but their specific effects can only be determined through obtaining the crystal structure and performing face indexing. The BFDH model, in particular, is a simple and informative tool for studying crystal morphology and systematically selecting additives that can positively influence the crystal shape.
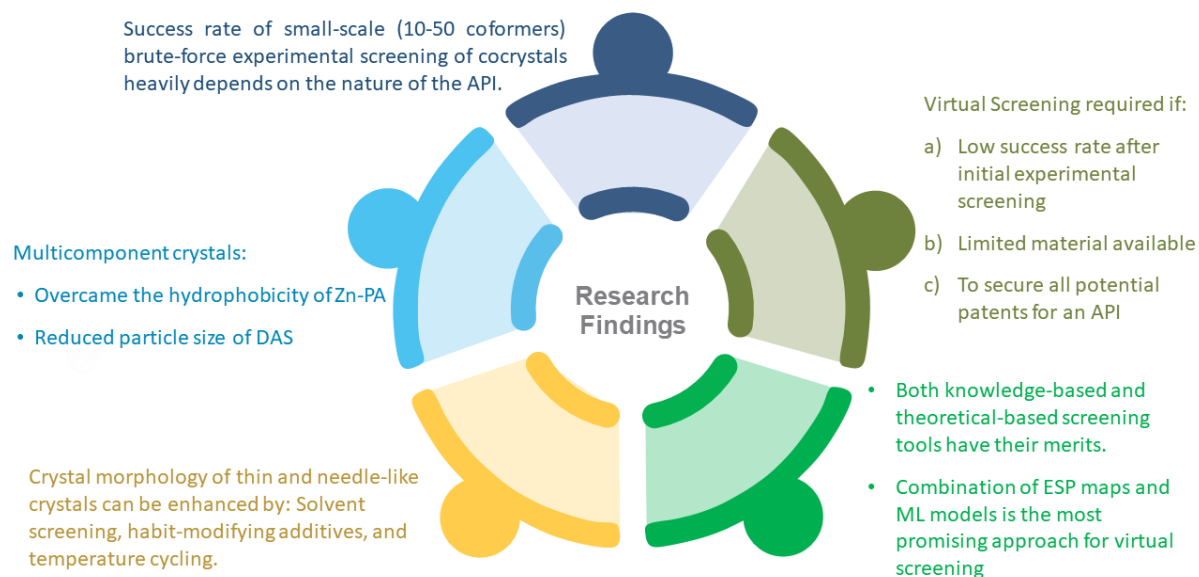
Figure 7-1 Summary of key research findings.

In summary, this thesis contributes to the field of pharmaceutical cocrystals. Through extensive research and experimentation, we advanced our understanding of cocrystal formation, explored their physicochemical properties, and established strategies for cocrystal design. We emphasized on utilizing ESP maps as a comprehensive representation of molecular features for cocrystal prediction, which has been further enhanced through the incorporation of machine learning models. Furthermore, the successful case studies focusing on mitigating hydrophobicity and utilizing transient solvates for particle size reduction exemplified the practical applications of multicomponent crystals in enhancing drug properties. Overall, this thesis sets the stage for future research and development in this pivotal realm of drug innovation.

7.2 Recommendation for Future Work

   i.   Cocrystal Production Scale-up

After discovering a new cocrystal through screening methods such as liquid-assisted grinding, the challenge of scaling up its production arises. Cooling crystallization is arguably the first choice for large-scale crystallization. Since the API and coformer exhibit different solubilities in the mother liquor, the development of a ternary phase diagram is necessary for each cocrystal production. Initially, slurries with varying API-to-coformer ratios are prepared. After reaching equilibrium, the solid phase is characterized by PXRD, and liquid concentrations are determined by HPLC (Chiarella et al., 2007). Once the ternary phase diagram is established, strategies for large-scale crystallization can be proposed. These studies can be conducted on cocrystals such as Zn-PA-INAM (Chapter 3), Rufinamide-Trimesic acid, and several Olanzapine cocrystals (Chapter 4), providing valuable insights into their production at an industrial scale.

   ii.   *In-silico* Preparation of Ternary Phase Diagram

The experimental route to obtain an accurate ternary phase diagram, as described in suggestion (i), is time-consuming and acts as a major stumbling block in solution cocrystallization studies. With thermodynamic understanding of cocrystals (Schartman, 2009), models can be developed to predict solubilities and solid outcome of crystallization. Statistical models and machine

learning techniques can be explored to facilitate an efficient *in-silico* preparation of the ternary phase diagram. This approach offers the potential for a more streamlined scale-up process of cocrystal manufacturing.

iii.   Incorporating Supramolecular Features in Cocrystal Prediction

Predicting whether a pair of API-coformer will form a cocrystal is a key focus of this thesis (Chapters 4 and 5). While we have examined molecular features, such as ESP maps, we acknowledge that when molecules combine to form a crystal, unique supramolecular features emerge. For example, lattice energy is a property exclusive to crystals. The challenge arises when attempting to extract these supramolecular features during the screening stage, as crystal structures of cocrystals are yet to be obtained at this stage. Although crystal structure prediction methods (Reilly et al., 2016) can be used to predict the crystal structure of an unknown cocrystal, they tend to be more time-consuming compared to experimental screening. One potential solution is the development of a machine learning algorithm that can predict specific supramolecular features of unknown crystal structures using available data from reported crystal/cocrystal structures in databases like CSD. The algorithm objective would be to predict, for instance, the lattice energy of a molecule pair (a regression problem), which in turn could serve as a measure for cocrystal prediction (a classification problem).

iv.   Cocrystal property prediction

Improving the physicochemical properties of APIs is the key objective for cocrystal synthesis. However, not all cocrystals yield favorable changes in properties, as they can alter API's properties in unpredictable ways. In a study of over 700 cocrystals, it was revealed that there is no clear trend in comparing cocrystal, API, and coformer melting points with 55.3% of the cocrystals showing melting points within the range of the individual components, 15.8% having higher melting points and 28.9% exhibiting lower melting points (Perlovich, 2015). Thus, predictive tools capable of estimating solubility, melting point, hydration stability, hygroscopicity, tabletability, permeability, and other relevant physicochemical properties is of great significance. These tools would provide valuable insights for future cocrystal design and synthesis.

7.3 References

Chiarella, R. A., Davey, R. J., & Peterson, M. L. (2007). Making Co-CrystalsThe Utility of Ternary Phase Diagrams. *Crystal Growth & Design*, *7*(7), 1223–1226.

Perlovich, G. L. (2015). Thermodynamic characteristics of cocrystal formation and melting points for rational design of pharmaceutical two-component systems. *CrystEngComm*, *17*(37), 7019–7028.

Reilly, A. M., Cooper, R. I., Adjiman, C. S., Bhattacharya, S., Boese, A. D., Brandenburg, J. G., Bygrave, P. J., Bylsma, R., Campbell, J. E., Car, R., Case, D. H., Chadha, R., Cole, J. C., Cosburn, K., Cuppen, H. M., Curtis, F., Day, G. M., DiStasio Jr, R. A., Dzyabchenko, A., … Groom, C. R. (2016). Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials*, *72*(4), 439–459.

Schartman, R. R. (2009). On the thermodynamics of cocrystal formation. *International Journal of Pharmaceutics*, *365*(1), 77–80.

Appendix 1

Molecular Dynamics Simulation of Homogeneous Nucleation of Supersaturated Potassium Chloride (KCl) in Aqueous Solutions

A1 Molecular Dynamics Simulation of Homogeneous Nucleation of Supersaturated Potassium Chloride (KCl) in Aqueous Solutions

Abstract

Molecular dynamics (MD) simulation is used to investigate the mechanism of crystal nucleation of potassium chloride (KCl) in a supersaturated aqueous solution at 293 K and 1 atm. Using radial distribution function (RDF), bond distance calculation, angle measurement, and configurational snapshots it was found that the newly emerged phase is face-centered cubic crystals with some water molecules trapped inside the crystal lattice. It was also shown that during early stages of nucleation, high local density of ions occurred, and within these areas, the nucleation started with a sequence of ionic additions, as suggested by classical nucleation theory. It was concluded that crystal nuclei form in a sequential manner, but this can only happen in places where the local density of ions is higher than solution concentration and the probability of having effective collisions increases, making these sites the primary candidates for nucleation.

A1.1 Introduction

In solution crystallization, solute supersaturation is the primary driving force of crystallization, but it is not sufficient cause for a system to begin to crystallize. Before formation of any crystal, a number of minute solid bodies (nuclei) or seed must exist as centers of crystallization. These early stages of crystallization, known as nucleation, play a decisive role in determining the crystal structure and crystal size distribution (CSD) (Mullin, 1997). Having a clearer view of nucleation process will result in the determination of optimum operating conditions to achieve high selectivity over the organization of particles within crystal lattice and consequently obtain uniform crystals reproducibly (Erdemir et al., 2009).

The main difficulty in studying nucleation is the small length- and time-scale of initial clusters; their size ranges from 100 to 1000 particles (Erdemir et al., 2009) which is difficult to detect experimentally, and they also move freely in the solution over short periods of time which is

very difficult to track. Thanks to the advanced experimental techniques, such as cryogenic transmission electron microscopy (cryo-TEM), which led to the Noble Prize in Chemistry in 2017, the early stages of crystallization have been observed in real time (Habraken et al., 2013). However, microscopic level insight of nucleation is still challenging to obtain, and the knowledge of crystal nucleation is still far from comprehensive (Karthika et al., 2016).

Nucleation can be classified into two major categories, as shown in Figure A1-1: primary and secondary. Primary nucleation refers to the systems that do not contain crystalline matter initially. On the other hand, nuclei are often generated in the vicinity of crystals present in a supersaturated system; this mechanism of nucleation is known as secondary nucleation. In the absence of crystals, primary nucleation can either be homogeneous or in the presence of foreign particles providing surface area for nucleation (heterogeneous nucleation) (Mullin, 1997). Mechanism of pure homogeneous nucleation and the formation of a stable crystal nucleus is not known with any degree of certainty; crystal nuclei should resist the tendency to redissolve, and they also have to become oriented into a fixed lattice. The stable nuclei are formed by very high local supersaturation or coagulation, but the mechanism is still to be determined since it is counterintuitive to assume a stable nucleus can be formed from the simultaneous collision of many particles, ordering themselves in a crystalline structure (Mullin, 1997).
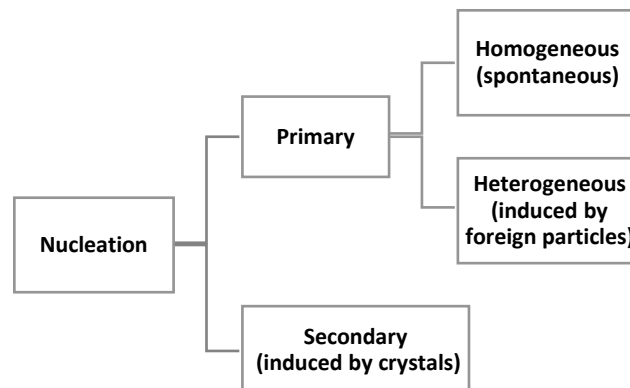
Figure A1-1 Classification of crystal nucleation processes.
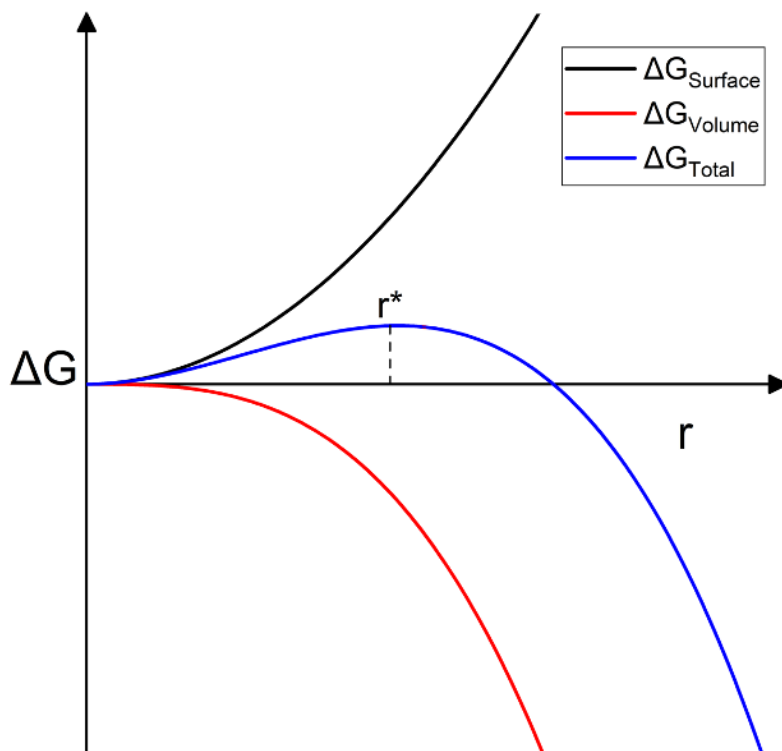
Figure A1-2 Schematic representation of Gibbs free energy barrier versus nuclei radius (r) according to classical nucleation theory.

There are several nucleation theories developed to describe this process, divided into two subcategories: classical nucleation theory (CNT) and non-classical nucleation pathways.

The CNT, most common theoretical model developed to understand nucleation, stems from the work of Volmer, Becker, and Frenkel in 1920s to 1940s (Becker & Döring, 1935; Frenkel, 1939; Volmer & Weber, 1926). CNT argues that it is more likely that a nucleus arises from a sequence of bimolecular additions until the Gibbs free energy of the system start to decay, and the crystallization process becomes favorable (Karthika et al., 2016). The free energy changes may be considered as the summation of two parts, as shown in Figure A1-2, interfacial free energy which is unfavorable and results in an increase in total Gibbs free energy and the volumetric part which is favorable and leads to negative Gibbs energy:

$$\Delta G_{Total} = \Delta G_{Surface} + \Delta G_{Volume} = 4\pi r^2 \sigma - \frac{4\pi r^3}{3v} \, KT \ln S$$

Where σ is specific surface energy of the interface between nucleus and surrounding, the $\frac{4\pi r^3}{3v}$ represents the number of molecules in a cluster of radius r with the volume of a single molecule as v and S is supersaturation defined by C/C*, where C is the solute concentration and C* is the solute saturated concentration.

The growth of clusters depends on the competition between ΔG_v, which favors growth, and ΔG_s, which favors dissolution. The blue line in Figure A1-2 shows the sum of these two competing ΔGs. For small nucleus sizes the surface energy dominates but at some point, shown as r* (critical radius), the favorable volumetric part surpasses the interfacial energy, and the total Gibbs energy starts to decline, making the nucleus stable.

Classical nucleation theory usually gives reasonable predictions of critical nucleus size and nucleation rates, and it is also capable of capturing the underlying physics of the phenomena. However, it is incapable of providing accurate quantitative predictions of key parameters such as free energy of formation of the cluster, its size, and rate (Karthika et al., 2016). These shortcomings encouraged researchers to develop non-classical nucleation theories. Advanced experimental techniques have demonstrated the emergence of intermediate stages before formation of the most thermodynamically favored phase. Oswald's rule of stages (Ostwald W, 1897) argues that an unstable supersaturated system does not transform directly to the most stable state, but to one which resembles its own energy and structure. This "multi-stage" nucleation, although widely upheld, has no theoretical foundation and is not universally accepted.
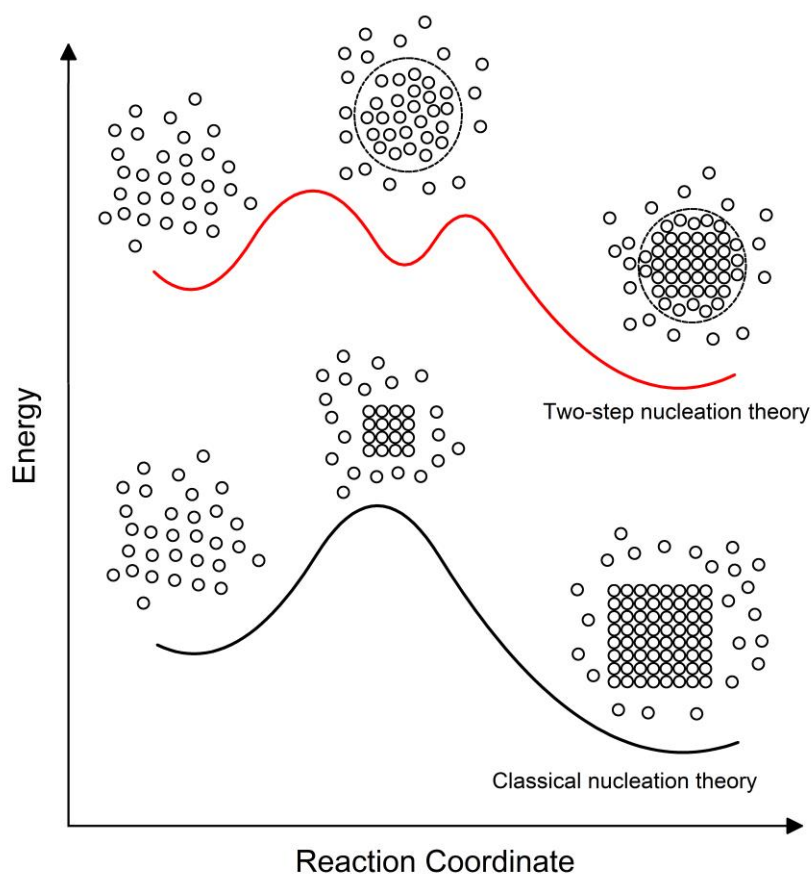
Figure A1-3 Schematic Comparison between classical nucleation and two-step nucleation theories. In the classical view of crystal nucleation, particles form an ordered arrangement from the beginning (bottom) while in two-step nucleation theory they first form a dense amorphous precursor then rearrange into crystalline form (top).

In a two-step nucleation theory, as shown by the red line in Figure A1-3, the first energy barrier corresponds to local density fluctuation of the solute molecules, forming a non-crystalline cluster which needs to overcome the second energy barrier to rearrange into a stable ordered crystalline nucleus (Karthika et al., 2016). The short lifetime of metastable intermediate phases was considered as the explanation for the lack of experimental evidence for multi-step nucleation.

Computer simulations can be considered as a powerful complementary technique to experimental methods, where concentration fluctuations in the solution lead to the formation

of amorphous prenucleation clusters with diffuse boundaries. Brute-force Simulations, such as Monte Carlo (MC), usually coupled with enhanced sampling techniques, can estimate the free energy barrier (Talanquer & Oxtoby, 1998). One of the first researches supporting the two-step mechanism was reported by Wolde and Frenkel, who studied homogeneous nucleation in a Lennard-Jones system of short-range attraction by Monte Carlo (Rein ten Wolde & Frenkel, 1999). The presence of an intermediate phase in the form of a high concentration liquid was shown to be a generic feature for substances that interact through sufficiently short-range interactions.

Molecular dynamics (MD) is another simulation method that can be used for nucleation studies with the advantage of having the temporal evolution of the process (Talanquer & Oxtoby, 1998). The most popular ensemble used for crystallization simulation is NPT, known as the isothermal-isobaric ensemble in which the Number of particles, Pressure, and Temperature are kept constant (NPT).

Sodium chloride (NaCl), in particular, is a well-studied crystallization system using MD simulation due to its relatively easy setup and availability of experimental nucleation rates. In the early 90s, Ohtaki and Fukushima employed molecular dynamics simulation to study the nucleation of NaCl and CsF crystals from aqueous solution on a very small system, consisting of only 448 particles, over 10 picoseconds (Ohtaki & Fukushima, 1991). In 2004, Zahn employed the path sampling approach to investigate nucleation of NaCl from aqueous solution, showing that the stable centers of nucleation consist of octahedrally coordinated ions (Zahn, 2004). Later, in 2008, another MD study of NaCl system for a range of temperatures of supercritical water, performed by Nahtigal et al., showed the dependence of NaCl crystal size distribution (CSD) on the density of the system, with lower densities providing larger crystals (Nahtigal et al., 2008). Moreover, they reported the size of the initial clusters to be between 14-24 ions with an amorphous structure. As computing power improved over time, larger systems for a more extended period of time were studied. Chakraborty and Patey in 2013 published a large- scale simulation of NaCl, consisting of 64,000 particles, and reported the evidence of two-step nucleation (Chakraborty & Patey, 2013). They employed SPC/E parameters as the water model, and for ions they used OPLS force field.

Metadynamics method has also been utilized for NaCl simulation; Giberti et al. used GROMOS force field and SPC/E water model to perform the metadynamics simulation (Giberti et al., 2013). Their finding was that during early stages a different polymorph of NaCl, namely a wurtzite-like structure formed rather than face- centered cubic of rock salt. A more in-depth study by Zimmermann et al. illustrated that the GROMOS force filed tends to overestimate the stability of wurtzite-like structure for NaCl (Zimmermann et al., 2015). In this study, they used the common SPC/E water model, but for ion parameters, values suggested by Joung and Cheatham were selected which resulted in reliable solubilities and chemical potential driving force. Additionally, the free energy barrier for nucleation as a function of supersaturation was also evaluated, which showed a substantial discrepancy with experimental results. In a recent study in Zimmermann's group in 2018, they re-examined seeded simulations of NaCl, providing nucleation rates and the possible sources of uncertainty in nucleation rate estimations (Zimmermann et al., 2018). Beside from NaCl studies, in a recent publication in 2019, Peng et al. investigated a supersaturated system of KCl, with 6 M concentration, using OPLS ion models and TIP4P/2005 water model (Peng et al., 2019). Two simulation boxes with sizes of 3 and 7 nm were considered, showing the strong size decency of nucleation time since nucleation was observed at 50 ns for 3 nm simulation while it occurred in less than 10 ns for the 7 nm simulation. They reported the formation of ionic clusters prior to nucleation as well as a spike in the local viscosity at the time of nucleation.

In addition to computational simulations, theoretical studies, such as density functional theory (DFT), have also provided evidence for the two-step nucleation mechanism (Talanquer & Oxtoby, 1998).

Despite all significant advances, both in experimental methods and simulation techniques, many open questions are still to be answered. Overall, the study of nucleation in more depth, besides scientific importance, would affect the industry as well; having a clearer view of nucleation will eventually result in the determination of optimum operating conditions to achieve high selectivity over the organization of particles within clusters and consequently obtain uniform crystals reproducibly.

In this study, we utilized SPC/E water model and OPLS force field parameters for KCl ions to study the nucleation of KCl from aqueous solutions with a large-scale MD simulation by GROMACS. We found out that the initial stable nuclei are in fact ordered octahedral structure of 26-ions formed from a sequence of ionic additions. A more in-depth investigation of local solution properties showed that the nucleation occurred in locations where the ionic supersaturation was higher than the solution concentration, leading to more frequent and effective ionic collisions.

A1.2 Methods

A1.2.1   Simulation Runs

All simulations were performed using GROMACS 2019.1(Abraham et al., 2015; Berendsen et al., 1995; Hess et al., 2008; Lindahl et al., 2001; Páll et al., 2015; Pronk et al., 2013; Van Der Spoel et al., 2005). A leap-frog algorithm was used for integrating Newtonian equations of motion with a 2 femtosecond (fs) time step. Periodic boundary conditions (PBC) were used to avoid boundary effects and make the simulation box an infinite-like system. For the long-range electrostatic interactions, the particle mesh Ewald (PME) approach (Darden et al., 1993) was utilized, and the LINCS algorithm (Hess, 2008) was used to constrain the hydrogen bonds. Simulations were carried out in the isothermal-isobaric ensemble (NPT), by the aid of velocity rescaling (v-rescale) thermostat (Bussi et al., 2009) to keep the temperature constant at 293 K and Parrinello-Rahman Barostat which kept the pressure constant at 1 atm (Parrinello & Rahman, 1981). The cutoff for non-bonded interactions (van der Waals and electrostatic) was set to 1 nm.

The experimental solubility (mass fraction) of KCl in water is reported to be 0.2648 at 298.15 K (Pinho & Macedo, 2005), which is equivalent to molarity of 3.4 M. It is important to note that solubility limit for KCl at 298.15 K in MD simulation may vary considerably from experimental measurements, based on the force field parameters for ions as well as the water model. To the best of authors' knowledge, the solubility of KCl for the OPLS force field and SPC/E water model is not reported in the literature.

Determination of solubility is a formidable task in MD simulations and can be a topic of separate research. However, in order to ensure that our target production simulation is neither undersaturated nor unreasonably supersaturated to reach spinodal decomposition, a series of MD simulations on a smaller system (9 nm cubic box with approximately 24,000 particles) was performed at different concentrations, the number of nuclei observed after each simulation observed, and shown in Table A1-1:

Table A1-1 List of simulations performed to select the appropriate concentration for production simulation.

Run	Simulation box edge (nm)	Number of water molecules	Number of ions	Concentration (M)	Number of nucleation sites	Simulation time (ns)
1	9	21,695	2,196	2.5	0	400 ns
2	9	21,257	2,634	3	1	250 ns
3	9	20,379	3,512	4	3	100 ns
4	9	19,501	4,390	5	5	50 ns
5	9	18,623	5,268	6	8	50 ns
6	9	17,745	6,146	7	Indefinite*	20 ns

*The spinodal decomposition was observed in this case; a particular case where the concentration of solute is too high making the supersaturated solution unstable compared to the crystalline phase; and crystallization proceeds without any energy barrier, leading to an indefinite number of crystals.

This set of simulations showed that the concentration of 2.5 M for KCl with OPLS force field and SPC/E water model was low enough that no nucleation was observed during 400 ns of NPT simulation, implying that the system is either in its equilibrium state or overcoming the energy barrier for nucleation needs more extended simulation which is unreachable with classical MD simulations. On the other hand, higher concentrations, such as 7 M, resulted in the spinodal decomposition at a very short simulation time, which is not appropriate for nucleation studies. Therefore, it can be concluded that the concentration used in this work (4.0 M) lies in the practical range of concentrations for MD simulation, without reaching the extreme condition of spinodal decomposition.

The analyses given in this article are based on the simulation on a relatively large cubic box with the edge of 13 nm, containing approximately 72,000 particles (61,163 water molecules

and 5,292 ion pairs) representing the concentration of 4 M for KCl, generated by genion command.

Initially, a potential energy minimization run was carried out using the steepest descent algorithm to relax the system, eliminate any steric conflicts between molecules, and obtain the initial configuration for the molecular dynamics simulation. The system was then equilibrated in a short NPT run, where the velocities were generated from a Maxwell distribution, and the desired temperature and pressure were achieved with the aid of proper thermostat and barostat. This equilibration simulation was then followed by a production run of 200 ns.

Visualization and analysis were carried out using GROMACS routines, Python scripting, and VMD (Humphrey et al., 1996; Stone, 1998) software.

A1.2.2   Force Fields

The all-atom optimized potentials for liquid simulations (OPLS-AA) (Åqvist, 1990; Chandrasekhar et al., 1984; Jorgensen et al., 1996) force field was used for water and KCl ions. The extended single point charge (SPC/E) (Berendsen et al., 1987) water model was used as the solvent. The list of parameters of this force field for the atoms/ions used in this work are listed in Table A1-2:

Table A1-2 Lennard-Jones Parameters, partial charges and masses for the model considered.

Atom Ion	Type	σ (nm)	ε $\left(\dfrac{kJ}{mol}\right)$	Atomic Mass $\left(\dfrac{g}{mol}\right)$	Partial charge (e)
H	opls_117	0.0	0.0	1.00800	+0.4238
O	opls_116	0.316557	0.650194	15.99940	-0.8476
K$^+$	opls_408	0.493463	0.00137235	39.09830	+1
Cl$^-$	opls_401	0.441724	0.492833	35.45300	-1

A1.2.3   Crystal Detection

As an attempt to detect the early stages of crystal formation, we need to find some criteria to distinguish liquid-like from solid- like clusters. This is generally achieved by measures on local parameters. Some of the most common local measurements include ion connectivity, ionic density, solvent density, and bond order parameters (Lanaro & Patey, 2016). In this work, local ion connectivity and a bond order parameter have been used to systematize the detection of crystalline ions and ions in the solution form.

A1.2.3.1   Ion Connectivity Method for Crystal Detection

In any ionic solution, regardless of the concentration, oppositely charged ions tend to connect and form a bond due to electrostatic attraction forces. However, in many cases, this connectivity is not strong enough and, therefore, is broken by the kinetic energy of neighboring molecules and electrostatic forces of partial charges on the solvent molecules. Each ion in the KCl solution can have 0 to 6 oppositely charged ions in its first solvation shell. The probability that an ion has the maximum allowable number of ions (6) is low enough that can be considered as the smallest meaningful cluster of ions that may lead to nucleation. The equilibrium position of the neighbors is determined by the Van der Waals radii of ions. However, it is expected that, initially, the neighboring ions are not fixed in their equilibrium distance. Therefore, in this study, a relatively small fluctuation is allowed, and 0.40 nm was chosen as the cut-off. As depicted in Figure A1-4.a, this cut-off is large enough to account for small vibrations, and at the same time, it is small enough to prevent the counting the next layer of ions due to the hindrance originated from Van der Waals radius of the 6 neighbors (Figure A1-4.b).
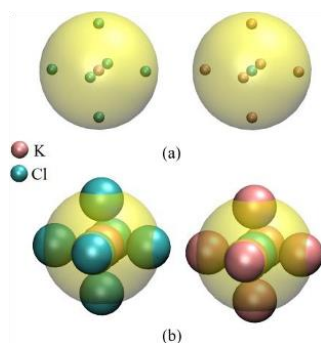
Figure A1-4 The maximum number, six, of oppositely charged ions surrounding a central ion, within 0.40 nm (yellow sphere) visualized by (a) point charges and, (b) Van der Waals spheres.

To perform the analysis of ion connectivity, we generated a Python script to calculate the distance of oppositely charged ions in the system within the cut-off of 0.40 nm; In case that an ion was surrounded by exactly 6 neighbors, it was labeled as a center and counted for each frame of the trajectory file with time step of 100 ps.

A1.2.3.2   Bond Order Parameter Method for Crystal Detection

The bond order parameter is a robust technique for crystal identification. In 2016, Lanaro and Patey used an order parameter to detect and follow solid-like NaCl clusters as potential nuclei (Lanaro & Patey, 2016). The utilized bond orientational order parameter was proposed by Steinhardt et al. (Steinhardt et al., 1983), based on the spherical harmonics (Ym). Spherical harmonics are in general function of angles of spherical coordinates (θ and ϕ) of the bond with respect to a reference point. It is worth noting that for even-l spherical harmonics, we do not need to associate a direction since they are invariant under inversion. In numerical studies for order parameter, the average quantities of spherical harmonics over a number of immediate neighbors of a central ion is considered:

$$Q_{lm} = \langle Q_{lm}(r) \rangle = \frac{1}{N} \sum_{r_1}^{r_N} Y_l^m \left(\theta(r_i), \phi(r_i) \right)$$

Where N is the number of neighbors, Y_l^m is the spherical harmonics which is a function of $\theta(r_i)$, polar angle, and $\phi(r_i)$, azimuthal angle. For any spherical harmonic, the degree of the harmonic (l) and the order of the harmonic (m), must be specified. The only constraints are that the degree of the harmonic must be a positive integer, and the absolute value of the order of the harmonic cannot be larger than its degree ($-l \leq m \leq l$).

For the purpose of order parameter studies, the function is summed over all possible m values ($\sum_{m=-l}^{l} |Q_{lm}|^2$), and normalized, making it a single variable parameter (Q_l):

$$Q_l = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |Q_{lm}|^2}$$

l is an independent variable that must be altered and then tested over known systems to verify its applicability to differentiate disordered ions in the solution form the ordered crystalline ions.

In order to make this order parameter more robust, a careful number of neighbors (N) must be selected. In the original paper of (Steinhardt et al., 1983) , they argued that radial distribution functions (RDFs) in dense liquid indicate that each atom is surrounded by approximately 12 particles in the first coordination shell. Therefore, the $\boldsymbol{Q_{lm}}$ was averaged over (N=12) in the above equations.
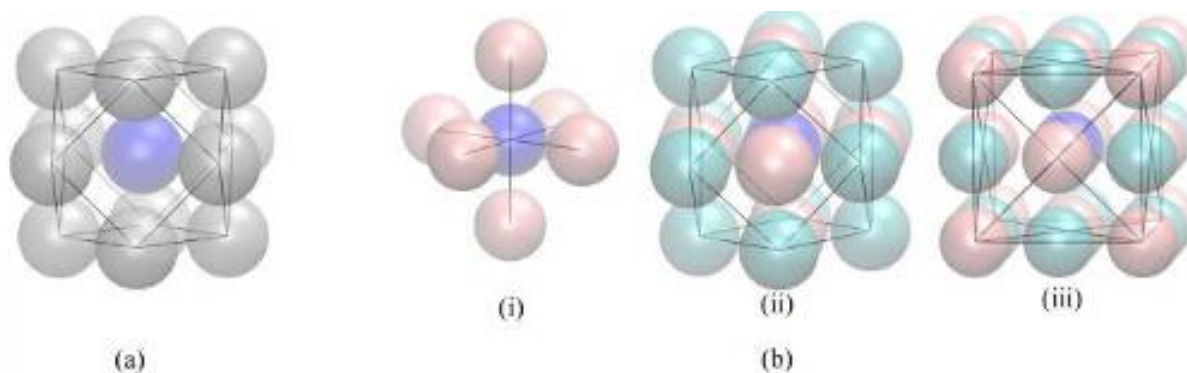
Figure A1-5 Possible number of neighbors for Q_{lm} averaging. (a) covalent FCC crystal (coordination number is 12), (b) ionic FCC crystals (coordination number is 6): (i) first coordination number, (ii) cumulative second coordination number, (iii) cumulative third coordination number.

However, we noticed a pitfall during the analysis. As can be seen in Figure A1-5.a, averaging over 12 neighbors is valid for an FCC covalent crystal. However, for ionic FCC crystals, where crystals consist of 2 types of particles, we should consider both ion types, by choosing the number of neighbors between 6, 18, and 26 neighbors, shown in Figure A1-5.b (i)-(iii), respectively.

A set of calculations for even values of l between 2 to 20, and averaging over 6, 18, and 26 neighbors showed that the optimum separation between crystal and solution could be obtained when l is 12 (Q_{12}), and the number of neighbors around a central ion is 18.
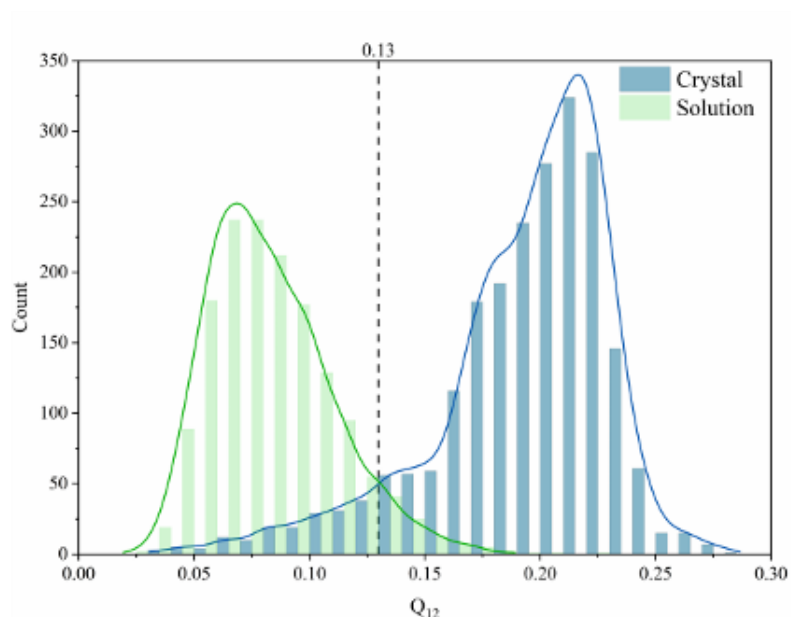
Figure A1-6 count of Q_{12} values in a sample crystal of over 2000 ions (blue) and a supersaturated solution with a KCl concentration of 2.5 molars consisting of about 2000 ions (green).

As illustrated in Figure A1-6, selection of 0.13 for Q_{12} enables us to distinguish the crystals from solution since 91.4% of the ions in the crystal have Q_{12} of more than 0.13, whereas 93.5% of ions in the solution have Q_{12} of less than 0.13.

Before finalizing the selected order parameter, namely Q_{12}, it is essential to ensure that it is independent of the solution concentration since solution concentration changes during crystallization. It is not expected to have any meaningful order in the arrangement of ions in the solution form, either saturated or supersaturated, so any proposed order parameter must acknowledge this fact. To confirm that Q_{12} distribution is invariant with solution concentration, 5 different concentrations of KCl, from 1.0 to 5.0 molar, were selected and their Q_{12} were calculated and compared in Figure A1-7. Because different concentrations consist of different numbers of ions, the probability density is used as the y-axis to achieve better comparison. The probability density normalizes the count, in a manner that the area under the histogram sums to 1.
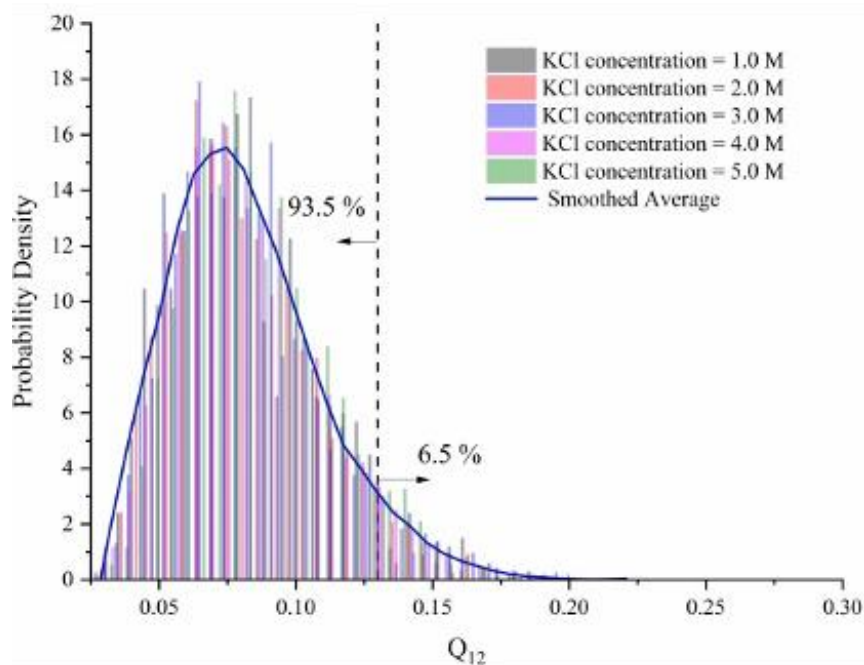
Figure A1-7 Comparison of the value of Q_{12} for different concentrations of KCl solutions.

Figure A1-7 illustrates that the chosen order parameter is independent of solution concentration with a complete overlap between 5 concentrations. As shown by the smoothed average (blue curve), 93.5 % of ions have Q_{12} of more than 0.13 and only 6.5% of the data lies in the over 0.13 portion of the graph.

In order to apply the proposed order parameter on the performed simulation in GROMACS, a Python script is written that takes the coordinates of ions at each frame and picks the closest 18 neighbors of each ion to perform the average order parameter calculation. After filtration of the 18 nearest neighbors, their Cartesian coordinates are converted to spherical coordinates. Next, the spherical harmonics function from SciPy library, a Python-based collection of open-source software, is used to calculate the spherical harmonics (Y_l^m), and finally obtaining the Q_{12}.

A1.3 Results

A1.3.1   Uncorrelated Measurements

Before drawing any conclusion of MD simulations, it is crucial to make sure the measured states are uncorrelated. Velocity autocorrelation function (VACF) is a criterion, which measures the dependence of the average velocity of all particles in the system on measurements in previous time steps. To have statistically independent sampling, the time of taking measurements must be longer than the time required for the VACF to decay to zero. This time is usually very short for small molecules. Using GROMACS velacc tool, it was found that for this system the VACF decays to zero in less than 1 ps. In all the analyses reported in this work, data is sampled every 100 ps that ensures the uncorrelated measurements.

A1.3.2   Changes in the Energy of the System

Crystallization of KCl is an exothermic process that experimentally releases 13.8 kJ energy per mole (Vacek & König, 1983), resulting in a temperature rise in the system. For quantitative modeling of this energy, the force field parameters should be modified accordingly. Since the OPLS force field has not been parametrized to capture the true energy of crystallization process, only qualitative interpretations are valid. When we use the v-rescale thermostat in MD simulation, the temperature is kept constant. By keeping the temperature constant in the NPT ensemble, the total energy of the system is no longer conserved. Therefore, the heat of crystallization affects the total energy of the system by decreasing the potential energy, while kinetic energy remains constant due to the constant temperature.

Figure A1-8 illustrates the changes in the total energy of the system during 200 ns of simulation. The black curve, which is sampled every 2 ps, is smoothened by using moving average method over 1 ns interval, shown by the red line. During the first 27 ns of simulation, the decrease in energy is very small, but it experiences a steep decline as crystal nuclei start to form. The steep descent continues until 100 ns and then slows down. This downward trend in the energy plot almost levels off towards the end of the simulation, indicating that no more ions are joining the crystal nuclei, the solution is no longer supersaturated, and the simulation has almost reached its equilibration state.
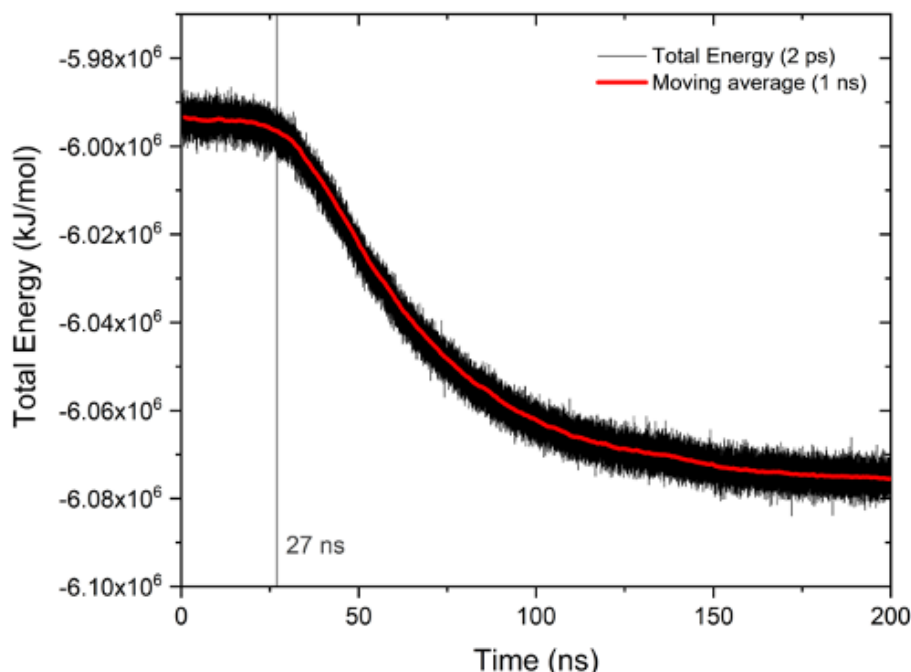
Figure A1-8 Change in the total energy of the simulation box during the crystal nuclei formation process.

A1.3.3 Radial Distribution Function (RDF)

The radial distribution function, g(r), or pair correlation function is a measure to demonstrate how density of selected group varies with respect to the distance from a set of reference particles, normalized to the average particle density of the selected group in the system with maximum radial distance, which in GROMACS is half of the box (6.5 nm in this case).

A1.3.3.1 Time Evolution of RDF

The time evolution of RDF for Cl ions is shown in Figure A1-9 for 3 nm distance from the reference group (K ions).
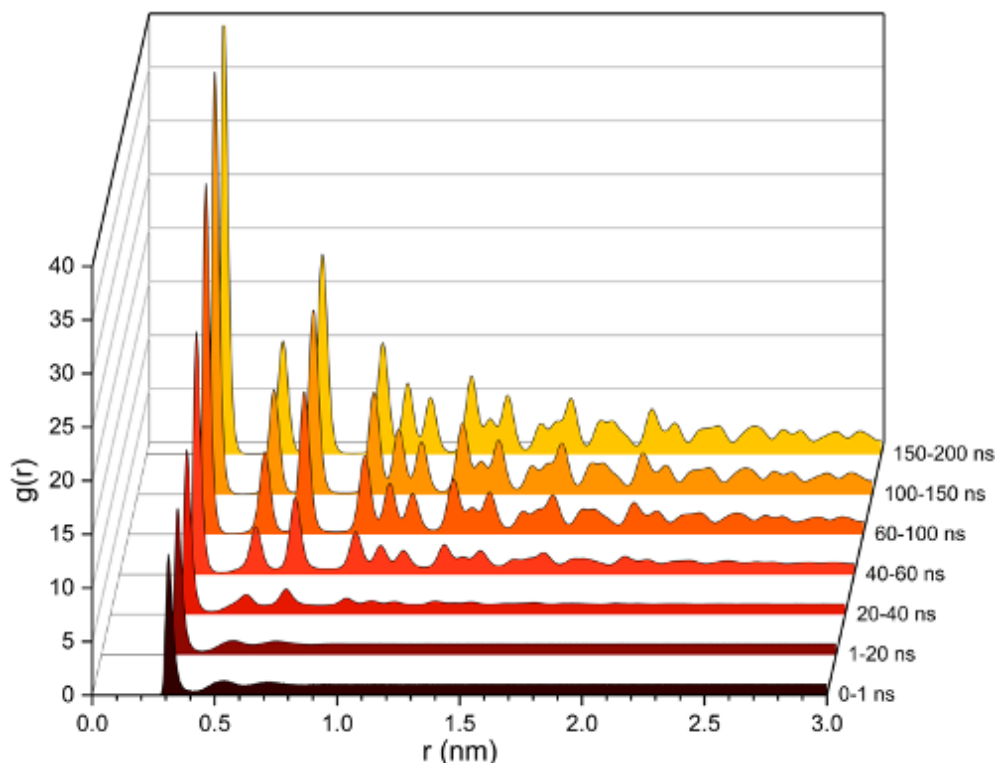
Figure A1-9 Radial distribution function (RDF) of K and Cl ions over time.

As one can see in Figure A1-9 , the initial configuration, black curve, exhibits only one noticeable peak. This peak is an indication of strong attraction between positive (K) and negative (Cl) ions, which results in the formation of diatomic chains of KCl even in the solution. These chains are not stable and easily break when hit by other particles in the system. The rest of the RDF, especially after 1 nm, is a plateau at value of 1 due to the normalization, which signifies the ions are totally uncorrelated at long distances. During the second time range, 1 to 20 ns, the RDF remains almost unchanged but in the third plot, 20 to 40 ns, the magnetite of the first peak as well as the extent in radius where peaks are observed increase. This is an indication of crystal formation since crystals are more ordered structures than the liquid solution and the amplitude of peaks in RDF must rise, and new peaks must emerge at longer distances. For the rest of the simulation, up to 200 ns, this trend continues, and the time evolution of these solid-like ordered regions can be seen.

A1.3.3.2 Crystal Structure Analysis Using RDF

The locations of peaks of an RDF correspond to the highest probable position of finding another ion. It should be a delta function if the material is a static crystal, but in a molecular dynamics simulation, particles are not stationary, and consequently, the RDF broaden, but the tip of each peak, connoting the most probable distance, can be considered as the equilibrium distance.

Alike ions (Cl-Cl and K-K) RDFs, Figure A1-10, in have complete overlap which is only observed for a crystalline structure; in solution, these peaks have very small magnitude and correspond to the momentarily formation of short chains, such as tri-ionic chains (Cl-K-Cl or K-Cl-K), and they peak at different locations based on the order of ions. Conversely, in a large 3D crystal, the peaks of like ions merge due to packing.
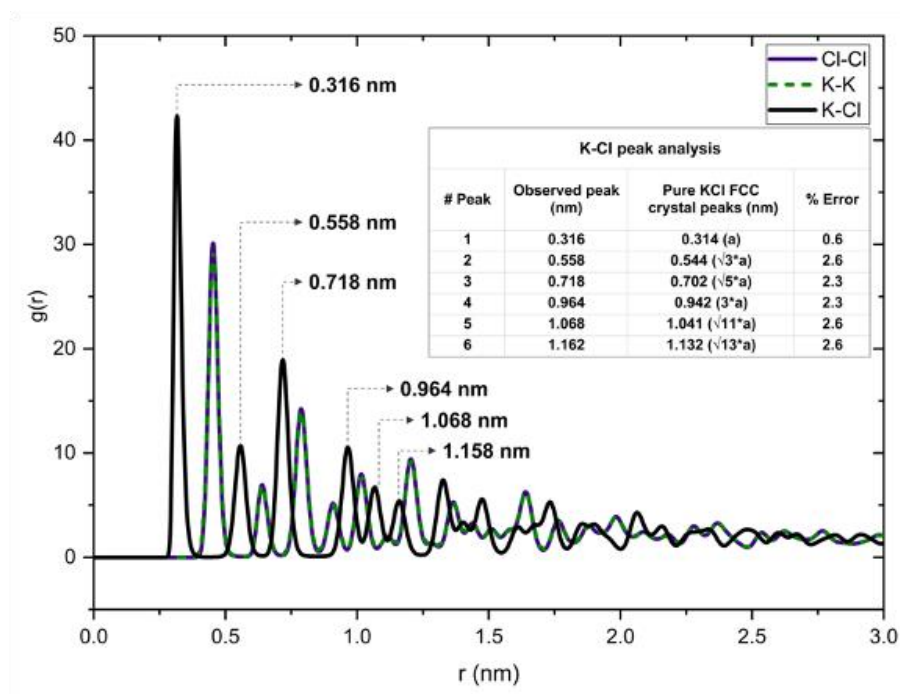
Figure A1-10 Radial distribution function (RDF) of Cl-Cl, K-K, K-Cl at the final configuration of the system for structural analysis.

The first 6 peaks of K-Cl RDF, shown as the black line in Figure A1-10, are labeled and compared with expectations from geometric analysis of a perfectly crystalline face-centered

cubic (FCC) of KCl with the bond distance of 0.314 nm (Vogt & Weiss, 2001). The simulation results are in good agreement with the theoretical values, with the average error of only 2.2 %. It can be concluded that the new phase formed during the simulation is FCC crystals of KCl.

A1.3.3.3   Water Inclusion Analysis Using RDF

In a pre-nucleation cluster, ions are surrounded with a large number of water molecules and as nuclei form the number of water molecules decreases but eventually some trapped water molecules remain in the lattice, and the crystal will not be anhydrous. Chakraborty and Patey also reported the same behavior in MD simulation of NaCl crystals (Chakraborty & Patey, 2013).

Pair distribution function, g(r), between water and ions, including both K and Cl, can be used to show the water inclusion in the lattice. The RDFs, calculated for three periods, initial configuration (0-5 ns), nucleation time (27-32 ns) and final configuration (195-200 ns), are shown in Figure A1-11. In the initial configuration of the supersaturated solution, the g(r) reaches to 1 after 1 nm of distance from the reference point, indicating the presence of water molecules in the vicinity of ions same as the solution concentration. During initial stages of nucleation (red line) a small decrease in RDF is observed, but for final configuration, it experiences a significant drop, especially for short ranges, the inner crystal regions. Although RDF decreases, it does not reach zero. This is an indication of existence of water molecules inside the lattice, as visualized in the side image of Figure A1-11. Note that, in the side image, ions are made transparent to visualize inner water molecules, causing the misinterpretation that the shown water molecules are outside of the crystal.
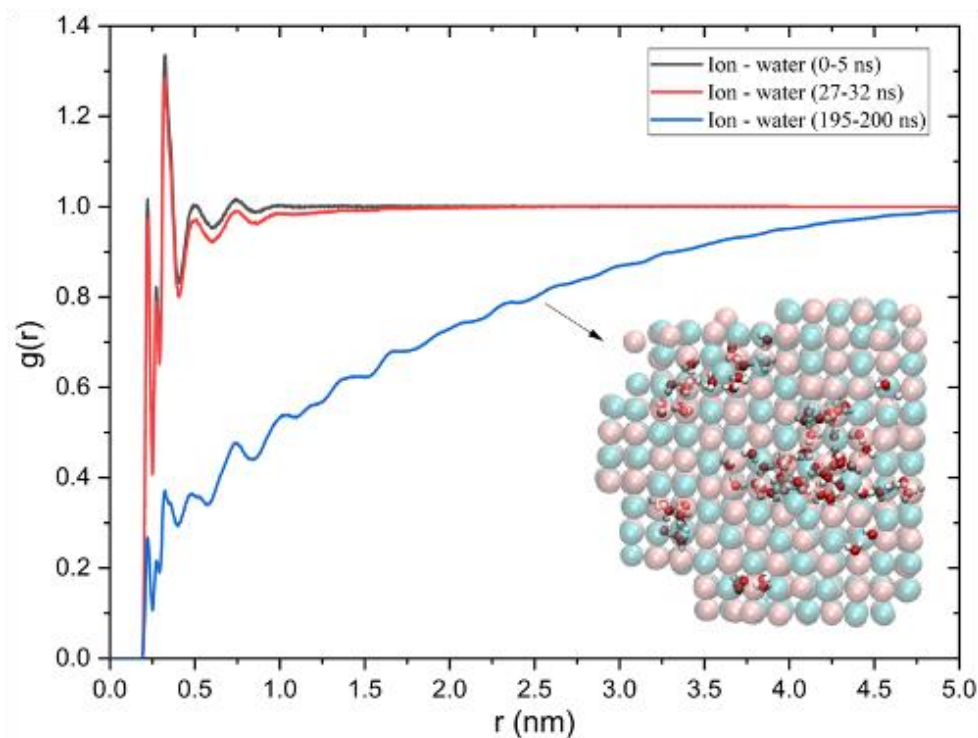
Figure A1-11 Water – ion radial distribution function.

In addition, a more accurate method for identifying the water molecules inside the lattice is developed, using water molecules' neighborhood analysis.

The analysis was performed by using a Python script to count the number of ions around each water molecule within a selected cut-off, followed by distinguishing and labeling the inner lattice water molecules. The inner lattice water molecules identification is based on the study of various scenarios for water molecules surrounded by ions within 5.5 Å of distance. As shown in Figure A1-12.b, water molecules on the surface of the crystal have encompassed 9 ions from the crystal surface within the cut-off. Using 9 ions as the criterion for identifying trapped water is not accurate since, at some points, some inner corners and holes were observed on the surface of the crystal due to the different growth rates of various faces. These locations are still considered as the surface of crystal, and water molecules inside these locations have up to 15 ions within the cut-off (Figure A1-12.c). Therefore, the filtration was done by 16 ionic neighborhoods since a water molecule trapped in the lattice is surrounded by at least 16 ions, as shown in Figure A1-12.d.
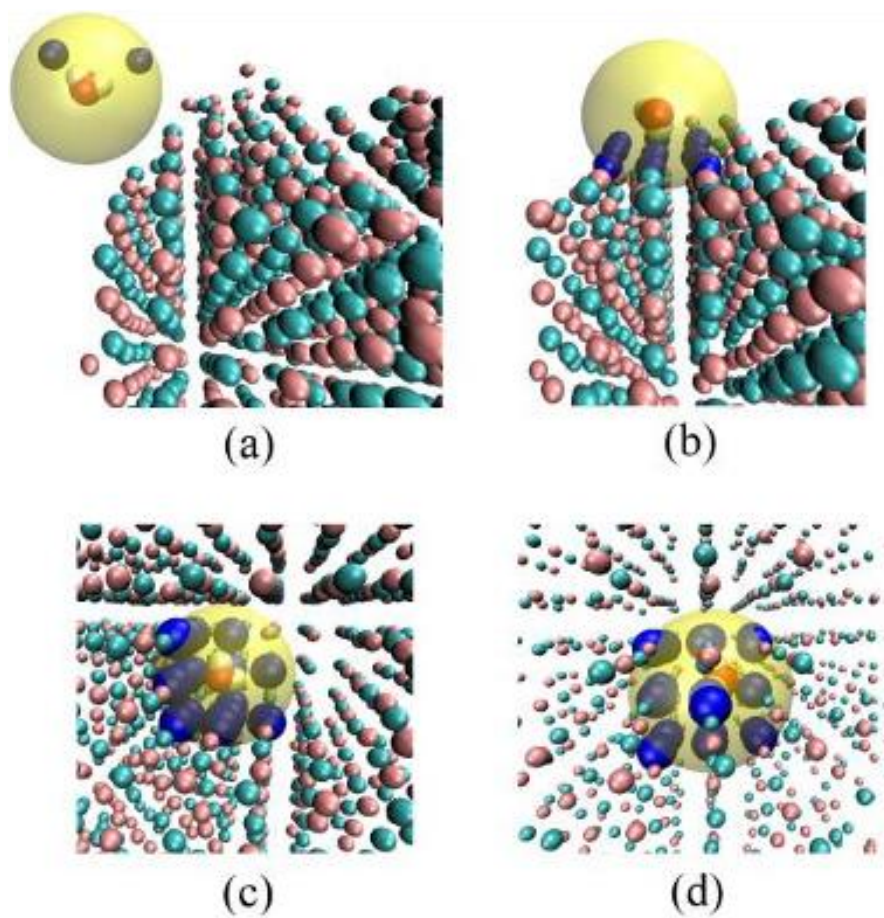
Figure A1-12 Marked ions (blue colour) within 5.5 Å of water molecules at different positions; (a) in the solution far from the crystal surface, (b) touching the surface of the crystal, (c) located inside the holes on the surface of the crystal, and (d) trapped inside the lattice.

The total number of 170 water molecules were identified, and a VMD visualization of the labeled molecules confirmed that the identification was successful.

Based on this analysis, the ratio of water molecules and ions in the crystal is 1:24, without any repeating pattern. Therefore, it can be concluded that the crystal is not a hydrate form of KCl crystal, and water molecules inside the lattice are just trapped due to kinetic effects.

A1.3.4 Ionic Bond Distance and Angle

The determination of the bond distance of a crystalline structure dates back to 1920s when Bragg suggested that in a crystal the atomic distances could be regarded as summation of the atomic radii (Bragg & Bell, 1921). In 1964, Slater proposed that, in fact, ionic radii must be used for bond distance calculation (Slater, 1964). He pointed out that usually positive ions are approximately 0.85 Å smaller than the atomic radii while negative ions are about 0.85 Å larger than the atomic radii and, therefore, in most cases, the sum of the atomic radii are almost same as the ionic radii of the same elements despite the fact that the ionic radii differ considerably from the atomic radii (Slater, 1964). The ionic radius of potassium and chlorine were reported to be 1.33 Å and 1.81 Å, respectively (Slater, 1964). Therefore, the bond distance of KCl in the crystal lattice must be 3.14 Å (0.314 nm).

In KCl crystal lattice, the coordination number of each ion is six, meaning that each ion is surrounded with 6 oppositely charged ions. For bond distance analysis, a set of 7 ions, including one central K and six surrounding Cls, as a nucleation point, is selected and the time evolution of these 6 pair ionic distances, between 23 to 33 ns, is shown in Figure A1-13.
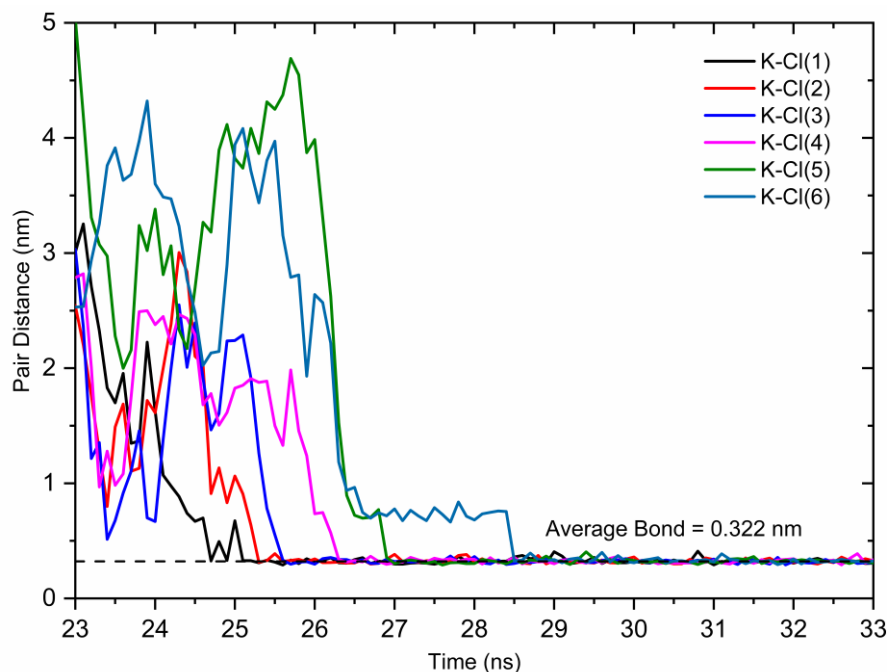
Figure A1-13 Bond distance of a central potassium ion with its 6 surrounding chlorine ions.

The time range was selected to focus on the nucleation process. You can see that initially, the distance is totally uncorrelated, and ions are moving freely across the simulation box, as far as 5 nm. However, after 29 ns, all of the 6 bonds are virtually formed, fixing the pair distance around the average value of 0.322 nm, which is in good agreement with the expected distance of 0.314 nm.

More interestingly, Figure A1-13 can shed light on nucleation process; as one can see, all six Cl ions attached to the central K ion in a sequential manner, starting from black line and ending at the blue line. This is a piece of evidence that classical nucleation theory (CNT) is valid for KCl crystallization and crystal nuclei form from a sequence of ionic additions. To check the general applicability of this observation, all crystalline ions were detected by the ion connectivity methodology, and their trajectories were studied throughout the simulation. More than 4,000 graphs were generated, proving that the process of single ionic addition to the nuclei is a ubiquitously upheld phenomenon in our MD simulations.

Besides the bond distance, we also need to check the angle between neighboring ions to confirm that the formed structure is crystalline. In a perfect face-centered cubic (FCC) crystal, which is the known crystal structure for KCl, the angles of neighboring ions are expected to be all 90 degrees.

The average angle of ions in the center of a selected crystal is plotted in Figure A1-14. One can see that during first 30 ns of the simulation, the average angle fluctuates considerably but around 30 ns, when the ions connect together, a significant reduction in the angle fluctuations is observed, stabilizing around 90 degrees for the rest of the simulation with the average of 89.94 degrees. The cause of the small fluctuations in the average angle after nucleation is due to dynamic simulation, statistical noise, and temperature of the system.

As a complementary analysis, the angle distribution of the neighboring ions is also plotted in Figure A1-15, signifying a Gaussian-like distribution with the highest probability of 90 degrees.
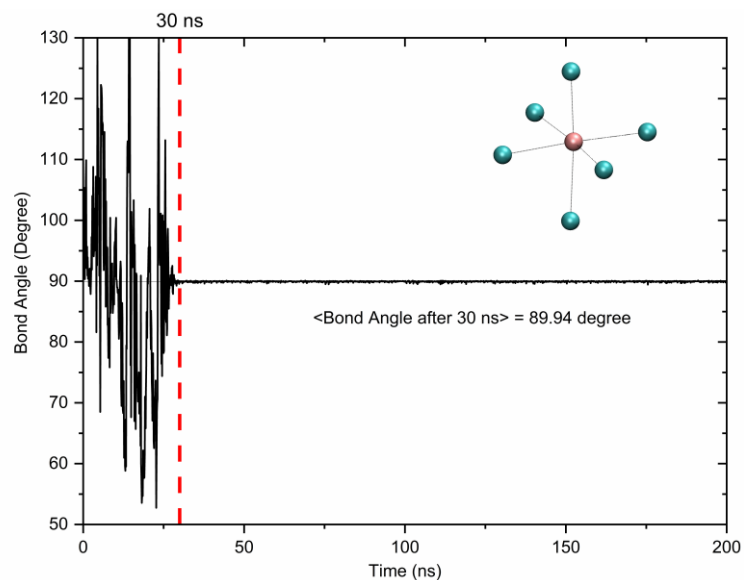
Figure A1-14 Time evolution of the averaged angle of neighboring chlorines of a central potassium ion.
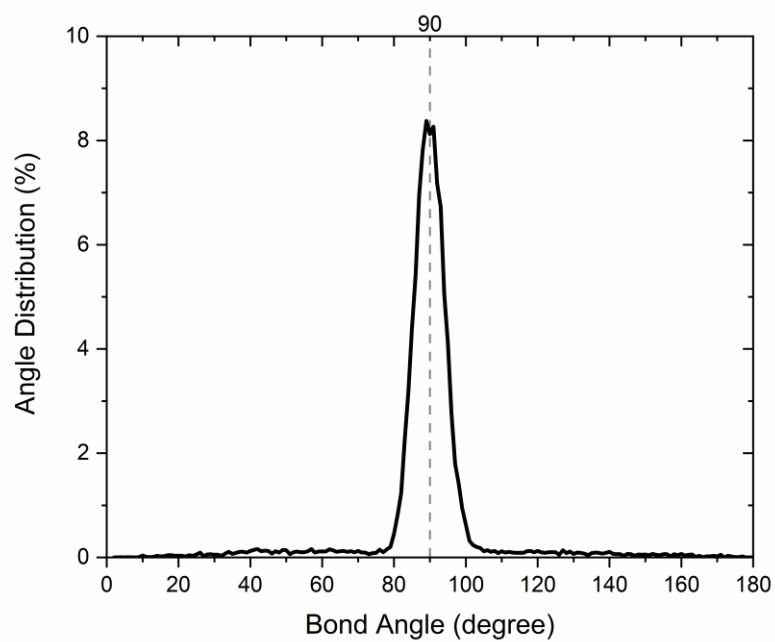
Figure A1-15 Angle distribution of neighboring chlorines of a central potassium ion over 200 ns of simulation.

A1.3.5   Local Supersaturation

Nucleation is a random event, and based on two-step nucleation theory, crystalline nuclei appear inside suspended pre-existing metastable clusters in the solution, consisting of dense liquid (Vekilov, 2010). In furtherance of the analysis of this simulation, a Python script was written to keep track of the local supersaturation around the already-known centers of nucleation. To achieve this aim, the number of ions in a sphere with a radius of 1 nm was counted. Diving the number of ions with the volume of the sphere results in local ionic density which is further converted to concentration by molecular weight. The result of this analysis, for the same nucleation center as Figure A1-13, is shown in Figure A1-16 that sheds light on early stage of crystallization.
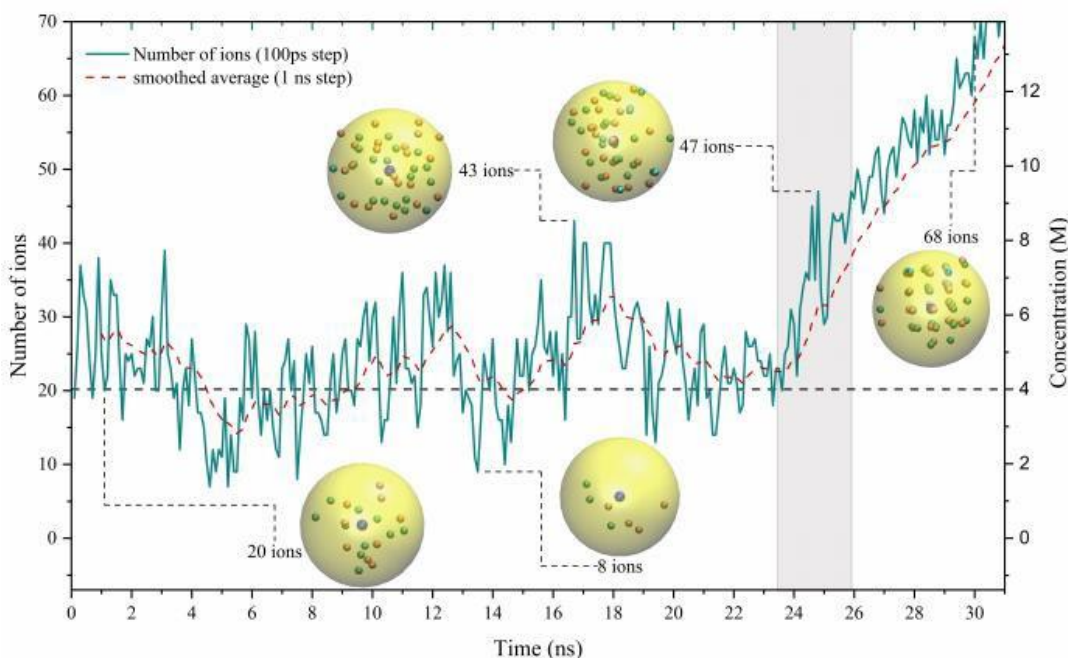
Figure A1-16 Number of ions and ionic concentration within 1 nm of a selected center of nucleation, before stabilization of the nucleus. The 6 ions in the first coordination shell of the stable nucleus were attached between 25 to 29 ns (refer to Figure A1-13).

As depicted in Figure A1-16, although significant fluctuations in local supersaturation are observed, mainly due to small size of the viewing sphere, the average concentration before nucleation is around the macroscopic concentration of the solution (4M). However, just before

the start of the nucleation at 25 ns (refer to Figure A1-13), a meaningful upward trend in local supersaturation starts, shown by the gray rectangle in Figure A1-16, and continues as nucleation occurs. A snapshot of the peak of local supersaturation within this transitional period (the gray rectangle) demonstrates the disorder orientation of ions, which finally leads to the formation of a well-ordered stable nucleus at 30 ns. This phenomenon is also observed for other nucleation sites.

In order to clarify that the structure of first, second, and third coordination numbers of the selected central ion are identical to those of a well-developed KCl crystal, even during nucleation, the CIF file of KCl (Persson, 2014) is converted to PDB file, and visualized with VMD to create the exact same colours and shapes as our simulation results for easier comparison.
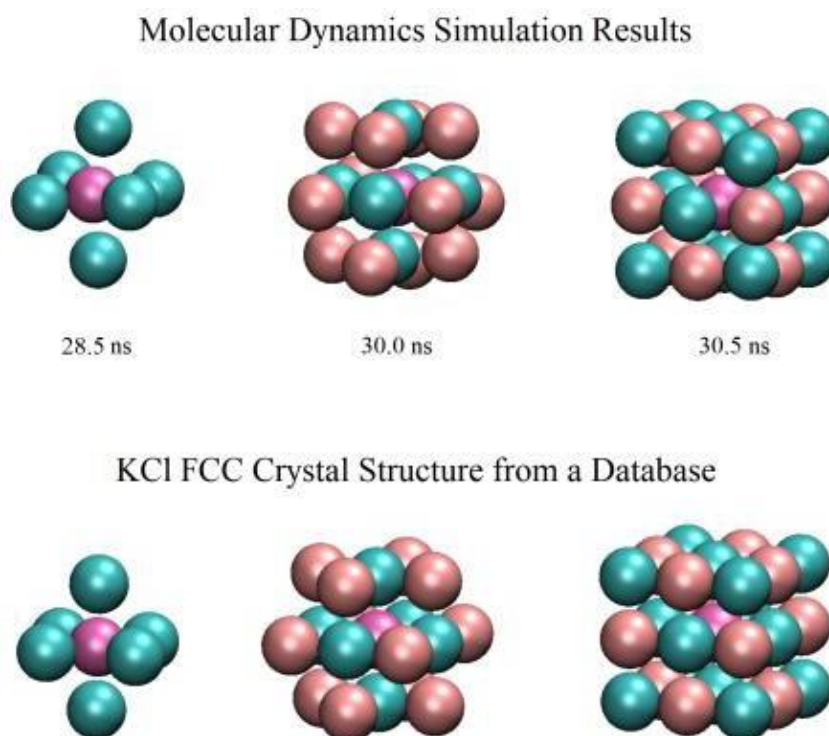
Figure A1-17 Comparison of 3 coordination numbers of a nucleation center between MD simulation and KCl face-centered cubic crystal structure form a database.

Figure A1-17 illustrates that the only difference between the early stages of crystal formation in the MD simulation and the well- equilibrated crystal is small fluctuations in the position of

some ions around the central atom, which is due to the dynamic nature of the simulation and the atomic vibrations. Otherwise, the structure, positions, and angles predicted by our simulation match with those of the database.

Configurational snapshots of the formation of the selected crystal from a larger viewing window with a radius of 2.5 nm are also shown in Figure A1-18 as complementary evidence. It can be seen that a disordered high-density region is formed at 27 ns and quickly, over 3 ns, results in a more ordered crystal embryo. This nucleation site then continued to grow over time, retaining its initial FCC structure, and finally formed a relatively large FCC structure, as shown in last snapshot.

| 0 ns | 27 ns | 30 ns | 50 ns | 200 ns |
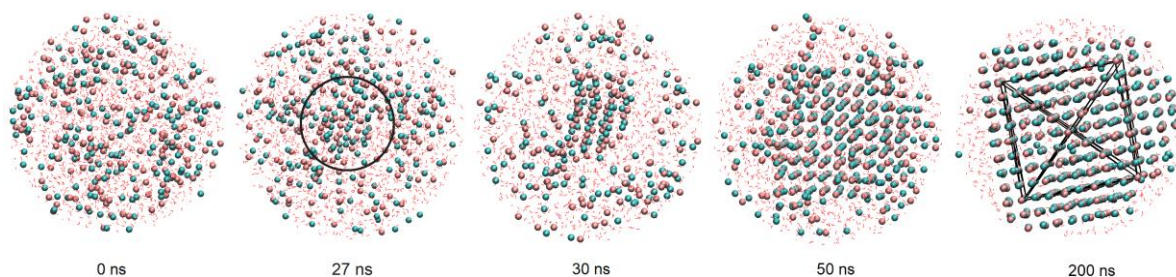
Figure A1-18 Configurational snapshots of the process of KCl crystal formation in aqueous solution over 200 ns of simulation within 2.5 nm of the central ion.

A1.3.6 Determination of Ionic Clusters and Crystals

Local ion connectivity analysis, as described in the section 2.3.1, with the aid of Python scripting to find and label the ions that are surrounded by 6 oppositely charged ions, is utilized to determine the formation of unstable clusters before the occurrence of nucleation, as well as identifying the first stable nuclei as the onset of nucleation.
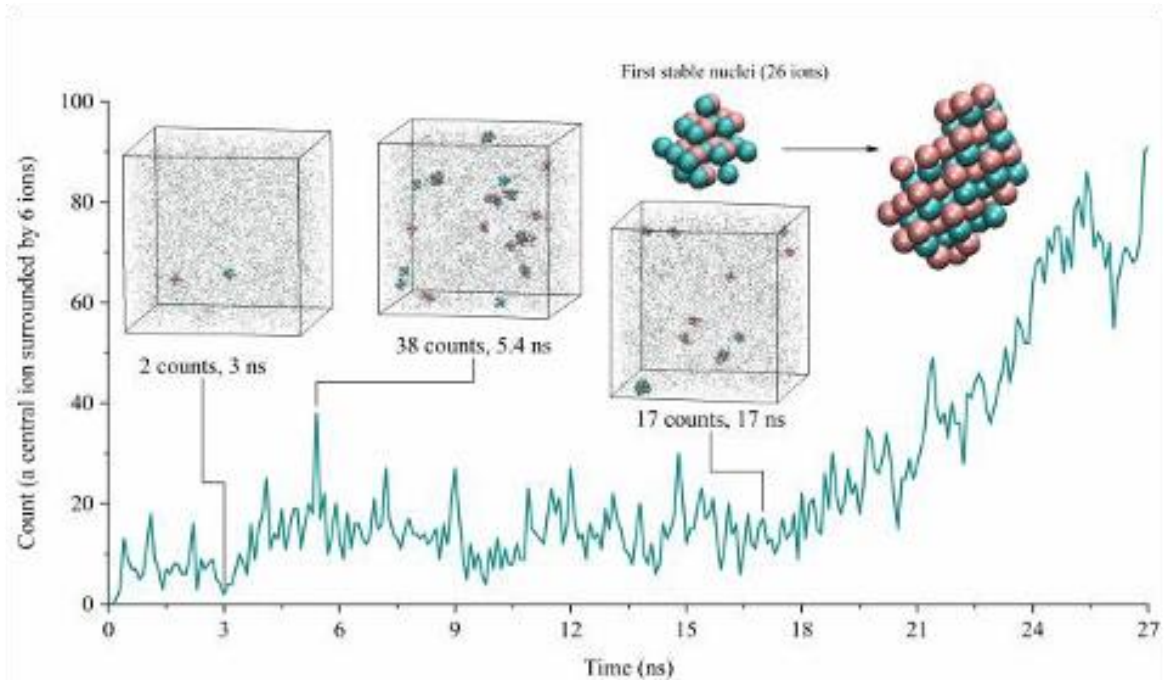
Figure A1-19 Count of central ions with 6 oppositely charged neighbors over initial stages of crystal formation.

As shown in Figure A1-19, at the initial configuration, there is no cluster. Soon afterward, several pre-nucleation clusters start to form and immediately disappear. For the first 20 ns, the average number of 7-ion clusters at each frame is 15 occurrences, ranging from 2 to 38. The configurational snapshots of some cases are attached to Figure A1-19. Notice that, in order to vividly visualize the clusters, the clusters are represented by larger spheres while other ions, presenting the solution, are shown by small points. Some of these centers are interconnected and formed larger clusters, but still, most of them did not reach the critical size of stability, and, consequently, redissolved in the solution. The first stable nucleus was observed at 17 ns, consisting of 26 ions formed from 6 intertwined centers. This cluster has a high degree of crystallinity based on our proposed order parameter (Q_{12}), which measures to be 0.202, which belongs to the crystalline portion of Figure A1-6.

Please note that the size of this nucleus was so small to have an impactful effect on the whole system. As this nucleus starts to grow, as well as the formation of other nucleation sites, the rate of change of crystal formation experienced a significant incline. At 27 ns onwards, shown in Figure A1-20, the number of ions surrounded by 6 oppositely charged neighbors

experiences a rapid incline, affecting the macroscopic properties of the system such as total energy (Figure A1-8) .
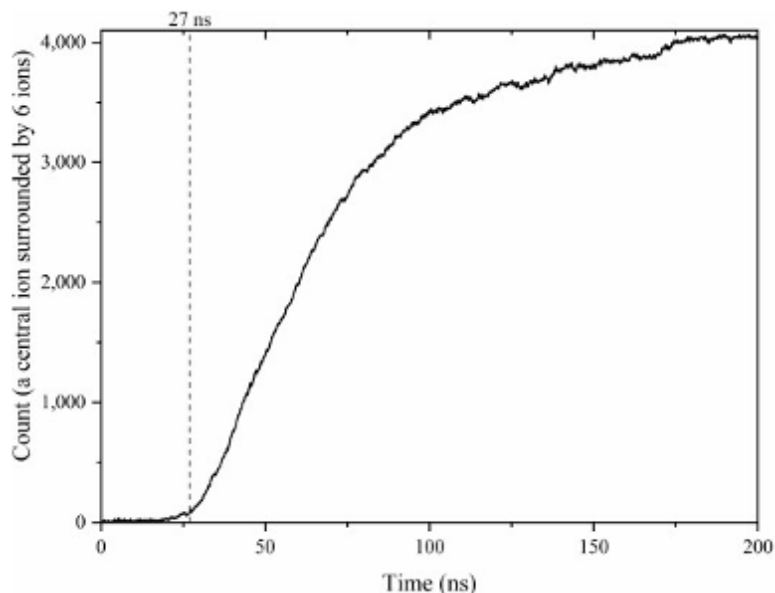
Figure A1-20 Count of central ions with 6 oppositely charged neighbors over 200 ns.

The count of central ions surrounded by 6 oppositely charged ions levels off at around 4000 ions (Figure A1-20), corresponding to the inner crystal ions and does not include the ions on the outermost layer of the crystals.

To identify all crystalline ions, the proposed order parameter of Q_{12}, described in section A2.3.2, is used to differentiate the ordered crystalline ions from the disordered ions in the solution. Figure A1-21 presents the filtered results of the order parameter calculation, illustrating the molar fraction of crystalline ions versus ions in solution.

As can be seen in Figure A1-21, during the pre-nucleation time of simulation, almost all of the ions belong to the solution. A small peak at 17 ns (the formation of the first stable nucleus) occurred, while at 27 ns, a significant incline in the number of crystalline ions started. This upward trend continues until 150 ns and then reaches a plateau with some minor fluctuations in the Q_{12} order parameter, resulted from the constant movement of the ions in a dynamic simulation. On average, during the last 50 ns of the simulation, 0.63 fractions of ions showed a high degree of crystallinity, and the remaining 0.37 percent remained in the solution form with a low order parameter.
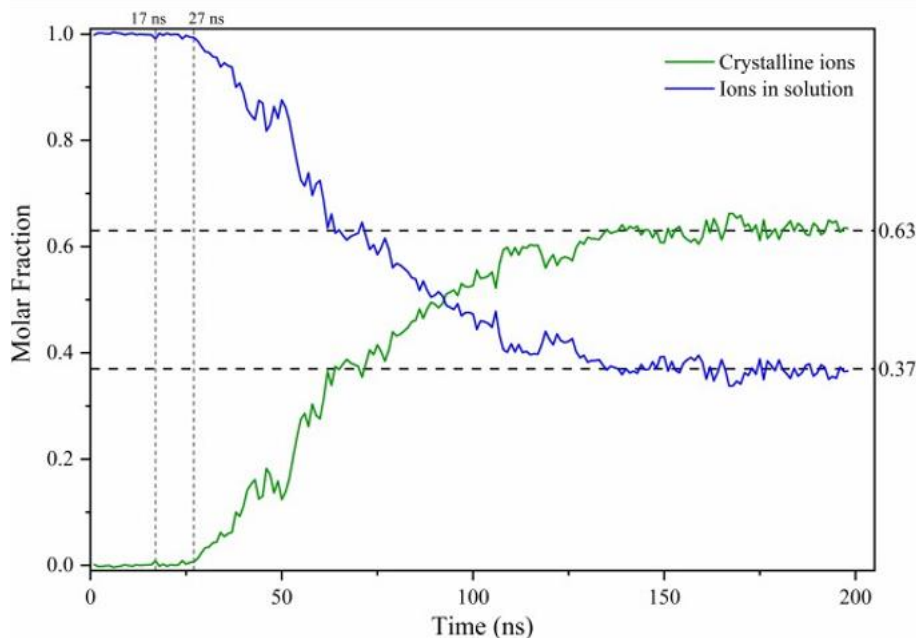
Figure A1-21 Mole fraction of ions in the simulation box, classified as crystalline ions versus ions in solution, using the proposed order parameter (Q_{12}).

A1.3.7   2D Number-density Maps

Another tool that can be used to investigate the crystallization process is 2D number-density planar map. GROMACS calculates the mass density of selected groups (in this case ions) and gives a plot of the density against a box axis. Distribution of ions across the interface can be determined by this method. The outcome of this analysis is shown in Figure A1-22. In Figure A1-22.a, one can see that initially the density of ions is evenly distributed in the system but during nucleation time (27 ns), shown in Figure A1-22.b, specific locations exhibit higher local density, shown by black contour line. This is an indication that during nucleation, higher local density regions, without having any ordered structure, are observed. The crystal nuclei can form in these sites and continue to grow over time, forming large crystalline areas with a high number density of ions, as shown in Figure A1-22.c, where large crystals are observed.
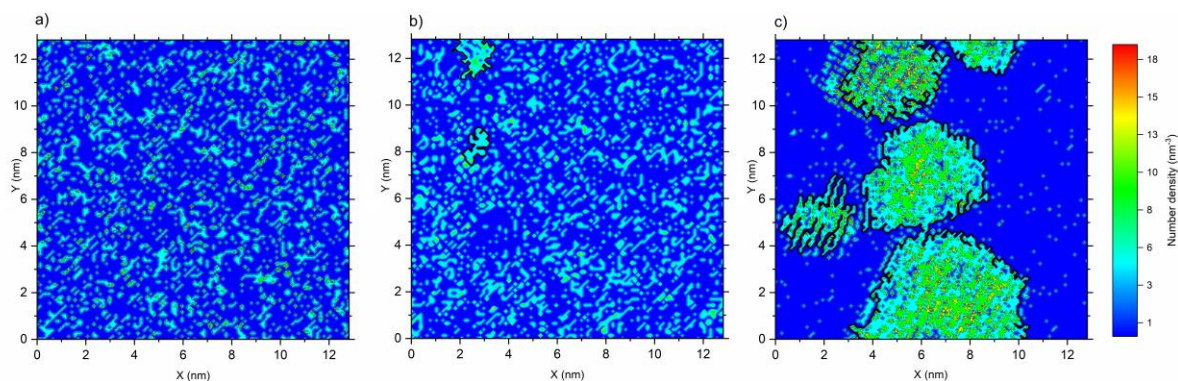
Figure A1-22 2D-number density of ions, averaged over a small portion of z-axis for a) initial configuration (0 ns), b) during nucleation time (27 ns), c) final configuration (200 ns).

In 2013, nucleation of calcium phosphate was experimentally investigated, concluding that the nanometer-scale pre-nucleation clusters are, in fact, calcium complexes that decrease the energy barrier for nucleation. This was considered as a demonstration to unite classical and two-step nucleation theories (Habraken et al., 2013).

The combination of Figures A1-13, 14, 17, and 19, suggesting classical nucleation theory, and on the other hand, Figure A1-16, 18, 22b, suggesting two-step nucleation theory, demonstrate the possibility to unite these two nucleation theories. In order words, crystal nuclei form as a result of a sequence of atomic, ionic, or molecular additions, but this can only happen in places that local density of ions is higher than solution concentration and the probability of having effective collisions increases, making these sites the primary candidates for nucleation.

A1.4 Conclusion

In this work, we have studied the crystal nucleation process of potassium chloride (KCl) in an aqueous solution at 293 K and 1 atm, using molecular dynamics (MD) simulation. Various analyses have been conducted to determine the structure of the newly emerged phase after equilibration, concluding that the new clusters are face- centered cubic crystals with some water molecules trapped inside the crystal lattice. The analyses of the early stages of the process showed that the nucleation occurred with a sequence of ionic additions, confirming

the classical nucleation theory. Further investigations, however, showed that local density of ions rises within the areas where nucleation started, confirming the two-step nucleation theory. Therefore, it was concluded that nucleation occurs within high-density regions, where the probability of effective collisions is higher, but the addition of ions happens in a sequential manner.

A1.5 References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–26.

Åqvist, J. (1990). Ion-water interaction potentials derived from free energy perturbation simulations. *The Journal of Physical Chemistry*, *94*(21), 8021–8024.

Becker, R., & Döring, W. (1935). Kinetische Behandlung der Keimbildung in übersättigten Dämpfen. *Annalen Der Physik*, *416*(8), 719–752.

Berendsen, H. J. C., Grigera, J. R., & Straatsma, T. P. (1987). The missing term in effective pair potentials. *The Journal of Physical Chemistry*, *91*(24), 6269–6271.

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*(1), 43–56.

Bragg, W. L., & Bell, H. (1921). The Dimensions of Atoms and Molecules. *Nature*, *107*(2682), 107.

Bussi, G., Zykova-Timan, T., & Parrinello, M. (2009). Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *The Journal of Chemical Physics*, *130*(7), 74101.

Chakraborty, D., & Patey, G. N. (2013). How Crystals Nucleate and Grow in Aqueous NaCl Solution. *The Journal of Physical Chemistry Letters*, *4*(4), 573–578.

Chandrasekhar, J., Spellmeyer, D. C., & Jorgensen, W. L. (1984). Energy component analysis for dilute aqueous solutions of lithium(1+), sodium(1+), fluoride(1-), and chloride(1-) ions. *Journal of the American Chemical Society*, *106*(4), 903–910.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12), 10089–10092.

Erdemir, D., Lee, A. Y., & Myerson, A. S. (2009). Nucleation of Crystals from Solution: Classical and Two-Step Models. *Accounts of Chemical Research*, *42*(5), 621–629.

Frenkel, J. (1939). A General Theory of Heterophase Fluctuations and Pretransition Phenomena. *The Journal of Chemical Physics*, *7*(7), 538–547.

Giberti, F., Tribello, G. A., & Parrinello, M. (2013). Transient Polymorphism in NaCl. *Journal of Chemical Theory and Computation*, *9*(6), 2526–2530.

Habraken, W. J. E. M., Tao, J., Brylka, L. J., Friedrich, H., Bertinetti, L., Schenk, A. S., Verch, A., Dmitrovic, V., Bomans, P. H. H., Frederik, P. M., Laven, J., van der Schoot, P., Aichmayer, B., de With, G., DeYoreo, J. J., & Sommerdijk, N. A. J. M. (2013). Ion-association complexes unite classical and non-classical theories for the biomimetic nucleation of calcium phosphate. *Nature Communications*, *4*, 1507.

Hess, B. (2008). P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *Journal of Chemical Theory and Computation*, *4*(1), 116–122.

Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, *4*(3), 435–447.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics*, *14*, 33–38.

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, *118*(45), 11225–11236.

Karthika, S., Radhakrishnan, T. K., & Kalaichelvi, P. (2016). A Review of Classical and Nonclassical Nucleation Theories. *Crystal Growth & Design*, *16*(11), 6663–6681.

Lanaro, G., & Patey, G. N. (2016). Birth of NaCl Crystals: Insights from Molecular Simulations. *The Journal of Physical Chemistry B*, *120*(34), 9076–9087.

Lindahl, E., Hess, B., & van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular Modeling Annual*, *7*(8), 306–317.

Mullin, J. W. (1997). *Crystallization: Third Edition*. Butterworth-Heinemann.

Nahtigal, I. G., Zasetsky, A. Y., & Svishchev, I. M. (2008). Nucleation of NaCl Nanoparticles in Supercritical Water: Molecular Dynamics Simulations. *The Journal of Physical Chemistry B*, *112*(25), 7537–7543.

Ohtaki, H., & Fukushima, N. (1991). Nucleation processes of NaCl and CsF crystals from aqueous solutions studied by molecular dynamics simulations. *Pure and Applied Chemistry*, *63*, 1743–1748.

Ostwald W. (1897). Studien über die Bildung und Umwandlung fester Körper. In *Zeitschrift für Physikalische Chemie* (Vol. 22U, p. 289).

Páll, S., Abraham, M. J., Kutzner, C., Hess, B., & Lindahl, E. (2015). *Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS BT - Solving Software Challenges for Exascale* (S. Markidis & E. Laure, Eds.; pp. 3–27). Springer International Publishing.

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, *52*(12), 7182–7190.

Peng, H., Gudgeon, J., & Vaughan, J. (2019). Nucleation phenomena of supersaturated KCl solutions revealing by molecular dynamic simulation: Implication of dehydration shell process. *Journal of Molecular Liquids*, *283*, 108–115.

Persson, K. (2014). *Materials Data on KCl (SG:225) by Materials Project*.

Pinho, S. P., & Macedo, E. A. (2005). Solubility of NaCl, NaBr, and KCl in Water, Methanol, Ethanol, and Their Mixed Solvents. *Journal of Chemical & Engineering Data*, *50*(1), 29–32.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., & Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, *29*(7), 845–854.

Rein ten Wolde, P., & Frenkel, D. (1999). Homogeneous nucleation and the Ostwald step rule. *Physical Chemistry Chemical Physics*, *1*(9), 2191–2196.

Slater, J. C. (1964). Atomic Radii in Crystals. *The Journal of Chemical Physics*, *41*(10), 3199–3204.

Steinhardt, P. J., Nelson, D. R., & Ronchetti, M. (1983). Bond-orientational order in liquids and glasses. *Physical Review B*, *28*(2), 784–805.

Stone, J. (1998). *An Efficient Library for Parallel Ray Tracing and Animation*. Computer Science Department, University of Missouri-Rolla.

Talanquer, V., & Oxtoby, D. W. (1998). Crystal nucleation in the presence of a metastable critical point. *The Journal of Chemical Physics*, *109*(1), 223–227.

Vacek, V., & König, A. (1983). Heats of crystallization of potassium chloride and magnesium chloride hexahydrate from aqueous solutions at 298.15 K. A critical review. *Thermochimica Acta*, *70*(1), 225–235.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, *26*(16), 1701–1718.

Vekilov, P. G. (2010). The two-step mechanism of nucleation of crystals in solution. *Nanoscale*, *2*(11), 2346–2357.

Vogt, J., & Weiss, H. (2001). The structure of NaCl(100) and KCl(100) single crystal surfaces: a tensor low energy electron diffraction analysis. *Surface Science*, *491*(1), 155–168.

Volmer, M., & Weber, A. Z. (1926). Nucleus Formation in Supersaturated Systems. *Zeitschrift Für Physikalische Chemie*, *119*, 277–301.

Zahn, D. (2004). Atomistic Mechanism of NaCl Nucleation from an Aqueous Solution. *Physical Review Letters*, *92*(4), 40801–1, 40801–40804.

Zimmermann, N. E. R., Vorselaars, B., Quigley, D., & Peters, B. (2015). Nucleation of NaCl from Aqueous Solution: Critical Sizes, Ion-Attachment Kinetics, and Rates. *Journal of the American Chemical Society*, *137*(41), 13352–13361.

Zimmermann, Nils. E. R., Vorselaars, B., Espinosa, J. R., Quigley, D., Smith, W. R., Sanz, E., Vega, C., & Peters, B. (2018). NaCl nucleation from brine in seeded simulations: Sources of uncertainty in rate estimates. *The Journal of Chemical Physics*, *148*(22), 222838.

Appendix 2

Cocrystals, Salts, and Salt-Solvates of Olanzapine; Selection of Coformers and Improved Solubility

A2 Cocrystals, Salts, and Salt-Solvates of Olanzapine; Selection of Coformers and Improved Solubility

Abstract

Pharmaceutical cocrystals and salts are extensively researched in recent years due to their ability to tune the physicochemical properties of active pharmaceutical ingredients (APIs). A model API, Olanzapine, an atypical antipsychotic drug classified as Biopharmaceutical Classification System class II, is used in this study. Cocrystals and salts of Olanzapine are discovered using solvent drop grinding and ball milling. Appropriate coformers were selected based on a combination of hydrogen-bond propensity (HBP) and hydrogen-bond coordination (HBC) calculations. Eight new multicomponent phases of Olanzapine, including one cocrystal hydrate with phenol; four anhydrous salts with salicylic acid, terephthalic acid, anthranilic acid, 3-hydroxybenzoic acid, and 2-aminoterephthalic acid; one salt dihydrate with terephthalic acid; and one salt solvate with 3-hydroxybenzoic acid and acetonitrile, have been discovered and characterized by PXRD and DSC. One reported cocrystal (Olanzapine-resorcinol) has also been considered for the dissolution test. All these newly formed solid phases followed the "ΔpK_a rule of 3". The crystal structures of cocrystal/salts were determined by single-crystal X-ray diffraction (SCXRD). With the collected single-crystal data, the crystal packings were found to be primarily stabilized via strong hydrogen bonds between carboxyl, phenolic hydroxyl of co-formers/salt-formers with the piperazine and diazepine nitrogen of Olanzapine, which confirmed the predicted result from the HBP and HBC calculations. HPLC coupled with UV-vis detector was used in the solubility and dissolution test instead of UV-vis spectroscopy, to avoid the peak overlap between Olanzapine and co-formers/salt-formers. A threefold increase in the solubility was observed in olanzapinium 3-hydroxybenzoate and olanzapinium anthranilate, and an almost fivefold increase in solubility of olanzapinium 2-aminoterephthalate.

A2.1 Introduction

In recent years, the pharmaceutical industry has increasingly utilized the potential of crystal engineering to tune the performance of active pharmaceutical ingredients (APIs) (Duggirala et al., 2016). Crystal engineering can be applied to most crystalline APIs to improve their stability, solubility, dissolution, and bioavailability (Almarsson & Zaworotko, 2004; Blagden et al., 2007). Therefore, screening and selection of the potential solid states of an API is one of the most critical early stages of drug development. A suitable solid-state of an API offers convenience in formulation and therapeutic efficacy (Duggirala et al., 2016; Karimi-Jafari et al., 2018; Rodrigues et al., 2018). The various crystalline solids of APIs include polymorphs, solvates and hydrates, cocrystals, and salts. Compared to other solid-state forms, pharmaceutical cocrystals usually have well-defined stoichiometry, higher thermal and humidity stability (Dalpiaz et al., 2018; Duggirala et al., 2016; Huang et al., 2019; L. S. Reddy et al., 2009; Weyna et al., 2009), and can be designed with a large number of potential cocrystal formers (coformers). Therefore, pharmaceutical cocrystals have recently emerged as an important pharmaceutical solid form selection with a concurrent patent activity rise (Clarke et al., 2012). Nevertheless, only a few API cocrystal formulations are approved in the current market. Thus, salt preparation has still the highest potential among all crystalline solid-state formulations (Nechipadappu et al., 2019; Sarmah et al., 2016, 2018a). Salt formulation can increase the solubility and dissolution of APIs in polar solvents. A detailed study of crystalline solid states of Olanzapine, as a model compound, has been researched in this paper using a large number of coformers/salt-formers to test the well-known ΔpK_a rule. As Olanzapine molecule has a strong unbonded hydrogen bond acceptor in its pure crystal, acidic functional groups were considered that have potential to form the hydrogen bond. Regarding carboxylic acid coformers, a number of olanzapinium salts have already been reported (Sarmah et al., 2018a). Therefore, in this paper, we focused on exploring phenol and benzoic acid derivatives as hydrogen bond donors. The experimental data have been confirmed/predicted by a combination of hydrogen-bond propensity (HBP) and hydrogen-bond coordination (HBC) calculations.

2-Methyl-4-(4-methyl-1-piperazinyl)-10H-thieno[2,3-b]-[1,5]benzodiazepine, well-known as Olanzapine (OLN), is an atypical antipsychotic drug widely used to treat schizophrenia and bipolar disorder (Bymaster et al., 1996; Nanubolu & Ravikumar, 2017). It represents a good crystal engineering model compound as it displays a low solubility and a high permeability (Clarke et al., 2012; Thakuria & Nangia, 2013), which belongs to the BCS (Biopharmaceutics Classification System) (Tsume et al., 2014) class II. For APIs belonging to this class, solid-state modification is expected to offer better efficacy and efficiency (Sarmah et al., 2016). Thus, previous studies have reported six anhydrous polymorphs(Clarke et al., 2012; Nanubolu & Ravikumar, 2017; Reutzel-Edens et al., 2003; Sarmah et al., 2016; Thakuria & Nangia, 2013; Tiwari et al., 2007), three polymorphic dihydrates, two polymorphic sesquihydrates, several solvates (Cavallari et al., 2013), mixed solvated forms (Wawrzycka-Gorczyca et al., 2007), cocrystals and salts (Clarke et al., 2012; Sarmah et al., 2016, 2018b; Surampudi et al., 2020; Thakuria & Nangia, 2013). Clarke et al. (2012) has systematically studied few quaternary solvated isostructural cocrystals of Olanzapine and categorized them based on the crystal packing arrangement. Furthermore, mechanochemical liquid-assisted grinding (LAG) approach was used to prepare olanzapinium salts with dicarboxylic acids to improve the hydration stability and solubility of OLN (Sarmah et al., 2018b, 2020). The elimination of side effects of OLN was realized by the design of Olanzapine-nateglinide and Olanzapine-pyrazinoic acid cocrystals (Sarmah et al., 2020), and the thermal properties of OLN were improved by dihydroxybenzene coformers (Surampudi et al., 2020).

Despite the existence of several publications on Olanzapine, it continues to serve as a useful model API. Herein in this paper, a systematic study on the prediction and selection of proper salt-formers and coformers, the synthesis of multi-component crystals of a model API, and the characterization of the targeted physicochemical property of OLN was proposed for first time. Several crystal engineering tools are studied on Olanzapine: first, with synthon analysis a set of coformers and salt-formers were selected, and their newly discovered crystals were fully characterized. The experimental screening was performed on more than 50 cocrystal/salt formers, only salicylic acid, terephthalic acid, anthranilic acid, 3-hydroxybenzoic acid, 2-aminoterephthalic acid, phenol and resorcinol resulted in new phases. The formation of these new crystalline forms agreed with hydrogen bonding propensity and coordination predictions,

as well as ΔpK_a rule of 3 to differentiate salts from cocrystals. Lastly, the solubility and dissolution tests of the discovered OLN multicomponent crystals were performed that showed promising improvements in these physicochemical properties.

A2.2 Experimental Section

A2.2.1   Preparation of Olanzapine Cocrystals, Salts and Salt-solvates

Olanzapine was donated by Apotex PharmaChem Inc. 3-Hydroxy benzoic acid (3HBA) (>99%) was purchased from ACROS Organics; anthranilic acid (AA) (>98%) and phenol (Phol) were purchased from Sigma-Aldrich; terephthalic acid (TA) (98+%), salicylic acid (SA) (99+%), resorcinol (Res) (99.0-100.5%), and 2-aminoterephthalic acid (2ATPA) (99%) were all purchased from Alfa Aesar. Coformers were used as purchased. Liquid-assisted grinding (LAG), also known as solvent drop grinding (SDG), was used by mechanical grinding of the API and the coformer/salt-former using an agate mortar and pestle, for the preparation of cocrystals/salts. Methanol was used as the solvent for grinding. Ball milling process was used in parallel to prepare bulk powders using a Retsch Mixer Miller 200 (Haan, Germany), a mixer mill equipped with a stainless-steel jar (5ml PTFE SmartSnap Jar), and two 5 mm Zirconia grinding balls. The mixture in the stainless-steel jar was blended for 10 min under a 12.5 Hz vibrational frequency on the Mixer Miller. A 1:1 stoichiometric ratio of the solid API and a coformer (200mg in total) was used in the initial studies for both ball milling and LAG. The product obtained from grinding was air-dried and characterized by PXRD and DSC. The product powder was then crystallized using different HPLC grade solvents (dichloromethane, chloroform, isopropanol, acetonitrile, ethanol, toluene, tetrahydrofuran, ethyl acetate and methanol) in 5.0 ml beakers and then kept for crystallization at room temperature (24°C) or at a lower temperature (5°C). Single crystals obtained from crystallization were then characterized structurally using SCXRD. After the cocrystal/salt structure was obtained by SCXRD, the observed stoichiometric ratio was used to produce the bulk cocrystals/salts (500 mg in total) through ball milling and LAG. The produced bulk cocrystals were characterized by PXRD. The PXRD patterns of these reproduced bulk samples were compared with the simulated PXRD patterns obtained from the SCXRD data to check the consistency. The detailed procedure of the production of the OLN multicomponent crystals is described below.

A2.2.1.1   OLN-SA, OLN-Res, OLN-AA , OLN-3HBA , OLN-2ATPA

A one-to-one stoichiometric ratio mixture of OLN with SA, Res, AA, 3HBA, and 2ATPA was manually ground with mortar and pestle using liquid assisted grinding conditions. Every 10 min, 8-10 drops of methanol were added, and the materials were ground for about 30 min until a dry homogenous powder was obtained. PXRD was then used to compare the X-ray pattern of the mixture with the starting compounds to ensure a new phase was formed.

A2.2.1.2   OLN-TA and OLN-TA-H2O

A mixture of 2 to 1 stoichiometric ratio of OLN and TA was manually ground in mortar and pestle using liquid assisted grinding with the same procedure discussed above, and the product was characterized by the PXRD to produce the anhydrous OLN-TA powder. Based on the single-crystal data, OLN-TA hydrate was detected in the asymmetric unit. Two drops of deionized water were introduced into half of the obtained ground bulk crystals and the mixture was then ground for another fifteen minutes to produce the hydrate phase. The products were then characterized by the PXRD. The diffraction patterns were compared to the simulated patterns from the SCXRD analysis.

A2.2.1.3   OLN-3HBA-ACN

Initially, the LAG process was used to produce the OLN-3HBA-ACN. Acetonitrile was used instead of methanol as the added solvent to produce the solvate. It was found that the obtained product was gummy and amorphous. Therefore, rotary vaporization was employed instead of LAG. Mixtures of 1:1 stoichiometric ratio of OLN and 3HBA were put in a 100 ml round-bottomed flask and were then dissolved in 50 ml HPLC grade acetonitrile with the help of ultrasound (110-120 V, 50/60 Hz, 10 min). The flask was set in a 50°C water bath for 1 h to evaporate the excess acetonitrile. The dried bulk OLN-3HBA-ACN was collected, characterized by the PXRD, and compared to the simulated pattern from the SCXRD.

A2.2.1.4   OLN-Phol-H_2O

A LAG process similar to the production of OLN-3HBA-ACN, was adopted to produce the OLN-Phol-H_2O. A 1:1 stoichiometric ratio of OLN and Phol mixture was ground in the

presence of methanol. After 30-min of grinding, a gummy mixture was formed as the product. Toluene, a non-polar solvent, was used to make the powder product. Ball milling with the assistance of methanol and several droplets of toluene led to the formation of OLN-Phol-H_2O powder. Crystals were dried after LAG and characterized by the PXRD.

A2.2.1.5   OLN-MeOH

Olanzapine methanol solvate was made to confirm the purity of the produced salts and cocrystals. Pure Olanzapine was ground with several droplets of methanol for thirty minutes. The product was characterized by the PXRD and DSC.

A2.2.2   Single Crystal X-ray Diffraction

The single crystal X-ray diffraction patterns of the obtained OLN multicomponent single crystals were obtained on the Bruker APEX-II CCD diffractometer using Mo Kα radiation (λ = 0.7107 Å). APEX2 software was used for data integration and reduction with SAINT. The data were collected at 110(2) K using an Oxford Cryostream low-temperature device. Based on the obtained information from SCXRD, crystal structures were solved by direct methods using SIR 2014 (Burla et al., 2015). Structure refinement was performed in the program package WinGX (Farrugia, 2012) and Olex2 1.3 (Dolomanov et al., 2009). All the non-hydrogen atoms were refined anisotropically by full-matrix least-squares calculations based on F2 with SHELXL-2016 (Sheldrick, 2015). All hydrogen atoms bonded to carbon were placed in the calculated positions, whereas acidic hydrogen atoms were located to confirm the salt formation. Mercury programs were utilized for structure analysis and the generation of crystal structures and packing diagrams. Linear acetonitrile introduces a strong disorder in the asymmetric unit of OLN-3HBA salt. To displace the disorder, we used Squeeze in the Olex2, and we removed the acetonitrile from the asymmetric unit. TWINABS-Version 2012/1-Bruker AXS scaling was used to identify 2 twin components of OLN-Phol-H_2O crystals.

A2.2.3   Powder X-ray Diffraction

The API, coformers, and all product cocrystals/salts samples were run on the powder X-ray diffraction (Rigaku, Miniflex) with the Cu-Kα source (λ for Kα= 1.54059 Å). Bulk powder of each sample was placed on a quartz-glass sample holder and measured using a continuous scan between 5° and 45° in 2θ with the 3°/min scan speed at 30 kV voltage and 15mA current.

A2.2.4   Differential Scanning Calorimetry

Analysis of the thermal properties of the API, coformers, cocrystals/salts was conducted using a differential scanning calorimetry (DSC, Mettler Toledo, Chicago, United States) under nitrogen gas atmosphere. Samples of precisely weighed cocrystals (5 mg to 10 mg) were placed in a non-hermetically sealed aluminum pan in a vacuum. Samples were scanned at a rate of 5°C/min in the range of 25−280°C under a dry nitrogen atmosphere at a flow rate of 100ml/min.

A2.2.5   Preparation of Buffer Solution

For the solubility tests, a phosphate buffer (0.1 M, pH 7.2) solution was prepared by dissolving monobasic potassium phosphate (KH_2PO_4, 27.2 g) and potassium hydroxide (KOH, 8.8 g) in 2 liters of distilled water at room temperature and adjusting the pH to 7.2 with 2N hydrochloric acid.

A2.2.6   Equilibrium Solubility and Powder Dissolution Measurement

Excess amounts (100-150 mg) of powder samples (OLN/ cocrystals/ salts/ solvates/ hydrates) were added to 10 ml of aqueous buffer medium (at pH 7.2), and the mixture was stirred gently at 3.2 Hz for 24 h at 37°C to measure the equilibrium solubility. After 24 h, the slurry was filtered using non-pyrogenic Filtropur S syringe filters with 0.45 µm pore size (Sarstedt AG & Co. KG, Germany) at room temperature.

It was found that the concentration of Olanzapine in the solution computed from the characteristic peak (262 nm^{-1}) in UV-vis spectroscopy interferes with coformers/salt-formers. Thus, High-Performance Liquid Chromatography (HPLC) was use in this study. Reddy et al. (2007) also applied HPLC method to determine the OLN concentration in the solution. The

reliability of the HPLC method was also verified by Basavaiah et al. (2014). After filtration, the clear solution was diluted with the aqueous buffer medium (10 times or 20 times) and was then quantified by HPLC coupled with UV-vis detector with the external standard method to determine its solubility. The peak area of OLN was traced and used for calculation.

Powder dissolution studies were performed in a similar way to the solubility test. Excess amounts (100-150 mg) of powder samples were placed in 15 ml glass vials, which were then filled with 10 ml of buffer solution and stirred at 3.2 Hz and 37°C. After specific time intervals, 0.2 ml of the slurry of samples were syringe filtered. The resultant solution was then diluted with the aqueous buffer medium (10 or 20 times) and quantified by HPLC to determine its concentration. Chromatographic separation was achieved on an HPLC Cartridge column (ChromSep, C18, 250x4.6 mm). The mobile phase was a 50:50 (v/v) mixture of 0.01M phosphate buffer and HPLC grade acetonitrile, the flowrate was fixed at 1.2 ml/min, and UV-detection was performed at 252nm (B. V. Reddy et al., 2007). The column temperature was settled at 25°C, and the analysis time was 15 min.

A standard calibration curve was used for solubility and dissolution calculations. The standard curve was determined with known concentrated OLN-buffer solution (OLN concentrations at 5.5, 4.4, 2.2, 1.1, and 0.55 mg/100ml). Ten microliters of the standard solution were injected automatically into the column in triplicate, and the chromatograms were recorded. Test solutions during the solubility and dissolution experiments were injected, and their corresponding OLN peak areas were used to compute the solution concentration. For higher accuracy, OLN-2ATPA in buffer solution was diluted twenty times before the measurement, and all other samples were diluted ten times in order to maintain the peak height lower than 1 mAU.

The pH of the buffer medium was measured at the beginning and the end of both equilibrium solubility and dissolution experiments using a Beckman pH meter. The excess solids remaining after the equilibrium solubility and dissolution experiments were dried and their PXRD and DSC patterns recorded.

A2.3 Results and Discussion

A2.3.1 Selection of Candidate Molecules for OLN Multi-component Crystals

The selection of coformers and salt-formers of OLN, in this study, was based on the hydrogen-bond propensity and hydrogen-bond coordination calculations. Typical synthons used in these calculations for OLN with phenol and benzoic acid are provided in Table A2-1 (Sandhu et al., 2018). CCDC Mercury (2020.2.0) (Macrae et al., 2020) was used for the calculation. The observed interaction in OLN crystal structure showed a rather low propensity of 0.25. This is an indication that the main hydrogen bond in OLN can be replaced with API-coformer interactions with higher propensities. As shown in Table A2-1.a, one can see that phenol and benzoic acid derivatives bond with significantly higher propensities (approximately 0.4) to OLN, which makes them ideal candidates for multicompetent crystallization. More specifically, the aromatic hydroxy and carboxy would act as the hydrogen bond donor and the piperazine N of OLN would act as the hydrogen bond acceptor (synthon A and B). For the OLN-benzoic acid, there exists another potential hydrogen bond former between the diazepine N (donor) and the carbonyl group (C=O) (acceptor) (synthon C).

Hydrogen bond coordination analysis, shown in Table A2-1.b, provided further evidence on candidacy of phenol and benzoic acid derivatives for formation of salts/cocrystals with OLN. The coordination number for the primary donor atom of OLN (0.852-0.855) to form one hydrogen bond is lower than the equivalent value for phenol (0.930) and benzoic acid (0.951). On the other hand, the main OLN acceptor (0.610-0.636) has a higher coordination number than benzoic acid (0.570) and phenol (0.203) acceptors. Therefore, paring of the coformers donor and API acceptor is favored, making these compounds excellent candidates for salts/cocrystal formation with OLN. According to the molecular structure of OLN, there are 3 acceptors (N) and only 1 donor (NH). This imbalance of hydrogen bond moieties leads to uncoordinated sites that favors the formation of multicomponent crystals.

All selected coformers and salt-formers followed the "ΔpK_a rule of 3" (Lemmerer et al., 2015), which states that cocrystals will form if $\Delta pK_a = pK_a$ (protonated base) $- pK_a$ (acid) is less than

0. On the other hand, molecular salts will form if the difference is greater than 3. The OLN molecule has two strong basic acceptors, piperazine N with a pK_a value of 7.37 and diazepine N with a pK_a value of 4.69. Based on the ΔpK_a rule, acids with pK_a value < 4.5 will form salts with OLN, and acids with pK_a value > 7 will form cocrystals.

Consequently, resorcinol (pK_{a1}= 9.26), phenol (pK_a= 10.02) as coformers; and 3-hydroxybenzoic acid (pK_a= 3.84), salicylic acid (pK_{a1}= 2.79), anthranilic acid (pK_a= 4.89), 2-aminoterephthalic acid (pK_a= 3.95), and terephthalic acid (pK_{a1}= 3.32) as salt-formers (see Figure A2-1) have been selected for this study. The hydrogen bond donors of these coformers and salt-formers (carboxyl, phenolic hydroxyl) participate in intermolecular hydrogen bonds with strong acceptors of OLN.
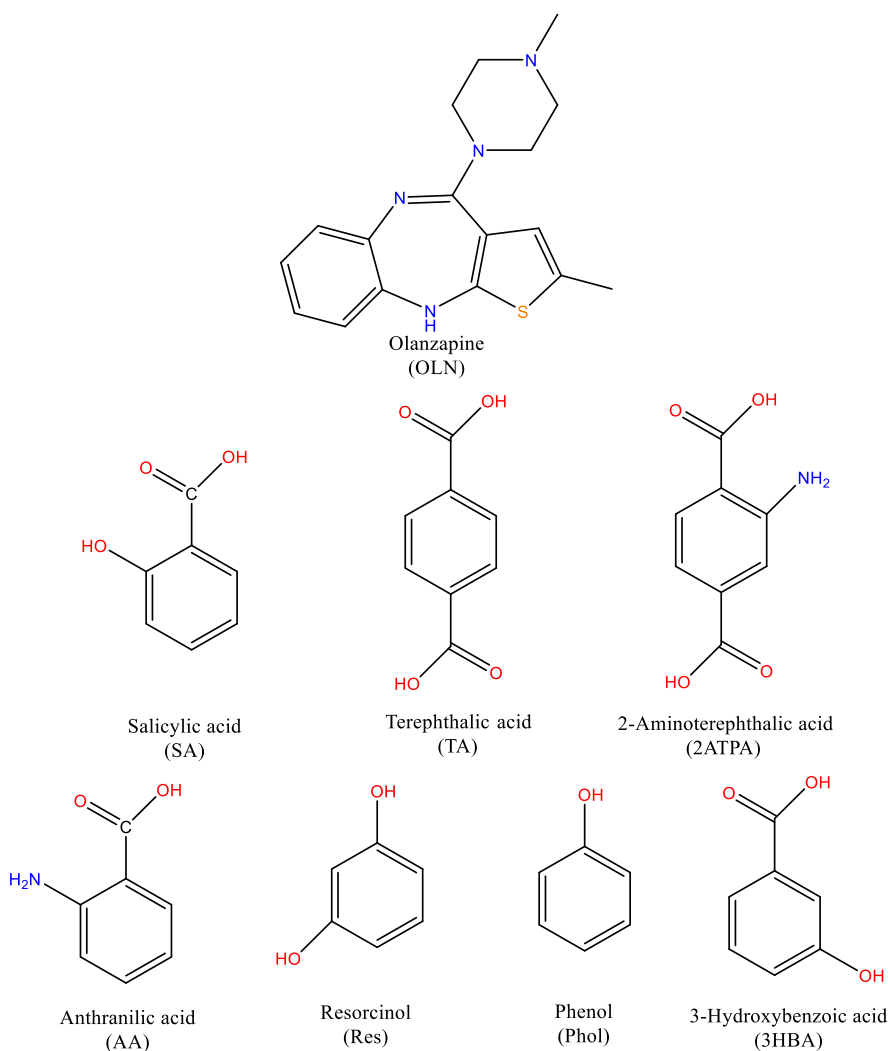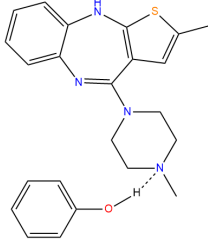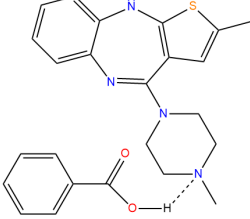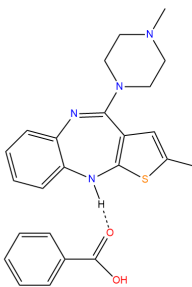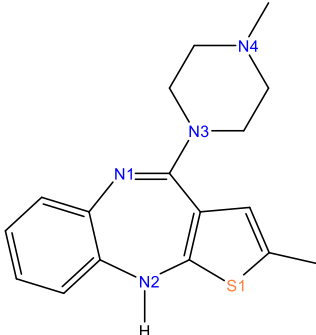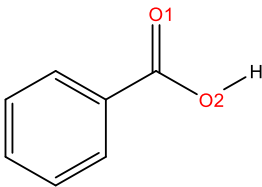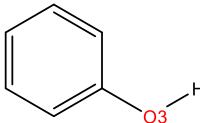
Figure A2-1 Molecular structure of OLN and coformers/salt formers used in this study.

Table A2-1 Predicted hydrogen-bond propensities (a), and hydrogen-bond coordination numbers (b), for OLN with phenol and benzoic acid.

a.	Synthon A	Synthon B	Synthon C

Hydrogen-bond propensities (lower bound, upper bound)

0.38 (0.27, 0.50)	0.39 (0.28, 0.52)	0.42 (0.28, 0.57)

b. Hydrogen-bond coordination numbers

Donor (D)/Acceptor (A) atom	Coordination number = 0	Coordination number = 1	Coordination number = 2
N1(A) in OLN-Phol	0.603	0.392	0.005
N1(A) in OLN-BA	0.657	0.339	0.004
N2 (D) in OLN-Phol	0.095	0.852	0.053
N2 (D) in OLN-BA	0.090	0.855	0.055
N3 (A) in OLN-Phol	0.940	0.060	0
N3 (A) in OLN-BA	0.946	0.054	0
N4 (A) in OLN-Phol	0.364	0.636	0
N4 (A) in OLN-BA	0.390	0.610	0
O1 (A) in OLN-BA	0.378	0.570	0.050
O2 (D) in OLN-BA	0.006	0.951	0.043
O3 (D) in OLN-Phol	0.020	0.930	0.050
O3 (A) in OLN- Phol	0.787	0.203	0.010

A2.3.2   Structural analysis

Single-crystal X-ray diffraction data confirmed the formation of new solid phases of OLN. The single crystal structure of Olanzapinium·Salicylate Benzene (1:1:1.5) (Thakuria & Nangia, 2013), Olanzapine·Resorcinol (1:1) (Surampudi et al., 2020), and the structure of anhydrous Olanzapine·Phenol (1:1) (Andrusenko et al., 2020) obtained by 3D electron diffraction have already been reported. Single crystals of anhydrous OLN with Res [OLN-Res, cocrystal]; anhydrous OLN with SA [OLN-SA, salt]; and AA [OLN-AA, salt] were successfully isolated and characterized. Meanwhile, hydrated cocrystal of OLN-Phol (water from air); hydrated salt of OLN-TA and solvated salt OLN-3HBA were also isolated and characterized. No suitable-size single crystal of OLN-2ATPA for the SCXRD could be isolated using various crystallization methods that were employed. The ORTEP diagrams and detailed crystal packing diagrams for each of the OLN multicomponent crystals are shown in Figure A2-2. As the OLN-Res structure was already reported by Surampudi et al. (2020), its structure was not analyzed by SCXRD. The list of primary intermolecular interactions of OLN multicomponent crystals is summarized in Table A2-2. All anhydrous cocrystals and salts followed the predicted synthons (synthon A, B and C) depicted in Table A2-1.a.

OLN–Phol–H$_2$O crystallizes in toluene solvent with the $P\bar{1}$ space group with one OLN molecule, one Phol molecule, and one water molecule in the asymmetric unit. One OLN molecule interacts with one Phol molecule by an O-H$\cdots$N hydrogen bond (Interaction I). The hydrogen bond donor of one water molecule, is connected to the piperazine N (hydrogen acceptor) O-H$\cdots$N) (Interaction II). The oxygen of the same water molecule, as a hydrogen bond acceptor, interacts with the diazepine through the N-H$\cdots$O (Interaction III). Another hydrogen of the same water molecule is connected to the Phol molecule through an O-H$\cdots$O hydrogen bond (Interaction IV). Two water molecules and two OLN molecules form a ring (see Table A2-2 and Figure A2-2).

OLN-SA crystallizes in ethanol solvent with the $P\bar{1}$ space group, and it consists of one olanzapinium and one salicylate ion in the asymmetric unit. An acidic proton from SA is transferred to the piperazine N atom of OLN, forming a salt through the N$^+$-H$\cdots$O$^-$ charge-

assisted hydrogen bond (Interaction I). The second O of the SA carboxyl group interacts with the diazepine N of the OLN via N-H···O⁻ (Interaction II) (see Table A2-2 and Figure A2-2).

OLN-AA crystallizes in dichloromethane with the $P2_1/c$ space group. It includes one olanzapinium and one anthranilate ion in the asymmetric unit. Similar to the OLN-SA, an acidic proton from AA is transferred to the piperazine N of OLN, forming the salt through the N^+-H···O⁻ charge-assisted hydrogen bond (Interaction I). The second O of the AA carboxyl group interacts with the diazepine N of the OLN via N-H···O⁻ (Interaction II) (see Table A2-2 and Figure A2-2).

OLN-3HBA crystallizes in acetonitrile with the $P\bar{1}$ space group, and it includes one olanzapinium and one 3-hydroxybenzoate ion in the asymmetric unit. The acidic proton from 3HBA carboxyl is transferred to the piperazine N of OLN, forming the salt through the N^+-H···O⁻ charge-assisted hydrogen bond (Interaction I). The second O of the 3HBA carboxyl group is connected to the diazepine N of OLN through the N-H···O⁻ hydrogen bond (Interaction II). The phenolic hydroxyl of 3HBA is connected to the diazepine N of another olanzapinium ion through the hydrogen bond O-H···N (Interaction III). Two 3HBA ions and two OLN ions form a ring (see Table A2-2 and Figure A2-2).

OLN-TA-H₂O crystallizes in methanol with $P2_1/n$ space group, and it constitutes one olanzapinium, half terephthalate ion, and two water molecules in the asymmetric unit. Two acidic protons from the two carboxyl groups of TA are transferred to two piperazine N of OLN, forming a 2:1 molar ratio salt through the N^+-H···O⁻ charge-assisted hydrogen bond (Interaction I). The second O of the TA carboxyl group from both sides is connected to two diazepine N of OLN through the N-H···O⁻ hydrogen bond (Interaction II). The first O of the TA carboxyl is also connected to a hydrogen atom of a water molecule as a hydrogen bond acceptor via the O-H···O⁻ hydrogen bond (Interaction III). Another hydrogen atom from this water molecule forms a different hydrogen bond (O-H4···O) (Interaction IV) with the second water molecule. The second water, which also interacts with the diazepine N of the OLN via O-H···N hydrogen bond (Interaction V) is connected to another pair of water molecules through the O-H···O hydrogen bond (Interaction VI) to form an 8-member ring as shown in the packing diagram Figure A2-2.e.

Table A2-2 List of intermolecular interactions of OLN multicomponent crystals

Interaction	D-X···A	Symmetry Code	X···A (Å)	D···A (Å)	∠D-X···A (°)
		OLN–Phol–H₂O			
I	O1-H1···N1	x, y, z	1.87	2.74	174
II	O2-H2C···N4	1-x, 2-y, 2-z	1.80	2.83	162
III	N2-H2···O2	x,-1+y, z	2.04	2.88	162
IV	O2-H2B···O1	x, y, z	2.12	2.97	169
		OLN-SA			
I	N4⁺-H4N···O1⁻	-1-x,-y, -z	1.64	2.66	166
II	N2-H2N···O2⁻	-1-x,-y, 1-z	2.04	2.92	171
		OLN-AA			
I	N4⁺-H4A···O1⁻	x, ½-y, ½+z	1.64	2.65	172
II	N2-H2A···O2⁻	-x,-y,-z	2.12	2.96	168
		OLN-3HBA			
I	N4⁺-H4···O1⁻	x, y, z	1.47	2.56	176
II	N2-H2···O2⁻	x, y,-1+z	2.07	2.95	169
III	O3-H3···N1	2-x, 1-y, 2-z	1.97	2.78	165
		OLN-TA-H₂O			
I	N4⁺-H4···O1⁻	1.5-x, ½+y, 1.5-z	1.77	2.68	165
II	N2-H2···O2⁻	½+x, ½-y, ½+z	1.95	2.85	172
III	O4-H4C···O1⁻	1.5-x, ½+y, 1.5-z	2.00	2.82	156
IV	O4-H4B···O3	1-x, 1-y, 1-z	1.97	2.80	159
V	O3-H3A···N1	x, y, z	2.24	3.15	159
VI	O3-H3B···O4	x, y, z	1.87	2.72	165

C
H
O
N
S
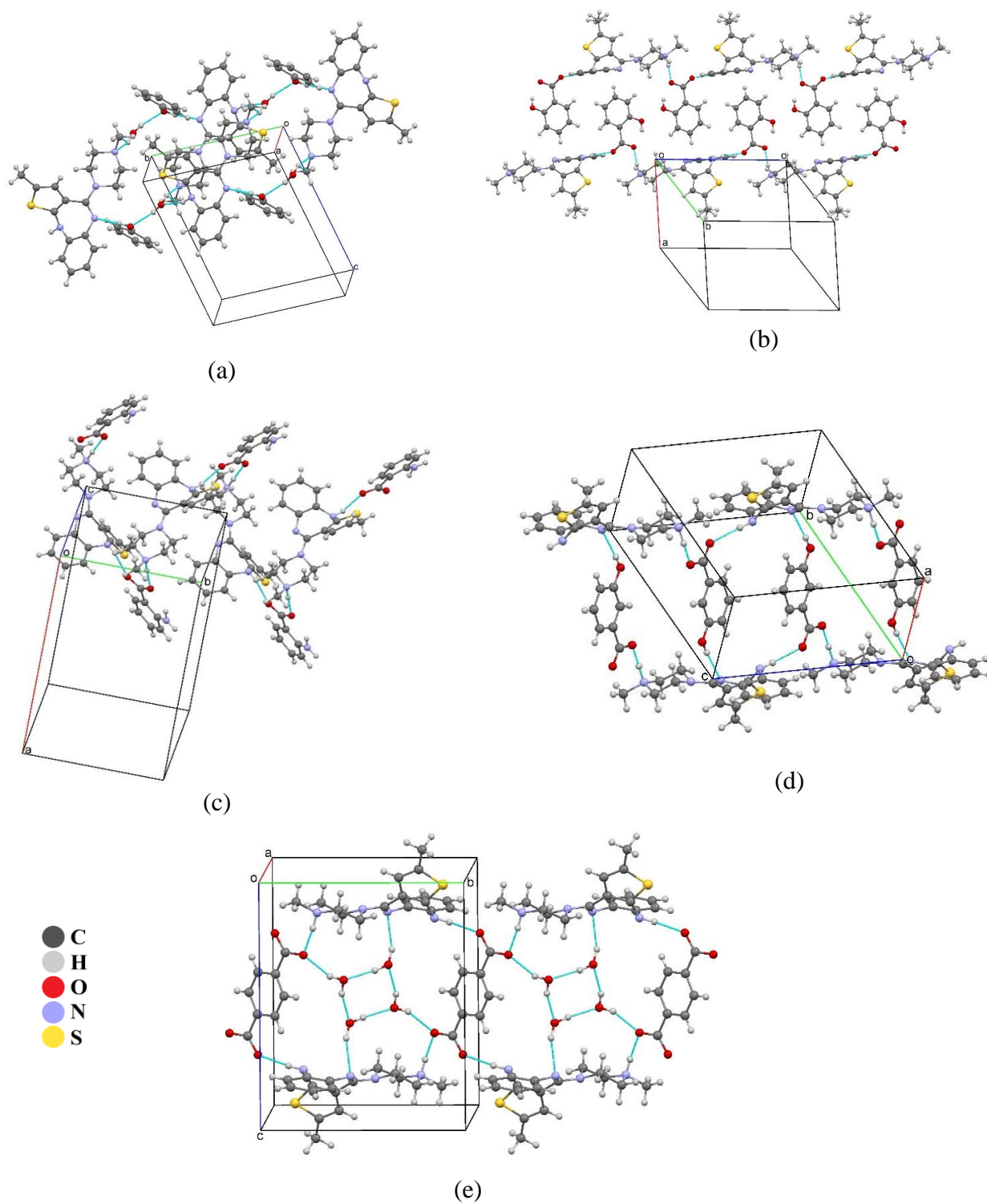
Figure A2-2 Crystal packing diagrams of (a) OLN-Phol-H_2O, (b) OLN-SA, (c) OLN-AA, (d) OLN-3HBA, and (e) OLN-TA-H_2O.

A2.3.3 Thermal Analysis

The DSC curves of OLN and its multicomponent crystals are shown in Figure A2-3 (split into two figures to reduce crowdedness). It was confirmed that each of the cocrystals and salts reported in this article displays a different onset value of the melting peak, compared to the pure OLN (onset value: 195°C) and the corresponding salt-formers/coformers. Hydrates and solvates formed (OLN-MeOH, OLN-3HBA-ACN, OLN-TA-H_2O) all show the loss of their corresponding solvents during DSC analysis. OLN-MeOH shows a broad endothermic peak ranging from 95°C to 133°C, suggesting the methanol evaporation, followed by another endothermic melting peak at 193°C, close to the melting point of pure OLN. The DSC curve of OLN-3HBA acetonitrile solvate indicates a similar melting point to OLN-3HBA at 202°C, following a broad endothermic peak starting at 100°C due to the loss of acetonitrile. The dihydrate of OLN-TA [OLN-TA-H_2O] shows a melting point at 260°C close to that of OLN-TA, following the broad endothermic peak at 73°C due to the loss of water. OLN-2ATPA shows an endothermic melting point of 224°C, different from both OLN and 2ATPA.
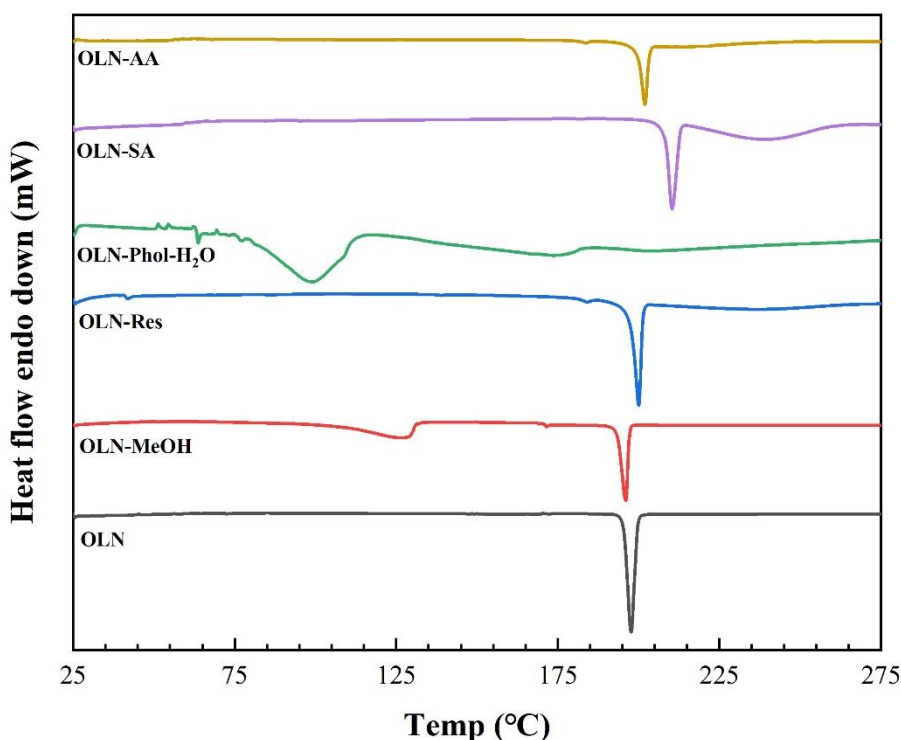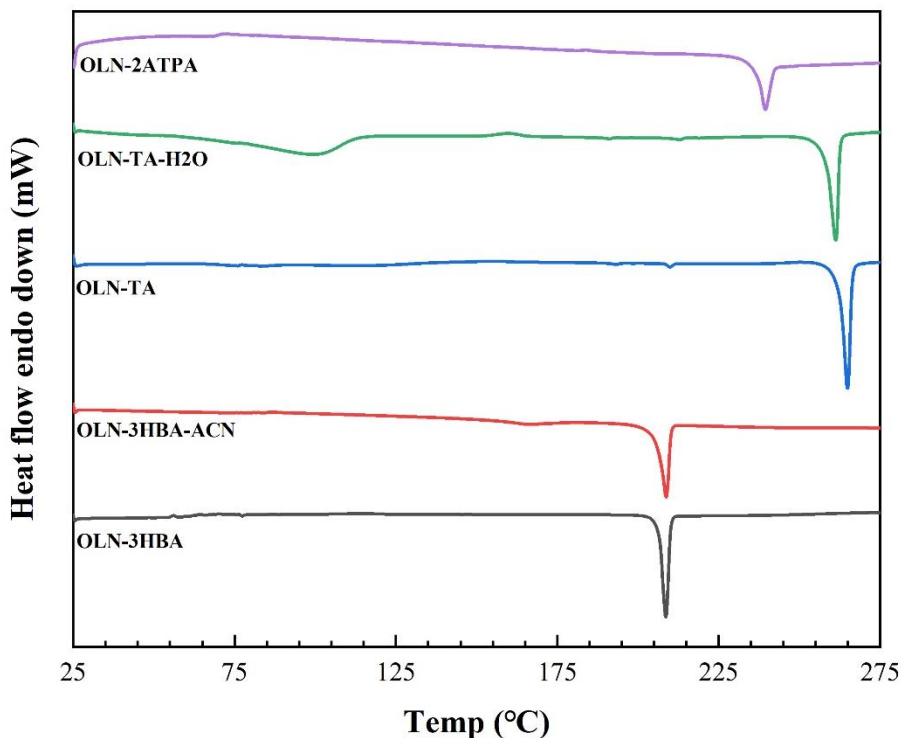
Figure A2-3 Thermographs of fresh OLN multicomponent crystals.

A2.3.4 Comparison of the PXRD Patterns

All ground bulk crystals exhibit unique PXRD patterns compared to OLN and their corresponding coformers/salt-formers, indicating the formation of new crystalline phases, as shown in Figure A2-4 (split into three figures to reduce crowdedness). Pure OLN and its methanol solvate were run as references. It was clearly shown that some of the main peaks at 8.6°, 17.0°, 17.8°, 18.7°, 19.8°, 21.0°, 22.2°, 23.9°, and 25.2° of pure OLN, and some of the main peaks at 16.3°, 22.9°, 24.6° of OLN methanolate disappeared in the new multicomponent crystal x-ray diffraction patterns.

The main peaks at 7.6°, 13.2°, 19.2°, are new in OLN-Phol-H_2O, compared to pure OLN.

Peaks at 19.4°, 23.7°, 26.6°, 28.3° of pure 3HBA disappeared in OLN-3HBA and OLN-3HBA-ACN. Instead, peaks at 13.6°, 16.7°, 18.2°, 20.4° in OLN-3HBA; and 12.7°, 18.0°, 22.5° in OLN-3HBA-ACN were detected.

Peaks at 17.4°, and 35.2° in TA disappeared, with new peaks appearing at 6.7°, 13.2°, 14.1°, 15.6°, 19.3°, 21.5° and 23.2° in OLN-TA. The PXRD pattern of OLN-TA-H_2O is similar to the OLN-TA anhydrate, with peaks at 6.7° and 23.2° disappearing and new peaks appearing at 9.8° and 20.0°.

Peaks at 13.8°, 16.4°, 18.6°, 20.2°, 27.6° in AA; 19.2°, 20.0° in Res; and 11.0°, 15.3° in SA disappeared, while new sharp peaks at 6.1°, 12.0°, 19.4° in OLN-AA; 13.0°, 18.5°, 21.7°, 24.4° in OLN-Res; and 8.0°, 13.7°, 19.1° in OLN-SA appeared. The characterized excess crystals after dissolution and solubility tests are the same as the PXRD patterns of anhydrates of OLN-AA, OLN-Res, OLN-SA, indicating no hydrates or solvates were formed in these multi-component crystals.

Peaks of pure 2ATPA at 15.0° and 27.1° disappeared in the PXRD patterns of OLN-2ATPA, while new peaks at 9.6°, 19.4°, 21.7°, 23.5° appeared.
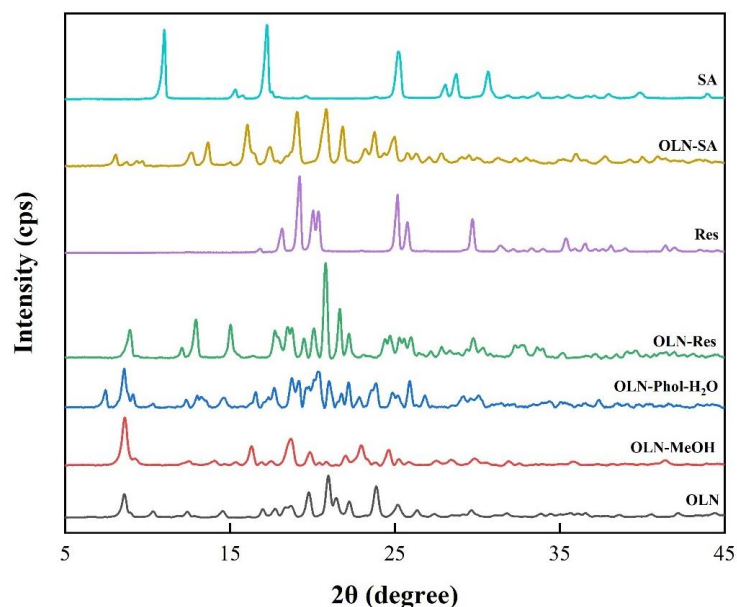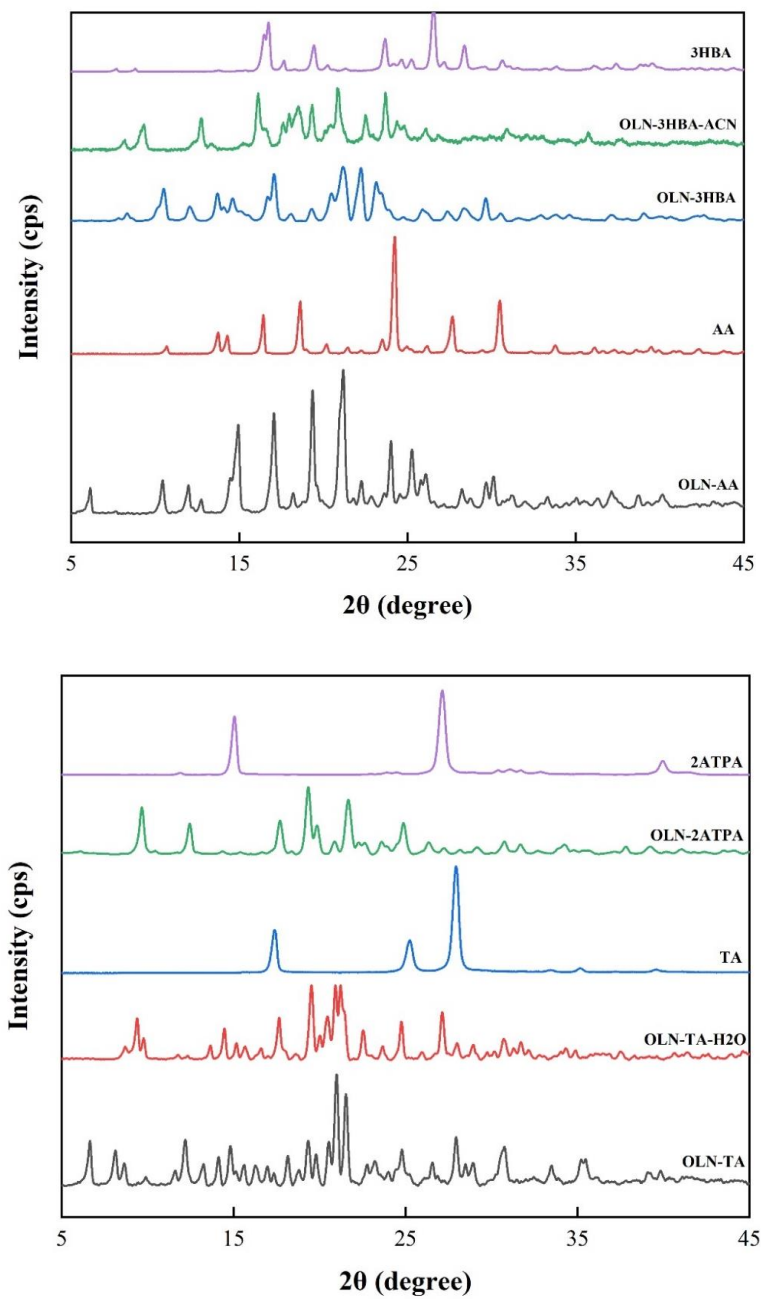
Figure A2-4 PXRD patterns of OLN and OLN multicomponent crystals.

A2.3.5 Solubility and Dissolution Test

In this research, we used the same solid phase of OLN-Res confirmed by SCXRD to test its solubility and dissolution. In addition, due to phenol's toxicity, only the single crystal structure of OLN-Phol-H$_2$O was examined and reported in this paper. The solubility of all other new multicomponent crystals is considered in this study.

The retention time of OLN was around 9 min. The linear correlation between area and concentration was found with a slope of 2575600 peak area/(mg/100ml). The R-squared of this linear regression was 0.9986, indicating high degree of correlation between the peak area and the OLN concertation.

The measured equilibrium solubility data of OLN and multicomponent crystals in the pH 7.2 phosphate buffer solution are presented in Table A2-3. It is clearly shown that, except OLN-Res, all the new-found OLN cocrystals/salts have improved solubility compared to pure OLN.

OLN-2ATPA displays the equilibrium solubility of 1.15±(0.10) mg/ml, which is almost five times the original OLN solubility. OLN-3HBA, OLN-AA have double the solubility, and the solubilities of OLN-SA and OLN-TA are 1.5 times of the pure OLN.

The left-over solid phases after the solubility and dissolution tests, were checked by PXRD and DSC. No phase transformation was noted for OLN-Res, OLN-SA, and OLN-AA samples. On the contrary, a phase change occurred during the test of OLN-2ATPA, OLN-TA, OLN-3HBA, and OLN-3HB-ACN. The PXRD patterns of the excess OLN-2ATPA crystals remaining in the solution indicated that part of the crystals formed OLN dihydrate (Reutzel-Edens et al., 2003) during the solubility and dissolution tests. The PXRD pattern is an overlap of both OLN dihydrate and OLN-2ATPA. The DSC pattern confirmed the formation of OLN dihydrate. The DSC data shows a broad endothermic peak starting from 92°C indicating the loss of water in the OLN dihydrate, following a small OLN melting peak and an OLN-2ATPA melting peak. Similarly, OLN dihydrate also formed during the solubility tests of OLN-3HBA and OLN-3HBA-ACN. From the PXRD analysis of the remaining solid samples after the solubility tests of OLN-3HBA and OLN-3HBA-ACN, a mixture of OLN dihydrate and OLN-3HBA was observed in both cases. The DSC curve of the remaining solid samples after the

solubility tests after filtration, indicated the melting peak of OLN-3HBA. The crystals were then washed with deionized water and characterized by DSC again. The DSC data showed a broad endothermic peak starting from 93℃ indicating the loss of water in the OLN dihydrate, followed by an OLN melting point and a small OLN-3HBA melting point, confirming the presence of the OLN dihydrate.

The PXRD analysis of the remaining solid samples after the solubility tests of both OLN-TA and OLN-TA-H$_2$O showed the same patterns as the OLN-TA-H$_2$O, indicating the OLN-TA-H$_2$O was formed during the solubility and dissolution tests of anhydrous OLN-TA.

As shown in Figure A2-5, most of the multicomponent crystals except OLN-Res improved the OLN dissolution at pH 7.2. The concentration of pure OLN and OLN-TA gradually decreased in the first two hours and remained stable at nearly 0.24±(0.05) mg/ml and 0.38±(0.05) mg/ml, respectively. This might be attributed to the fact that during the dissolution tests, hydrates of OLN and OLN-TA were gradually formed. Compared to the hydrates, anhydrous crystals of OLN and OLN-TA are easier to dissolve in water. Therefore, the dissolution was high initially; with the formation of the hydrate, the OLN concentration decreased and finally reached a plateau.

The major difference between the dissolution profiles of OLN-3HBA and OLN-3HBA-ACN was at t=30 min. The dissolution of OLN-3HBA-ACN peaked at 30 min before dropping back to the dissolution profile of OLN-3HBA at 1 hour. To explain this trend, excess crystals of OLN-3HBA and OLN-3HBA-ACN after dissolution tests were washed by the buffer solution and characterized by PXRD. A new diffraction pattern, different from OLN-3HBA and OLN-3HBA-ACN, emerged in both experiments. The new observed PXRD pattern was determined to be a mixture of anhydrous OLN-3HBA and Olanzapine dihydrate. Therefore, the peak in concentration in the dissolution test of OLN-3HBA-ACN at t=30 min can be associated with the loss of acetonitrile.

The OLN-2ATPA showed (Figure A2-5) a continuous decrease in the dissolution during the whole four hours, probably due to the partial dissociation of the salt and the formation of OLN

hydrate. The formed OLN hydrate displays a lower solubility compared to the OLN-2ATPA salt. Thus, a decrease in the dissolution was observed during the experiment.

The pH of the buffer solution was measured before and after the equilibrium solubility tests to determine the influence of the ionizing components. As shown in Table A2-3, it was observed that the pH of the buffer medium varied differently after the 24-hour equilibrium solubility tests. The pH of the buffer solution did not change during the equilibrium solubility study of pure OLN and OLN-Res at 37°C due to the similar pH value of the OLN aqueous solution. As OLN is a weak base, the salts of OLN reported here all belong to acidic salts. They all show a significant decrease in pH as the hydrolysis of olanzapinium will produce hydrogen ions. As the solubility increased, more olanzapinium ions were formed in the buffer solution. Thus, the pH of the buffer medium from the OLN-2ATPA solubility test was the lowest, as OLN-2ATPA displays the highest equilibrium solubility (1.149 mg/ml).

Table A2-3 Solubility of OLN cocrystals/salts/solvates.

Name	OLN Solubility (mg/ml)	pH before/after the equilibrium solubility experiment
OLN	0.24±(0.05)	7.2/7.2
OLN-Res	0.19±(0.05)	7.2/7.1
OLN-SA	0.37±(0.05)	7.2/7.0
OLN-AA	0.61±(0.05)	7.2/6.8
OLN-3HBA-ACN	0.69±(0.05)	7.2/6.7
OLN-TA	0.32±(0.05)	7.2/6.9
OLN-TA-H$_2$O	0.31±(0.05)	7.2/7.0
OLN-3HBA	0.71±(0.05)	7.2/6.9
OLN-2ATPA	1.15±(0.10)	7.2/6.2
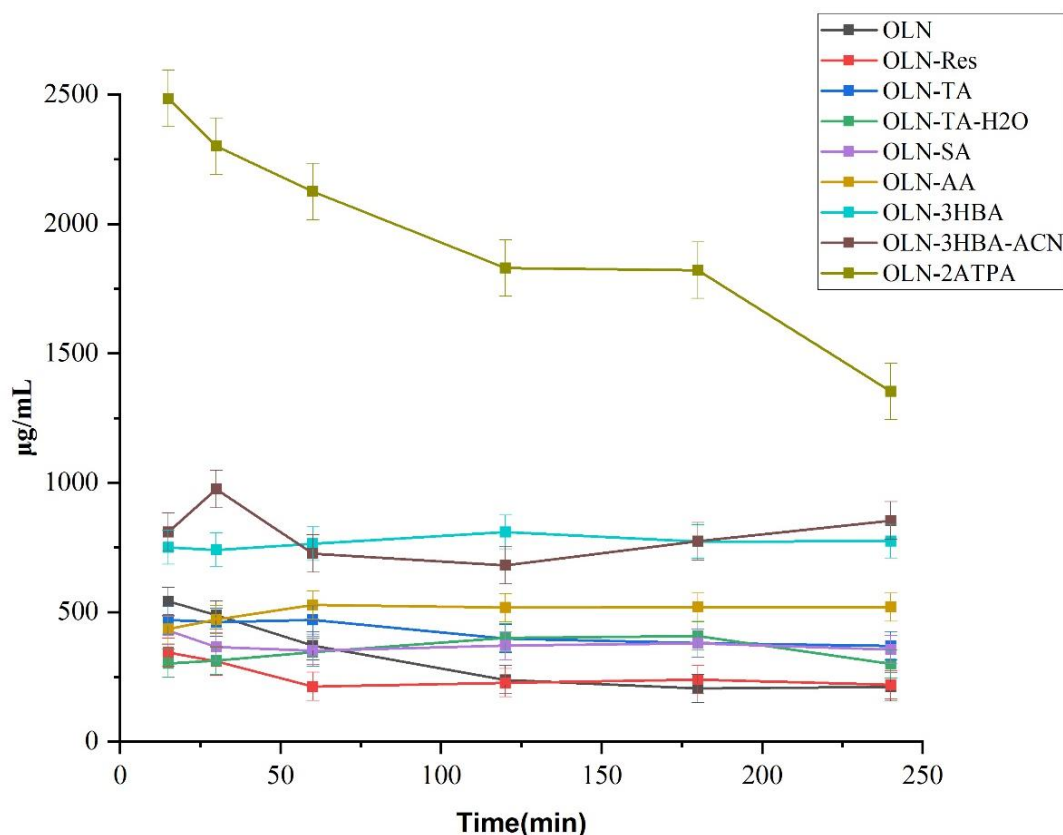
Figure A2-5 Dissolution measurement of OLN cocrystals/salts/solvates.

A2.4 Conclusion

Based on the prediction using hydrogen bond propensity and coordination (HBP and HBC) calculations, multicomponent crystals of OLN with coformers (Phol, Res) and salt-formers (SA, AA, 3HBA, 2ATPA) were successfully obtained through liquid-assisted grinding and ball millings and were characterized through PXRD and DSC. All achieved cocrystals and salts followed the "ΔpK_a rule of 3". Except for OLN-2ATPA, single crystals of the discovered OLN multicomponent solid-phases were obtained, and the salt formations/cocrystal formations were confirmed with their crystal structure data. The crystal packing analyses were conducted to shed light on the primary intermolecular interactions. Olanzapinium salts showed similar charge-assisted bonds and hydrogen bonds in their structures. The acidic proton of carboxylic acid transferred to the piperidine N, and the carbonyl group also interacted with the diazepine N of OLN as a hydrogen acceptor.

The solubility and dissolution tests of these new OLN multicomponent crystals were obtained by HPLC. It was found that all salts show significant improvements in the solubility; OLN-2ATPA showed almost a fivefold increase, and OLN-AA, OLN-3HBA showed a threefold increase in the solubility. This was reflected by the pH value of the solution after tests as a result of the hydrolysis of the olanzapinium ion. No significant variation in the solubility was found in the OLN-Res cocrystal.

A2.5 References

Almarsson, Örn, & 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*, *17*, 1889.

Andrusenko, I., Potticary, J., Hall, S. R., & Gemmi, M. (2020). A new olanzapine cocrystal obtained from volatile deep eutectic solvents and determined by 3D electron diffraction. *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials*, *76*(6), 1036–1044.

Basavaiah, K., Rajendraprasad, N., & Vinay, K. B. (2014). Isocratic High-Performance Liquid Chromatographic Assay of Olanzapine: Method Development and Validation. *ISRN Analytical Chemistry*, *2014*, 1–6.

Blagden, N., de Matas, M., Gavan, P. T., & York, P. (2007). Crystal engineering of active pharmaceutical ingredients to improve solubility and dissolution rates. *Advanced Drug Delivery Reviews*, *59*(7), 617–630.

Burla, M. C., Caliandro, R., Carrozzini, B., Cascarano, G. L., Cuocci, C., Giacovazzo, C., Mallamo, M., Mazzone, A., & Polidori, G. (2015). Crystal structure determination and refinement via SIR2014. *Journal of Applied Crystallography*, *48*(1), Article 1.

Bymaster, F. P., Calligaro, D. O., Falcone, J. F., Marsh, R. D., Moore, N. A., Tye, N. C., Seeman, P., & Wong, D. T. (1996). Radioreceptor Binding Profile of the Atypical Antipsychotic Olanzapine. *Neuropsychopharmacology*, *14*(2), 87–96.

Cavallari, C., Santos, B. P.-A., & Fini, A. (2013). Olanzapine Solvates. *Journal of Pharmaceutical Sciences*, *102*(11), 4046–4056.

Clarke, H. D., Hickey, M. B., Moulton, B., Perman, J. A., Peterson, M. L., Wojtas, Ł., Almarsson, Ö., & Zaworotko, M. J. (2012). Crystal Engineering of Isostructural Quaternary Multicomponent Crystal Forms of Olanzapine. *Crystal Growth & Design*, *12*(8), 4194–4201.

Dalpiaz, A., Ferretti, V., Bertolasi, V., Pavan, B., Monari, A., & Pastore, M. (2018). From Physical Mixtures to Co-Crystals: How the Coformers Can Modify Solubility and Biological Activity of Carbamazepine. *Molecular Pharmaceutics*, *15*(1), 268–278.

Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K., & Puschmann, H. (2009). OLEX2: A complete structure solution, refinement and analysis program. *Journal of Applied Crystallography*, *42*(2), 339–341.

Duggirala, N. K., Perry, M. L., Almarsson, Ö., & Zaworotko, M. J. (2016). Pharmaceutical cocrystals: Along the path to improved medicines. *Chemical Communications*, *52*(4), 640–655.

Farrugia, L. J. (2012). WinGX and ORTEP for Windows: An update. *Journal of Applied Crystallography*, *45*(4), 849–854.

Huang, Y., Kuminek, G., Roy, L., Cavanagh, K. L., Yin, Q., & Rodríguez-Hornedo, N. (2019). Cocrystal Solubility Advantage Diagrams as a Means to Control Dissolution, Supersaturation, and Precipitation. *Molecular Pharmaceutics*, *16*(9), 3887–3895.

Karimi-Jafari, M., Padrela, L., Walker, G. M., & Croker, D. M. (2018). Creating Cocrystals: A Review of Pharmaceutical Cocrystal Preparation Routes and Applications. *Crystal Growth & Design*, *18*(10), 6370–6387.

Lemmerer, A., Govindraju, S., Johnston, M., Motloung, X., & Savig, K. L. (2015). Co-crystals and molecular salts of carboxylic acid/pyridine complexes: Can calculated pK_a's predict proton transfer? A case study of nine complexes. *CrystEngComm*, *17*(19), 3591–3595.

Macrae, C. F., Sovago, I., Cottrell, S. J., Galek, P. T. A., McCabe, P., Pidcock, E., Platings, M., Shields, G. P., Stevens, J. S., Towler, M., & Wood, P. A. (2020). *Mercury 4.0*: From visualization to analysis, design and prediction. *Journal of Applied Crystallography*, *53*(1), 226–235.

Nanubolu, J. B., & Ravikumar, K. (2017). Designing a new cocrystal of olanzapine drug and observation of concomitant polymorphism in a ternary cocrystal system. *CrystEngComm*, *19*(2), 355–366.

Nechipadappu, S. K., Reddy, I. R., Tarafder, K., & Trivedi, D. R. (2019). Salt/Cocrystal of Anti-Fibrinolytic Hemostatic Drug Tranexamic acid: Structural, DFT, and Stability Study of Salt/Cocrystal with GRAS Molecules. *Crystal Growth & Design*, *19*(1), 347–361.

Reddy, B. V., Reddy, K. V. N. S., Sreeramulu, J., & Kanumula, G. V. (2007). Simultaneous Determination of Olanzapine and Fluoxetine by HPLC. *Chromatographia*, *66*(1–2), 111–114.

Reddy, L. S., Bethune, S. J., Kampf, J. W., & Rodríguez-Hornedo, N. (2009). Cocrystals and Salts of Gabapentin: PH Dependent Cocrystal Stability and Solubility. *Crystal Growth & Design*, *9*(1), 378–385.

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*(6), 897–907.

Rodrigues, M., Baptista, B., Lopes, J. A., & Sarraguça, M. C. (2018). Pharmaceutical cocrystallization techniques. Advances and challenges. *International Journal of Pharmaceutics*, *547*(1–2), 404–420.

Sandhu, B., McLean, A., Sinha, A. S., Desper, J., Sarjeant, A. A., Vyas, S., Reutzel-Edens, S. M., & Aakeröy, C. B. (2018). Evaluating Competing Intermolecular Interactions through Molecular Electrostatic Potentials and Hydrogen-Bond Propensities. *Crystal Growth & Design*, *18*(1), 466–478.

Sarmah, K. K., Nath, N., Rao, D. R., & Thakuria, R. (2020). Mechanochemical synthesis of

drug–drug and drug–nutraceutical multicomponent solids of olanzapine. *CrystEngComm*, *22*(6), 1120–1130.

Sarmah, K. K., Sarma, A., Roy, K., Rao, D. R., & Thakuria, R. (2016). Olanzapine Salts and Diversity in Molecular Packing. *Crystal Growth & Design*, *16*(2), 1047–1055.

Sarmah, K. K., Sarma, P., Rao, D. R., Gupta, P., Nath, N. K., Arhangelskis, M., & Thakuria, R. (2018a). Mechanochemical Synthesis of Olanzapine Salts and Their Hydration Stability Study Using Powder X-ray Diffraction. *Crystal Growth & Design*, *18*(4), 2138–2150.

Sarmah, K. K., Sarma, P., Rao, D. R., Gupta, P., Nath, N. K., Arhangelskis, M., & Thakuria, R. (2018b). Mechanochemical Synthesis of Olanzapine Salts and Their Hydration Stability Study Using Powder X-ray Diffraction. *Crystal Growth & Design*, *18*(4), 2138–2150.

Sheldrick, G. M. (2015). Crystal structure refinement with SHELXL. *Acta Crystallographica Section C: Structural Chemistry*, *71*(1), Article 1.

Surampudi, A. V. S. D., Rajendrakumar, S., Nanubolu, J. B., Balasubramanian, S., Surov, A. O., Voronin, A. P., & Perlovich, G. L. (2020). Influence of crystal packing on the thermal properties of cocrystals and cocrystal solvates of olanzapine: Insights from computations. *CrystEngComm*, *22*(39), 6536–6558.

Thakuria, R., & Nangia, A. (2013). Olanzapinium Salts, Isostructural Solvates, and Their Physicochemical Properties. *Crystal Growth & Design*, *13*(8), 3672–3680.

The Cambridge Structural Database (CSD) - The Cambridge Crystallographic Data Centre (CCDC). (n.d.). Retrieved March 5, 2021, from https://www.ccdc.cam.ac.uk/solutions/csd-core/components/csd/.

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*(3), 865–872.

Tsume, Y., Mudie, D. M., Langguth, P., Amidon, G. E., & Amidon, G. L. (2014). The

Biopharmaceutics Classification System: Subclasses for in vivo predictive dissolution (IPD) methodology and IVIVC. *European Journal of Pharmaceutical Sciences*, *57*, 152–163.

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*(1–3), 188–197.

Weyna, D. R., Shattock, T., Vishweshwar, P., & Zaworotko, M. J. (2009). Synthesis and Structural Characterization of Cocrystals and Pharmaceutical Cocrystals: Mechanochemistry vs Slow Evaporation from Solution. *Crystal Growth & Design*, *9*(2), 1106–1123.

Appendix 3

Supporting Information for Chapter 3

Overcoming the Hydrophobic Nature of

Zinc Phenylacetate through

Cocrystallization with Isonicotinamide

A3 Supporting Information for Chapter 3
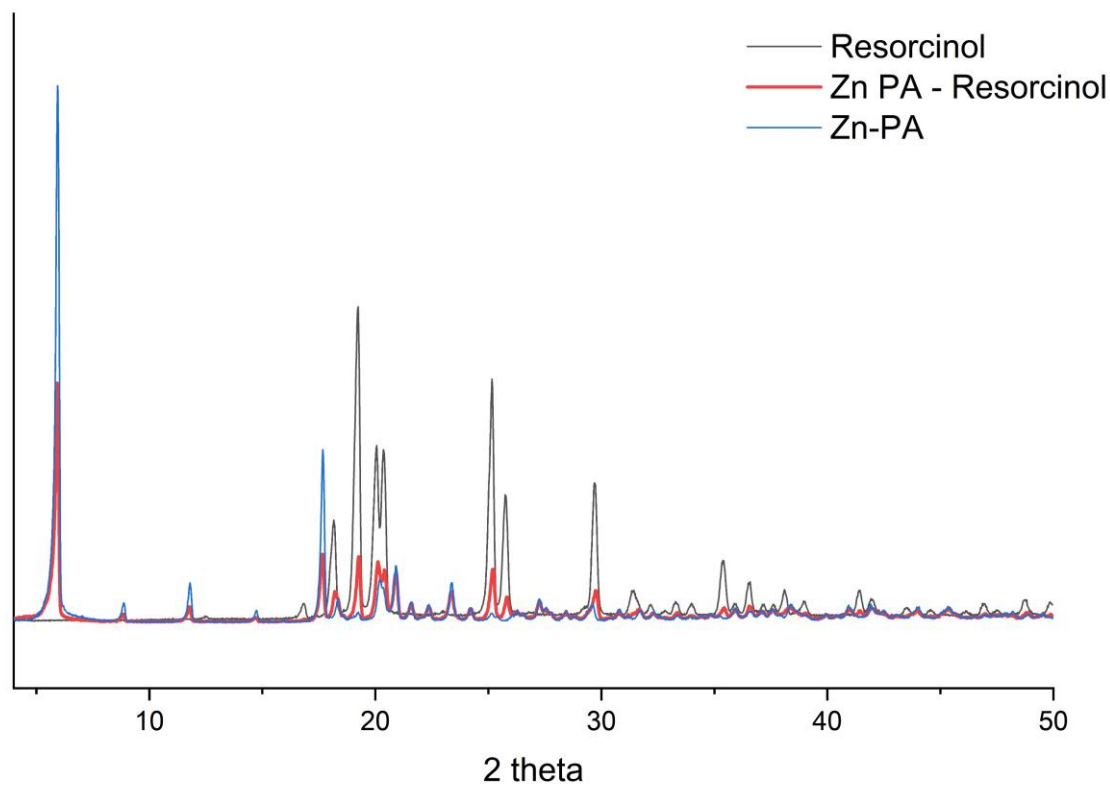
Figure A3-1 PXRD analysis of unsuccessful co-crystallization attempt of Zn-PA with Resorcinol that ended up as a physical mixture. Both Resorcinol peaks (black) and Zn-PA peaks (blue) are present in the mixture (red) with no new peak being observed.
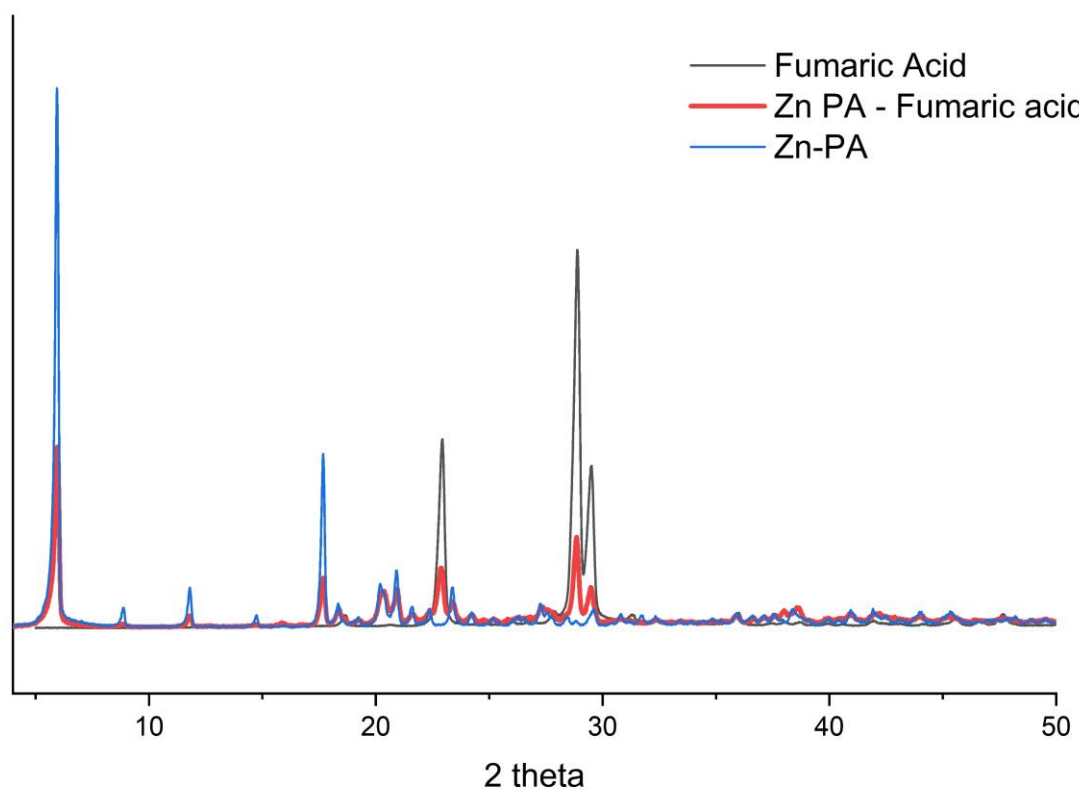
Figure A3-2 PXRD analysis of unsuccessful co-crystallization attempt of Zn-PA with Fumaric acid that ended up as a physical mixture. Both Fumaric acid peaks (black) and Zn-PA peaks (blue) are present in the mixture (red) with no new peak being observed.
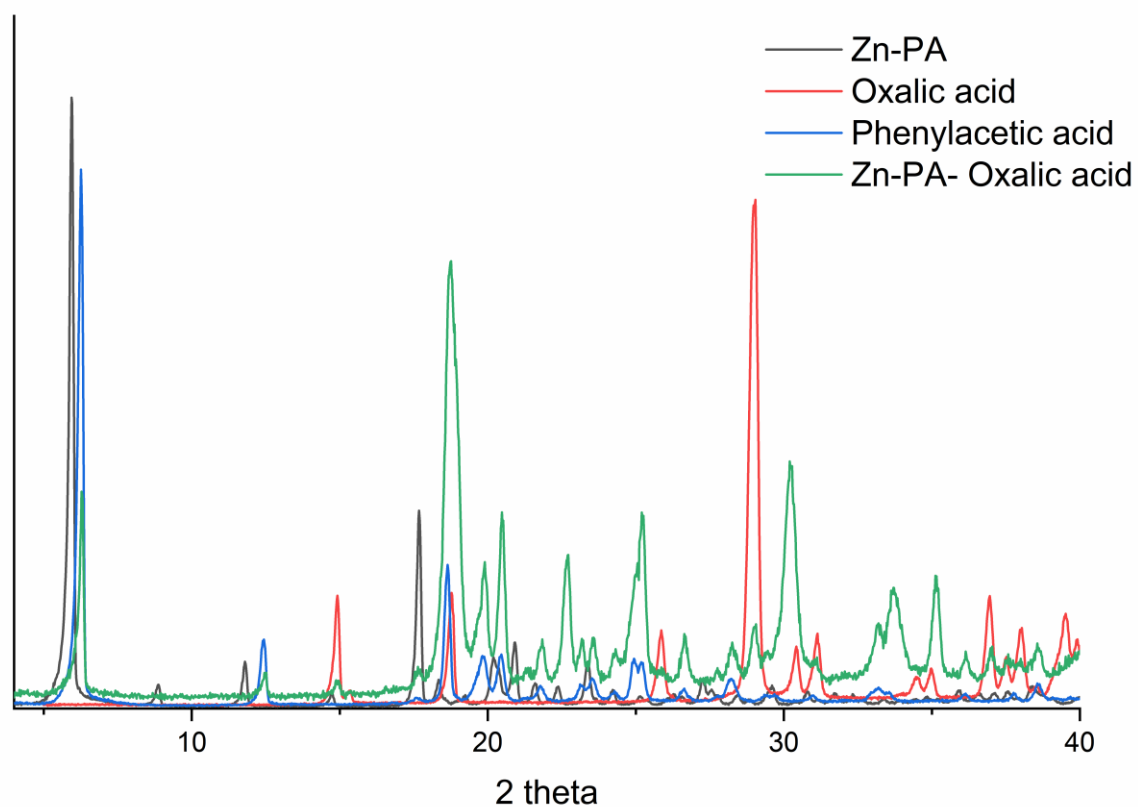
Figure A3-3 PXRD analysis of unsuccessful co-crystallization attempt of Zn-PA with Oxalic acid that ended up as a ligand substitution. The characteristic peaks of phenylacetic acid (blue) are observed in the mixture while Oxalic acid peaks have shrunken. Zn-PA peaks disappeared, and new peaks emerged (tentatively assigned to zinc oxalate).
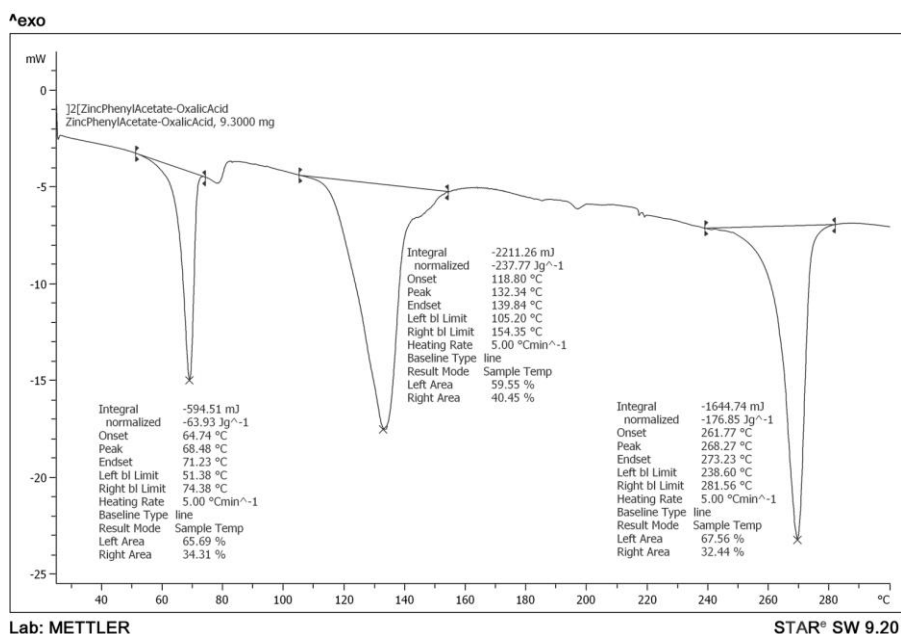
Figure A3-4 DSC analysis of the ground mixture of Zn-PA and Oxalic acid, showing multiple distinguished peaks, indicating a reaction has occurred and materials with new melting points are formed.
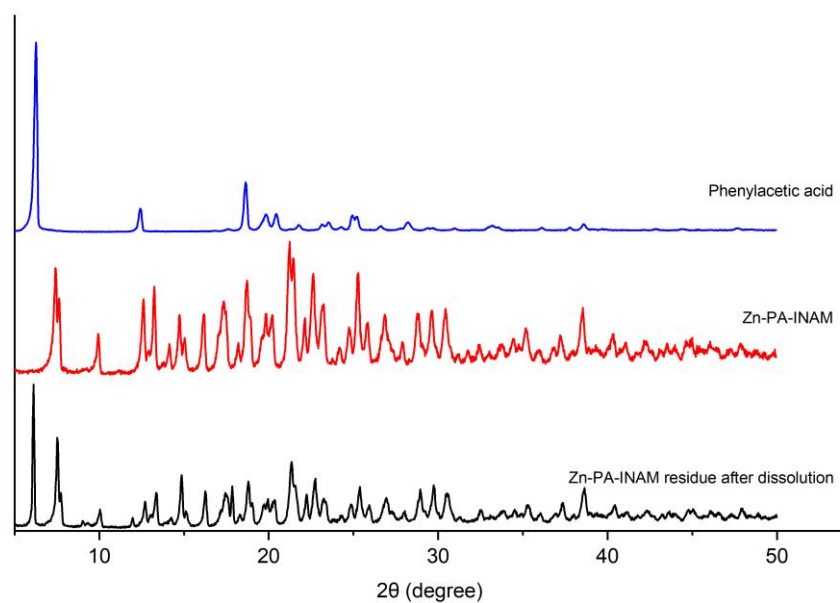
Figure A3-5 Comparison between the PXRD pattern of the residue powder after dissolution experiment of Zn-PA-INAM and the initial Zn-PA-INAM and Phenylacetic acid (PAH).
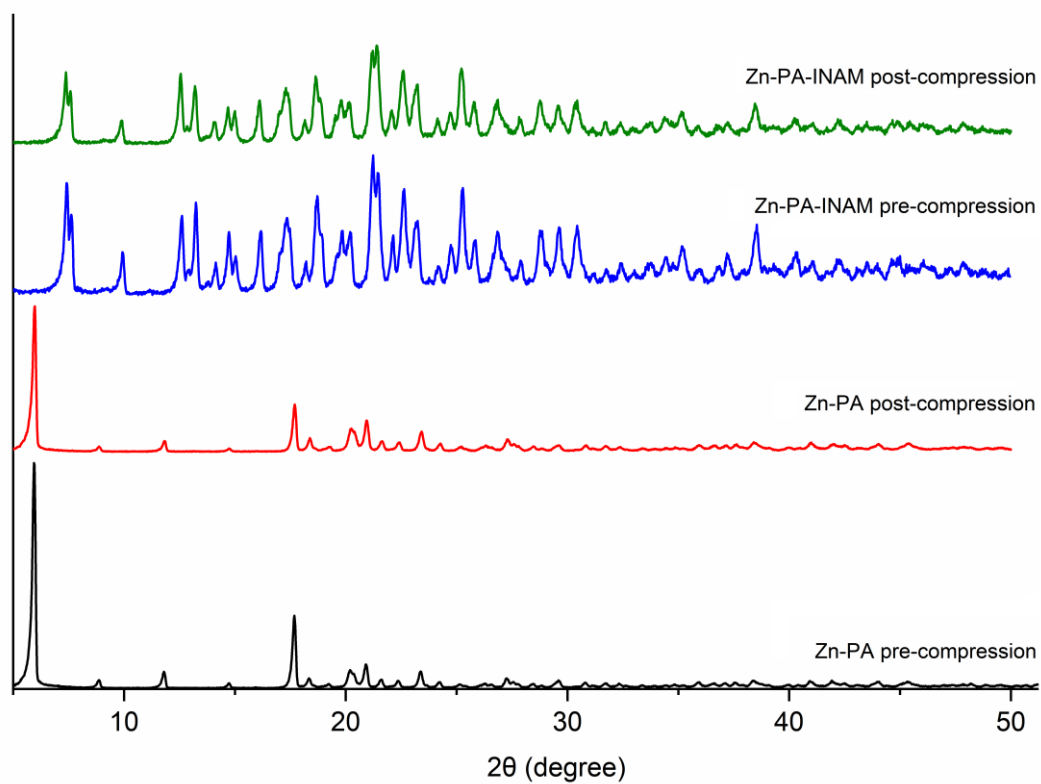
Figure A3-6 PXRD Comparison between the patterns of Zn-PA-INAM and Zn-PA, pre- and post-compression for tableting shows no structural changes were caused by the hydraulic press mechanical force.
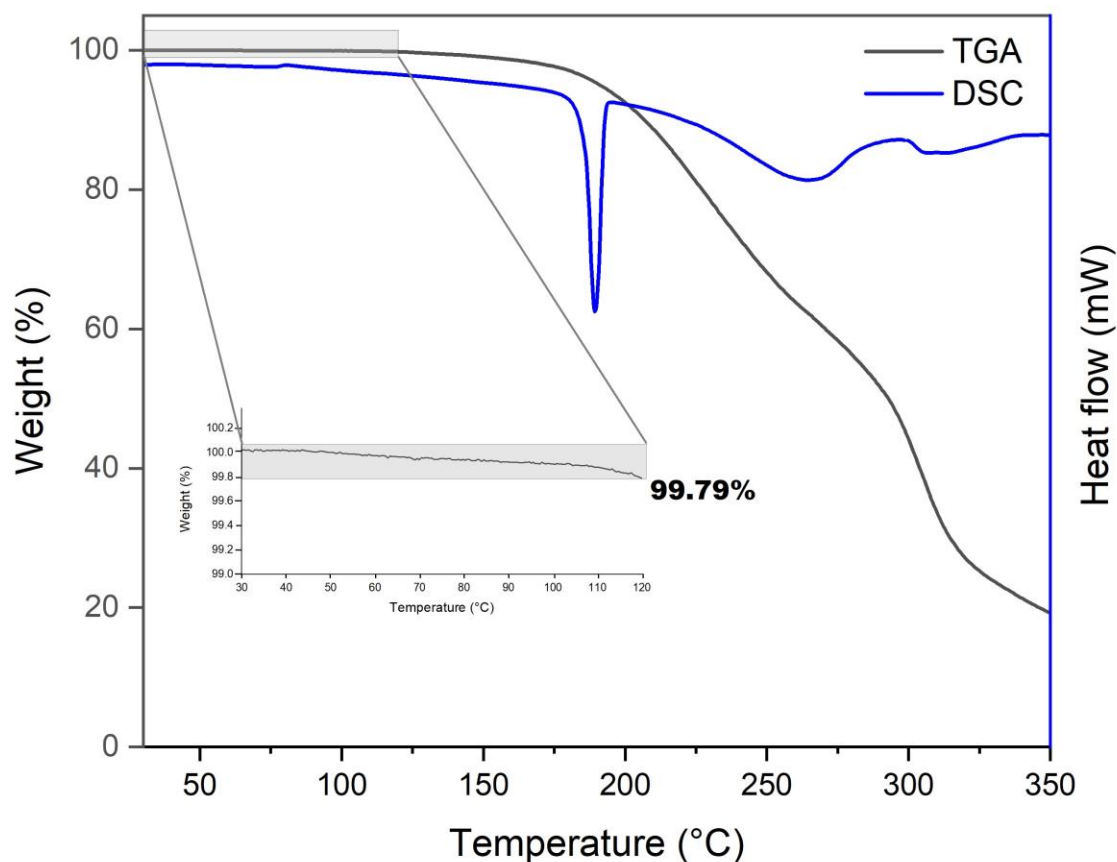
Figure A3-7 Thermogravimetric Analysis (TGA) and Differential Scanning Calorimetry (DSC) graphs for Zn-PA-INAM.

Table A3-1 Summary of onset/peak/endset temperatures, and enthalpy of melting (normalized integral of the peaks) of the main compounds related to the new crystal (Zn-PA-INAM) from DSC analysis.

Compound	Onset (°C)	Peak (°C)	Endset (°C)	Enthalpy of Melting (J/g)
PAH	75.57	76.76	79.39	-141.01
INAM	154.18	155.66	158.68	-206.76
Zn-PA	244.04	252.08	257.58	-149.15
Zn-PA-INAM	186.48	187.60	191.24	-143.45

Input data for Gaussian simulation of Zn-PA-INAM:

#MP2/6-31G** guess=core nosym density=MP2 pop=esp cube=cards cube=frozencore formcheck

#b22136xx 'P 21/c fragm 1

0 1
```
Zn   0.291116  -0.172057   0.555942
O    0.177645  -5.385618   5.167318
O    1.179779   2.571373   0.628981
O    3.110217   0.014902  -5.775386
O    1.039505   0.884771   2.056649
O   -1.342096   0.704655  -0.018181
O   -1.599339  -1.043886  -1.370061
N    0.115998  -1.914292   1.600500
N    1.444232  -0.344010  -1.092760
N   -1.667152  -5.987754   3.996207
N    4.726238  -1.245299  -4.827371
C   -0.346051  -4.111380   3.241246
C    1.298013   2.085775   1.769078
C    1.484893   0.682154  -1.949072
C    3.583342  -0.554052  -4.797437
C   -0.597399  -5.226069   4.214393
C   -0.614682  -4.170002   1.874353
C    2.848728  -1.592579  -2.589892
C    0.773160   4.037584   3.274245
C    2.859803  -0.524595  -3.473831
C    2.111695  -1.477851  -1.414179
C    2.158303   0.635828  -3.144526
C   -1.917865   0.079906  -0.973579
C   -0.365063  -3.061614   1.096619
C   -3.877714  -0.542389  -3.574715
C   -0.127761   3.805887   4.295979
C    0.741507   5.243935   2.592997
C   -3.283527   0.613032  -3.079229
C   -0.174882   6.228181   2.958198
C    0.190758  -2.934255   3.754859
C   -3.096237   0.827758  -1.601202
C   -2.843112   1.571044  -3.985507
C    0.386927  -1.868795   2.915125
C   -2.991046   1.378285  -5.357890
C    1.800406   2.967006   2.906085
C   -4.016853  -0.742533  -4.943727
C   -3.581119   0.213289  -5.830712
C   -1.059733   5.995770   3.979229
C   -1.047286   4.795944   4.650589
H    0.962982   1.590781  -1.687764
```

```
H  -1.012680  -5.071418   1.432176
H   3.400441  -2.494797  -2.808725
H   2.070937  -2.318889  -0.737733
H   2.145400   1.480889  -3.816848
H  -0.564680  -3.117596   0.036708
H  -4.235912  -1.294389  -2.887145
H  -0.121032   2.861291   4.819330
H   1.427720   5.421367   1.778192
H  -0.188065   7.173900   2.437000
H   0.449217  -2.861446   4.801036
H  -2.958951   1.884624  -1.426610
H  -4.000434   0.520781  -1.096566
H  -2.380420   2.476271  -3.621237
H   0.777126  -0.950999   3.329522
H  -2.648070   2.132085  -6.051103
H   2.723305   3.443082   2.609431
H   2.004471   2.353765   3.771294
H  -4.468743  -1.651948  -5.311371
H  -3.697448   0.058476  -6.893191
H  -1.770643   6.759558   4.257845
H  -1.748708   4.618772   5.452488
H  -1.883832  -6.739046   4.619485
H  -2.254005  -5.806421   3.207046
H   5.245784  -1.319417  -5.678618
H   5.062692  -1.690117  -3.997380
```

b22136xxa.den

```
  0  -7.799999  -9.400000  -9.320000
100   0.160000   0.000000   0.000000
120   0.000000   0.160000   0.000000
110   0.000000   0.000000   0.160000
```

Appendix 4

Supporting Information for Chapter 4

Virtual Multicomponent Crystal Screening: Hydrogen Bonding Revisited

A4 Supporting Information for Chapter 4

Table A4-1 Selected list of coformers for experimental screening of Olanzapine and Rufinamide

1,7-Phenanthroline	Benzoic acid	Phenanthridine
2-2'-Bipyridine	Catechol	Phenol
2-Aminoterephthalic acid	Citric acid	Resorcinol
2-Ethylimidazole	Dextrose	Saccharin
3-Aminobenzoic acid	Fumaric acid	Salicylic acid
3-Chloromandelic acid	Hydroquinone	Stearic acid
3-Hydroxybenzoic acid	Imidazole	Succinic acid
4,7-Phenanthroline	Isonicotinamide	Sulfamethazine
4-Aminobenzoic acid	Isonicotinic acid	Sulfanilic acid
4-Azabenzimidazole	Maleic acid	Tartaric acid
5-Fluorouracil	Malic acid	Terephthalic acid
9-Formylphenanthrene	Malonic acid	Thioacetamide
Acesulfame	Mandelic acid	Thiourea
Acetaminophen	Melamine	Trimesic acid
Adenine	Menthol	Urea
Anthranilic acid	Nicotinamide	Vanillin
Ascorbic acid	Oxalic acid	
Benzimidazole	Phenanthrenequinone	

Table A4-2 List of coformers used for ESP analysis.

(+)-camphoric acid	2-Amino-2-hydroxymethyl-propane-1,3-diol
1,10-phenanthroline	2-Amino-4,6-dimethylpyrimidine
1,2-Bis(4-pyridyl)acetylene	2-amino-5-nitropyridine
1,3,5-trinitrobenzene	2-aminobenzamide
1,3-Adamantanedicarboxylic acid	2-aminobenzoic acid
1,3-Di(4-pyridyl)propane	2-aminopyridine
1,3-Dicyanobenzene	2-aminopyrimidine
1,3-dimethylurea	2-aminoterephthalic acid
1,4-Dicyanobenzene	2-Carboxymethyl-benzoic acid
1,5-Diaminonaphthalene	2-Cresol
1,5-dihydroxynaphthalene	2-hydroxycaproic acid
1,5-Naphthalenedisulfonic acid	2-Hydroxymethylphenol
1,7-phenanthroline	2-methoxyphenol
14-cyclohexanedi1	2-Nitrobenzoic acid
1-Cyanonaphthalene	2-picolinamide
1-hydroxy-2-naphthoic acid	2-picoline
1-Hydroxynaphthalene	2-Picolinic acid
1-Phenylethylammonium	2-Piperidone
2,2'-bipyridine	2-pyridone
2,2-Dimethylsuccinic acid	2-pyrrolidinone
2,2'-ethene-1,2-diyl dipyridine	3-(4-Hydroxyphenyl)propionic acid
2,3,4,5-Tetrafluorobenzoic acid	3,3-tetramethyleneglutarimide
2,3,4-Trifluorobenzoic acid	3,4,5-trifluorobenzoic acid
2,3,5,6-tetramethyl pyrazine	3,4,5-trimethoxybenzoic acid
2,3-dihydroxybenzoic acid	3,4-diaminopyridine
2,4- dichlorophenoxyacetic acid	3,4-dichlorobenzoic acid
2,4, 6-trimethylpyridine	3,4-difluorobenzoic acid
2,4,5-Trichlorophenoxyacetic acid	3,4-dihydroxybenzoic acid
2,4,5-trifluorobenzoic acid	3,5-dihydroxybenzoate
2,4,6-trifluorobenzoic acid	3,5-dihydroxybenzoic acid
2,4,6-trimethylbenzoic acid	3,5-dimethylpyrazole
2,4,6-Trinitrobenzoic acid	3,5-dinitrobenzoic acid
2,4-Dichloro-5-fluorophenyl-acetic acid	3,5-dinitrosalicylic acid
2,4-dihydroxybenzoicacid	3-Amino-1,2,4-triazole
2,4-dinitrophenol	3-aminobenzoic acid
2,4-trifluoromethyl phenylsulfanyl benzoic acid	3-carboxy-4-hydroxy-Benzenesulfonic acid
2,5-Dichlorothiophene-3-carboxylic acid	3-chlorobenzoic acid
2,5difluorobenzoic acid	3-Chlorothiophene-2-carboxylic acid
2,5-Dihydroxybenzoic acid	3-Cyanophenol
2,6-dihydroxybenzoic acid	3-Cyanopyridine

2-{[3-(trifluoromethyl)phenyl]amino}benzoic acid	3-Hydroxy-2-naphthoic acid
3-hydroxybenzoic acid	5-amino-1-naphthol
3-hydroxypyridine	5-Chlorosalicylic acid
3-hydroxysalicylaldoxime	5-Fluorocytosine
3-methoxy-2-pyridone	5-fluorouracil
3-methylbenzoicacid	5-Hydroxisoquinoline
3-nitrobenzoicacid	5-Methyl-2-thiophenecarboxylic acid
3-picoline	5-Methylresorcinol
4,4'-Azopyridine	5-nitroisophthalic acid
4,4'-Biphenol	6-Hydroxy-2-naphthoic acid
4,4'-Bipyridine	6-methyl-2-pyridone
4,4'-ethane1,2-diyldipyridine	7,7,8,8-Tetracyanoquinodimethane
4,4'-Vinylenedipyridine	9-(4-Methoxyphenyl)xanthen-9-ol
4-amino-2-hydroxybenzoic acid	Acesulfame
4-Aminobenzoic Acid	Acetamide
4-aminophenol	Acetaminophen
4-Aminopyridine	Acetazolamide
4-chloro-3-nitrobenzoicacid	Acetylenedicarboxylic acid
4-Cresol	Acridine
4-Cyanophenol	Adamantane-1,3,5,7-tetracarboxylic acid
4-Cyanopyridine	Adenine
4-dimethylaminopyridine	Adipic acid
4-fluoro-3-nitrobenzoic acid	Anthranilic acid
4-fluorobenzoic acid	Ascorbic acid
4-hydroxyacetophenone	Aspirin
4-hydroxybenzamide	Aza-2-cyclooctanone
4-hydroxybenzoic acid	Azelaic acid
4-methoxybenzoic acid	Azepan2-one
4-methy-1-aminobenzoic acid	Baicalein
4-methylbenzene-1-sulfonic acid	Barbituric acid
4-methylbenzoic acid	Benzamide
4-methylimidazole	Benzamidine
4-nitroaniline	Benzene1,2,3-triol
4-nitrobenzoic acid	Benzene1,2-diol
4-nitroimidazole	Benzene-1,3-dicarboxylic acid
4-nitrophenol	benzene1,4-diol
4-nitropyridine-N-oxide	benzenesulfonic acid
4-phenylimidazole	Benzoic acid
4-phenylpyridine	Benzylamine
4-picoline	biphenyl-2,2'-dicarboxylicacid
4-toluamide	Caffeic acid
5-(4-Chlorophenyl)-6-ethylpyrimidine-2,4-diamine	Caffeine

camphorsulfonic acid	isopropyl hydrogen sulphate
Carbamazepine	Ketoglutaric acid
Carbanilide	lamotrigine
Chloro-acetic acid	levetiracetam
chlorothiazide	Lonidamine
Chrysin	luteolin
Cinnamic acid	MaleicAcid
Citric acid	malic acid
Crotonic acid	Malonamide
Cytosine	Malonic Acid
Decanoic Acid	Mandelic acid
Dextrose	mandelicacid
D-fructose	Mefenamic acid
Diacetamide	Mefenamic acid
Dibenzamide	Melamine
Diflunisal	Menthol
Dihydromyricetin	Mesaconic acid
Dipicolinic acid	Methyl 3-amino 4-hydroxybenzoic acid
Dipropionamide	methyl gallate
entacapone	methyl nicotinate
ethyl gallate	Methyl-4-Hydroxybenzoic acid
Ethylparaben	methylbenzoic acid
Fenamic acid	methylbenzylamine
ferulic acid	Morpholine
Flavone	N,N-dimethylacetamide
flufenamic acid	N,N'dimethylpiperazine
Fumaric Acid	N,N'ethane12diyldiformamide
Gentisic Acid	N+methylbenzamide
Glutaric Acid	N-6-acetamidopyridin-2-yl acetamide
Glycolic acid	N-acetylbenzamide
Heptanoic Acid	Naphthol
Hexamine	Niclosamide
hexanoic acid	Nicotinamide
Hydrocaffeic acid	Nicotinicacid
Hydrocinnamic acid	N-isobutyrylbenzamide
hydroquinone	N-methyl-2-pyrrolidone
imidazole	N-methyl-4-nitroaniline
isethionic acid	N-methylmorpholine
isoniazid	Nonanoic Acid
Isonicotinamide	N-phenylacetamide
isonicotinic acid	N-propionylbenzamide
n-propyl gallate	Resorcinol

Octanoic Acid	Saccharin
Oleic acid	Salicylamide
orcinol	Salicylic Acid
Orotic acid	Sebacic acid
Oxalic Acid	Sorbic Acid
p-coumaric acid	sorbitol
pentafluorobenzoic acid	Stearic acid
pentanoic acid	suberic acid
pentoxifylline	Succinamic acid
Phenazine	Succinamide
Phenobarbital	Succinamide
phenol	Succinic acid
Phenoxyacetic acid	Sulfacetamide
Phenylacetic acid	Sulfamethazine
Phenylsuccinic acid	sulfamethoxypyridazine
Phloroglucinol	sulfametrole
Phthalazine	Sulfathiazole
phthalic acid	Syringic acid
Picolinamide	Tartaric Acid
Picric acid	t-butylhydroquinone
Pimelic acid	tegafur
Piperazine	Terephthalaldehyde
Piroxicam	Terephthalic acid
p-phenylene diacetic acid	Tetramethylpyrazine
propionamide	theobromine
p-terostilbene	Theophylline
p-toluenesulfonic acid	thiomorpholine
Pyrazinamide	Thiosaccharin
pyrazine	thiourea
pyrazine-N,N′-dioxide	tolfenamic acid
Pyrazinoic acid	trans-1,4-Diaminocyclohexane
Pyrazole-3,5-dicarboxylic Acid	Triamterene
pyridine-2,6-dicarboxylic acid	trimesic acid
Pyridine-2-carboxamide	Trimethoprim
pyridine-2,6-diamine	Triphenylphosphine oxide
Pyridoxine	Urea
pyridoxine	valproic acid
Pyroglutamic acid	valpromide
Quercetin	Vanillic Acid
quinoxaline-N,N'-dioxide	Vanillin

Table A4-3 Multicomponent crystals of six APIs with reported 3D crystal structures. CSD identifiers/Refcodes of crystal structures are provided in parenthesis.

Acetaminophen			
4,4'-bipyridine (MUPQAP)	4,4'-ethane-1,2-diyldipyridine (WIGBUL)	Morpholine (AHEPUY)	N-methylmorpholine (MUPPOC)
Piperazine (MUPPUI)	trans-1,4-Diaminocyclohexane (WIGCEW)	2,4-pyridinedicarboxylic acid (SUTVAF)	1,4-diazabicyclo[2.2.2]octane (NIDPUN)
N,N'-dimethylpiperazine (MUPPIW)	Oxalic acid (LUJTAM)	Phenazine (LUJSOZ)	
Acetazolamide			
Azepan-2-one (MADSIW)	4-Hydroxybenzoic acid (RUYGIC)	Nicotinamide (MADTAP)	pyridine-2-carboxamide (DATFIP)
Salicylamide (DATFEL)	6-methyl-2-pyridone (MADSUI)	Theophylline (YEVMUK)	3-methoxy-2-pyridone (MADGEG)
2-aminobenzamide (DATFAH)	valerolactam (MADGIK)	2-pyridone (MADSES)	2,3-Dihydroxybenzoic acid (DATDUZ)
Carbamazepine			
Adamantane-1,3,5,7-tetracarboxylic acid (UNIBIC)	Adipic acid (MOXVEB)	Benzene-1,4-diol (ABOQUF)	Benzoic acid (MOXVAX)
(+)-Camphoric acid (MOXXAZ)	Chlorothiazide (VEJZUI)	Pyrimethamine (KICWOK)	5-nitroisophthalic acid (UNIBEY)
4-aminobenzoic acid (XAQRAJ)	4-amino-2-hydroxybenzoic acid (FAYXOV)	4,4'-bipyridine (XAQQUC)	4-hydroxybenzamide (SOGSEP)
4-nitropyridine N-oxide (JIQKUS)	Fumaric acid (WEYFEN)	Isonicotinamide (LOFKIB)	L-1-hydroxy-2-naphthoic acid (MOXWEC)
Malonic acid (XOBCEX)	N-methylpyrrolidone (KIWBIC)	Pterostilbene (YABHIU)	pyrazine-N,N'-dioxide (VIGGUO)
2,6-pyridinedicarboxylic acid (XAQRIR)	quinoxaline-N,N'-dioxide (VIGGOI)	Succinic acid (XOBCIB)	DL-tartaric acid (MOXWIG)
Thiosaccharin (YAJGEY)	Thiourea (UWAZID)	Trimesic acid (UNIBAU)	2-aminopyrimidine (JIQLAZ)
Vanillic acid (JIQLED)			
Lamotrigine			
Adipic acid (NESBAQ)	Cinnamic acid (HUQVIA)	4-fluorobenzoic acid (OVUMEY)	4-hydroxybenzoic acid (LIBYAY)
Fumaric acid (FUHVOU)	Isethionic acid (QEJHUI)	Malonic acid (VECTUV)	Methylparaben (WUVKEE)
Phenobarbital (VECVEH)	Pimelic acid (NESBIY)	Saccharin (ROJKOS)	Sebacic acid (VECVAD)
Sorbic acid (PEZKEM)	Succinic acid (FOXLUA)	DL-tartaric acid (FOXMEL)	Theophylline (FISNIH)
Thiobarbituric acid (OVUMOI)	Picoline (OVUMUO)	Valproic acid (ROSYIK)	Vanillin (YUYQAM)

Theophylline			
Acetamide (RABXOK)	Acesulfame (BUKYOX)	Acetaminophen (KIGLUI)	Acetazolamide (YEVMUK)
Anthranilic acid (WUTHEA)	Baicalein (KAMQIB)	Benzamide (RABXIE)	Benzoic acid (VAXSAQ01)
Caffeine (NEHJER)	Cinnamic acid (WOCHUT)	Citric acid (KIGKAN)	(+)-dihydromyricetin (OQIFUR)
Entacapone (XIPNOC)	5-Chlorosalicylic (CSATEO)	5-fluorouracil (ZAYLOA)	4-aminosalicylic acid (YUJLOG)
4-chloro-3-nitrobenzoic acid (YEVNIZ)	4-fluoro-3-nitrobenzoic acid (ZEKPIR)	4-hydroxybenzamide (YEVMIY)	Glutaric acid (XEJXIU)
Maleic acid (XEJXEQ)	Malonic acid (XEJXAM)	Melamine (GIDBAY)	Methyl gallate (GAFTUE)
Niclosamide (HEBFEB)	9-(2-methoxyphenyl)-9Hxanthen-9-ol (NIJJIB)	p-nitrophenol (TOPPNP)	N,N'-ethane-1,2-diyldiformamide (DUWXAW)
Orcinol (WOCHAZ)	Oxalic acid (XEJWUF)	p-coumaric acid (IJIBEJ)	Pentafluorobenzoic acid (OPUMAP01)
Phloroglucinol (WOCGUS)	Phthalic acid (LUKXUL)	Picolinamide (YEVNEV)	p-nitroaniline (ZEXTIF)
Pyrazinamide (RACFIN)	Pyrazole-3,5-dicarboxylic acid (UNITER)	Pyridine-2,6-diamine (DEDMEH)	pyridine-2,6-dicarboxylic acid (UNITIV)
Pyridoxine (SETQOA)	Resorcinol (WOCGOM)	Saccharin (XOBCUN)	Sulfacetamide (HOCCUZ)
Sulfamethazine (AWIJEW)	Sulfathiazole (SULTHE01)	DL-tartaric acid (NUJCEC)	Aspirin (DIPJAQ)
Tegafur (DOXFIH)	Thiosaccharin (YAJHID)	3-aminobenzoic acid (WUTJAY)	3,5-dihydroxybenzoic acid (WOCHIH)
3,4 dichlorobenzoic acid (YEVNAR)	3,4-difluorobenzoic acid (OPUNIY)	3,4,5-trifluorobenzoic acid (OPULOC)	3-hydroxybenzoic acid (DOPMUS)
3-hydroxysalicylaldoxime (RUYXUG)	2-aminobenzoic acid (WUTHEA)	2,5-difluorobenzoic acid (OPUNEU)	2,4-dihydroxybenzoic acid (DEYREF)
2,4,5-trifluorobenzoic acid (OPUKUH)	2,4,6-trifluorobenzoic acid (OPULES)	Pyrrolidin-2-one (PICMIA)	2,3-dihydroxybenzoic acid (DOPNAZ)
2,3,4,5-tetrafluorobenzoic Acid (OPULUI)	2,3,4-trifluorobenzoic acid (OPUMIX)	2-((3-(trifluoromethyl)phenyl)amino)benzoic acid (ZIQDUA)	2,2'-(benzene-1,4-diylbis{methanediyl[(pyridin-2-ylmethyl)imino]}) diacetamide (XENGAB)
Urea (DUXZAX)	Diflunisal (OPOGAD)		
Trimethoprim			
Adipic acid (SEMNEE)	Barbituric acid (GIGQIX)	Benzenesulfonic acid (LUWHIU)	Benzoic acid (CUCSEY01)
Cinnamic acid (VASFUS)	Dipicolinic acid (NATHEW)	Flufenamic acid (PORTUO)	4-nitrobenzoic acid (PARWUB)
Fumaric acid (CURSAL)	Glutaric acid (CACBOY)	DL-malic acid (QOVROJ)	Maleic acid (QIKDIX)
Malonic acid (HAMYIE)	Mefenamic acid (PORVEA)	Anthracene-9-carboxylic acid (FEVKID)	Phthalic acid (SEMNAA)
p-toluenesulfonic acid (LUWHAM)	Salicylic acid (MIFWUT)	Sorbic acid (KADFUR)	Succinic acid (YECNEA)

| Sulfamethazine (RIWLOY)	Sulfamethoxypyridazine (QASHEX)	Sulfametrole (HEKRUK)	Sulfanilic acid (LUWHEQ)
Terephthalic acid (VADVOM)	3-bromothiophene-2-carboxylic acid (FEVJUO)	(3-carboxy-4-hydroxybenzene)sulfonic acid (LUWHOA)	3- chlorobenzoic acid (HURMOW)
3-chlorothiophene-2-carboxylic acid (FEVKAV)	3,5-dinitrosalicylic acid (HILPOI)	3-nitrobenzoic acid (PARWOV)	3,3-tetramethyleneglutarimide (GOLDOC)
Tolfenamic acid (PORVAW)	2-aminoterephthalic acid (LIBCOQ)	2,5-dichlorothiophene-3-carboxylic acid (FEVJOI)	2-nitrobenzoic acid (KADGAY)
2- picolinic acid (HEGHIL)			
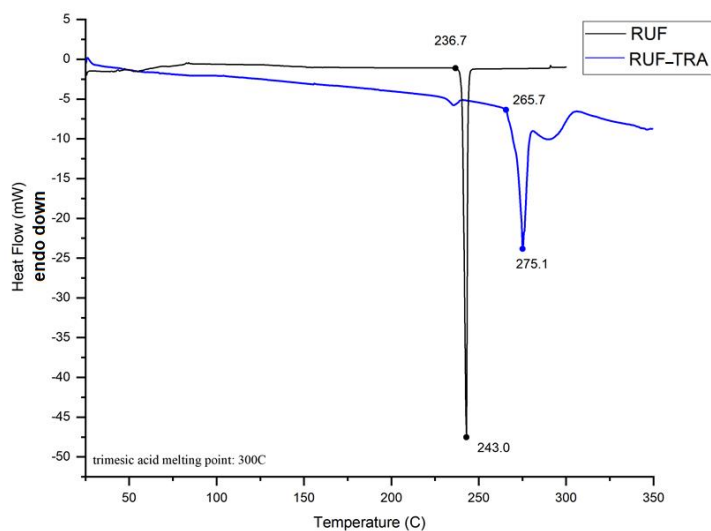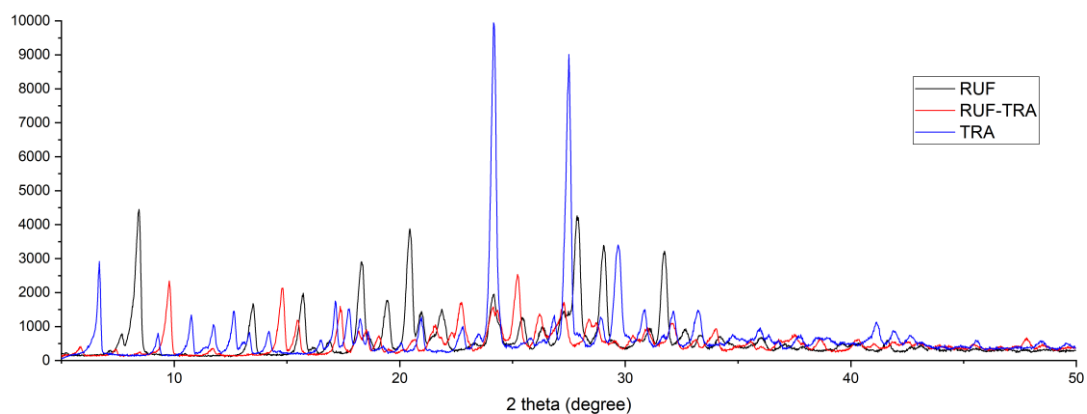
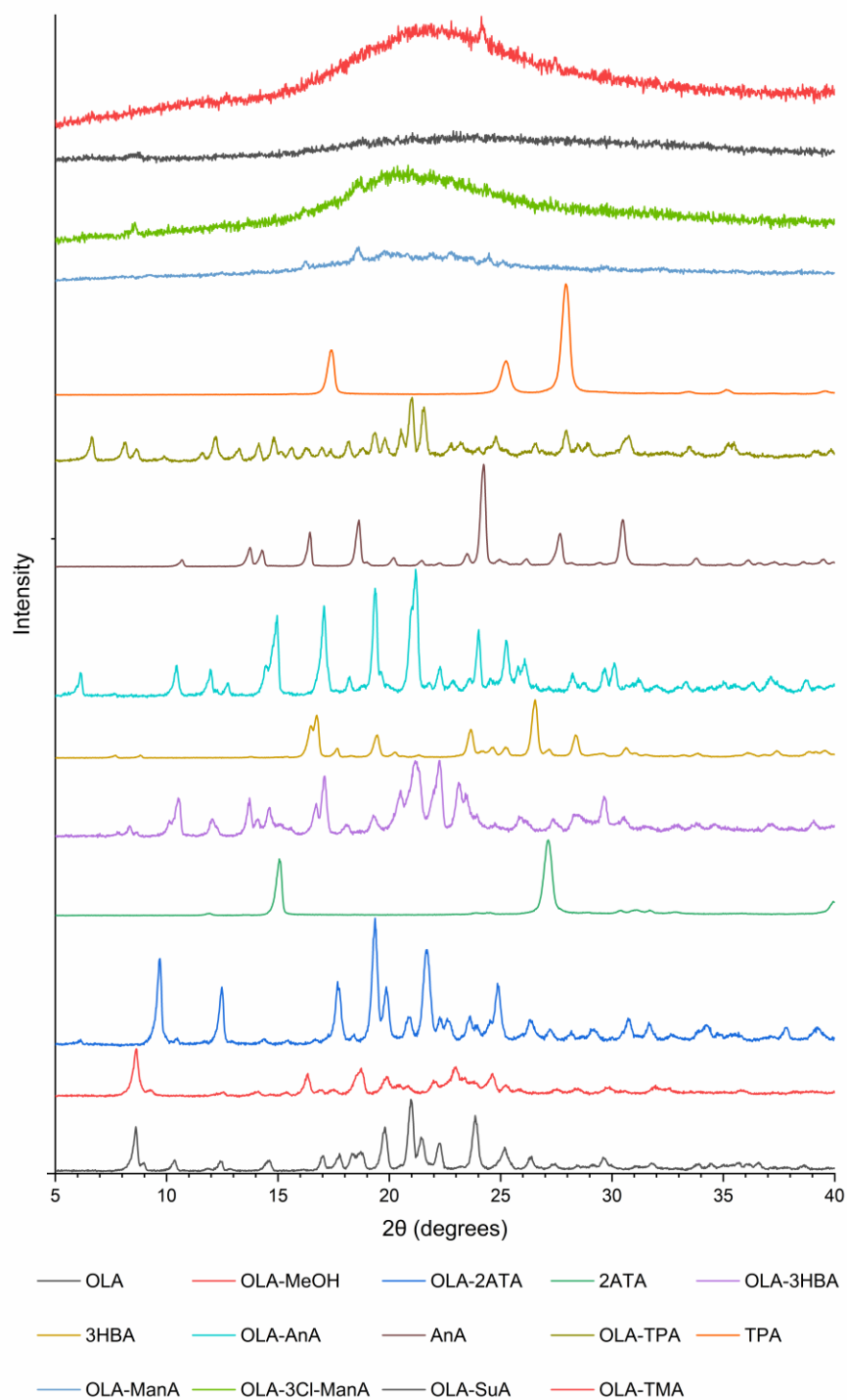Figure A4-1 PXRD (top) and DSC (bottom) of Rufinamide and Trimesic acid (TMA) cocrystal.

Figure A4-2 PXRD diffractogram of Olanzapine, coformers, and multi-component entities. OLA stands for Olanzapine and coformer abbreviations are based on the manuscript.

Table A4-4 The list of identifiers and their weights obtained by logistic regression for the Multivariate Logistic Regression model.

IDENTIFIER	WEIGHTS
E_{A-A}	-1.198
E_{C-C}	0.092
$E_{cocrystal}$	0.153
P_{A-A}	-0.310
P_{C-C}	-0.371
P_{A-C}	-0.340
E_{A-C}	-0.725
E_{C-A}	-0.008
# α_A-β_C	0.287
# α_C-β_A	-1.177

Table A4-5 Salt vs. cocrystal recognition for Olanzapine system based on ΔpKa calculations. ΔpKa values are based on Olanzapine conjugated base – Coformer best acid pKa.

Coformer	Experimental Outcome (Salt/Cocrystal)	ΔpKa	Probability of cocrystal	Probability of salt	ΔpKa Prediction (True/False)
3-Hydroxybenzoic-Acid	Salt	3.4	14.2	**85.8**	True
p-Aminobenzoic-Acid	Salt	2.47	30.0	**70.0**	True
Anthranilic-acid	Salt	2.35	32.0	**68.0**	True
Benzoic-Acid	Salt	3.16	18.3	**81.7**	True
Catechol	Cocrystal	-2.1	**99.1**	0.9	True
Fumaric-Acid	Salt	3.89	5.9	**94.1**	True
hydroquinone	Cocrystal	-2.44	**99.1**	0.9	True
Maleic-Acid	Salt	4.39	0.8	**99.2**	True
Malic-acid	Salt	4.04	0.8	**99.2**	True
Malonic-Acid	Salt	4.81	0.8	**99.2**	True
Oxalic-Acid	Salt	5.88	0.8	**99.2**	True
Phenol	Cocrystal	-2.78	**99.1**	0.9	True
Resorcinol	Cocrystal	-2.02	**99.1**	0.9	True
Salicylic-Acid	Salt	4.45	0.8	**99.2**	True
Succinic-acid	Salt	3.69	9.3	**90.7**	True
Tartaric-Acid	Salt	4.52	0.8	**99.2**	True
Terephthalic-acid	Salt	3.92	5.4	**94.6**	True

Appendix 5

Supporting Information for Chapter 5

Machine Learning-guided Prediction of Cocrystals from DFT-derived Point Clouds

A5 Supporting Information for Chapter 5

A5.1 Data Compilation Procedure

The workflow for compiling positive cases of cocrystallization from CSD begins with searching for entries containing two unique residues within the ConQuest application. Additional filters include "3D coordinates determined," "No error," "No polymeric," "no ions," "only single crystal structures," and "only organics." After collecting the Refcodes from the search results, the individual molecules within the CIF file can be extracted and saved using the following Python code:

```python
from ccdc import io
import os

save_directory = r"C:\Users\Desktop"
list_of_cocrystals = r"C:\Users\Desktop\cocrystals.gcd"

molecule_reader = io.MoleculeReader(list_of_cocrystals,
format='identifiers')

for molecule in molecule_reader:
    mol = molecule
    for component in enumerate(mol.components):
        with io.MoleculeWriter(os.path.join(save_dir, '{}-
{}.sdf'.format(mol.identifier, component[0]+1))) as mol_writer:
            mol_writer.write(component[1])
```

If one of the constituents was a solvent or gas molecule, it was removed from the database since cocrystals are defined as a new crystalline phase in which two compounds are solid at room temperature.

For the negative cases, the CID codes were searched in PubChem, and the 3D structures of the molecules were downloaded in SDF format.

After compiling all chemical structures of the molecules in the positive and negative databases as SDF files, the Gaussian input files (.gjf) were prepared and submitted to Compute Canada for calculations.

A5.2 Hyperparameter Tuning

We utilized Bayesian optimization, a powerful hyperparameter tuning method, within the Keras library to identify the optimal model architecture. A summary of the selected hyperparameters and their ranges can be found in Table A4-1.

Table A5-1 Search spaces of hyper-parameters for PointNet.

Hyperparameter	Range	step
Number of 1-D convolution layers	[2, 6]	1
Number of ith convolution layer neurons	[32, 512]	16
Number of fully connected layers	[2, 6]	1
Number of neurons of the 1st dense layer of classification network	[128, 512]	32
Number of neurons in the hidden layers of the classification network	Number of 1st layer neurons divided by [1,8]	1
Learning rate	[0.0001, 0.1]	Log scale
Dropout Rate	[0.1, 0.55]	0.05
Activation function type	ReLU, Sigmoid	
Batch size	[16, 64]	16

A5.3 Metrics Formula

To ensure unbiased evaluation of the classification model due to the imbalanced ratio of positive and negative samples, we employed the following metrics. True Positive Rate (TPR) and True Negative Rate (TNR) were independently calculated for each class, while Balanced Accuracy (BACC) was used to consider the accuracies of both positive and negative samples. The equations for these metrics are as follows:

TPR = TP / (TP + FN) (1)
TNR = TN / (FP + TN) (2)
BACC = (TPR + TNR) / 2 (3)

In the above formulas, TP, TN, FP, and FN stand for True Positive, True Negative, False Positive, and False Negative, respectively. These metrics provide a comprehensive assessment of the model's performance, addressing the challenges posed by the imbalanced sample ratio and lead to a fair evaluation of its predictive capabilities.

Curriculum Vitae

Name:	Soroush Ahmadi Nasrabadi
Post-secondary Education and Degrees:	Bahá'í Institute for Higher Education (BIHE) Tehran, Iran 2011-2017 B.Sc. The University of Western Ontario London, Ontario, Canada 2018-2023 Ph.D. University of Zurich Zurich, Switzerland 2023 Invited visiting student (Ph.D.)
Honours and Awards:	Bahieh Moayyad-Khademi Scholarship Award 2017 Mitacs Accelerate Internship 2021-2022 Mitacs Globalink Research Award 2023
Related Work Experience	Teaching Assistant Bahá'í Institute for Higher Education (BIHE) 2012-2018 Lecturer Bahá'í Institute for Higher Education (BIHE) 2019-2022 Teaching Assistant The University of Western Ontario 2019-2023

Publications:

- Ahmadi, S., & Rohani, S. (2023). Overcoming the Hydrophobic Nature of Zinc Phenylacetate Through Co-Crystallization with Isonicotinamide. *Journal of Pharmaceutical Sciences*, *112*(7), 1929–1938.

- Gong, W., Mondal, P. K., Ahmadi, S., Wu, Y., & Rohani, S. (2021). Cocrystals, Salts, and Salt-Solvates of Olanzapine; selection of coformers and improved solubility. *International Journal of Pharmaceutics*, 608, 121063.

- Ahmadi, S., Mondal, P. K., Wu, Y., Gong, W., Mirmehrabi, M., & Rohani, S. (2021). Virtual Multicomponent Crystal Screening: Hydrogen Bonding Revisited. *Crystal Growth & Design*, *21*(10), 5862–5872

- Ahmadi, S., Mondal, P. K., Mirmehrabi, M., & Rohani, S. (2021). Desolvation of dasatinib methanolate: an improved anhydrous polymorph. *CrystEngComm*, *23*(24), 4272–4283.

- Ahmadi, S., Wu, Y., & Rohani, S. (2019). Molecular dynamics simulation of homogeneous nucleation of supersaturated potassium chloride (KCl) in aqueous solutions. *CrystEngComm*, 21(48), 7507–7518.

- Ahmadi, S., & Sefidvash, F. (2018, April 2). Study of Pressure Drop in Fixed Bed Reactor Using a Computational Fluid Dynamics (CFD) Code. *ChemEngineering*, 2(2), 14.