L-Università ta' Malta

WHEN COMPUTATION MEETS EXPERIMENT:

A COMBINED APPROACH TOWARDS UNDERSTANDING CRYSTALLINE MATERIALS PROPERTIES

A thesis submitted to the Faculty of Science for the degree of Doctor of Philosophy at the University of Malta.

Supervisors: Prof. Ulrich Baisch and Prof. Liana Vella-Żarb

Lorella Spiteri
2023

FACULTY~~/INSTITUTE/CENTRE/SCHOOL~~ Science
DECLARATION OF AUTHENTICITY FOR DOCTORAL STUDENTS

Student's Code ________________

Student's Name & Surname ____________________________

Course Doctor of Philosophy

Title of Dissertation/Thesis
When computation meets experiment: A cominbed approach
towards understanding crystalline materials properties

(a) Authenticity of Thesis/Dissertation

I hereby declare that I am the legitimate author of this Thesis/Dissertation and that it is my original work.
No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.
I hold the University of Malta harmless against any third party claims with regard to copyright violation, breach of confidentiality, defamation and any other third party right infringement.

(b) Research Code of Practice and Ethics Review Procedure

I declare that I have abided by the University's Research Ethics Review Procedures. Research Ethics & Data Protection form code SCI-2021-00061.

☑ As a Ph.D. student, as per Regulation 66 of the Doctor of Philosophy Regulations, I accept that my thesis be made publicly available on the University of Malta Institutional Repository.

☐ As a Doctor of Sacred Theology student, as per Regulation 17 (3) of the Doctor of Sacred Theology Regulations, I accept that my thesis be made publicly available on the University of Malta Institutional Repository.

☐ As a Doctor of Music student, as per Regulation 26 (2) of the Doctor of Music Regulations, I accept that my dissertation be made publicly available on the University of Malta Institutional Repository.

☐ As a Professional Doctorate student, as per Regulation 54 of the Professional Doctorate Regulations, I accept that my dissertation be made publicly available on the University of Malta Institutional Repository.

____
Signature of Student

____
Name in Full (in Caps)

____
Date

To my late nanna Kikka,

may I be able to embrace life and bring joy, just as you did

ACKNOWLEDGEMENTS

Firstly, I wish to thank Prof. Liana Vella-Żarb and Prof. Ulrich Baisch for giving me the opportunity to embark on this journey within the CrEMa research group, as well as for their trust in my capabilities. I would like to thank Dr Fiona Sammut at the Department of Statistics for her technical input with regards to the statistical methods used in chapter three, and also for her time dedicated to helpful discussions.

I would like to show appreciation towards the lab officers at the Department of Chemistry, and my fellow research group members for their assistance and encouragement. I would also like to thank the members of online scientific communities, particularly fellow software users, who provided feedback or information whenever needed. This definitely lightened the burden of working on a PhD during a pandemic.

I would like to thank my friends for their support and patience throughout this entire journey. I wish to show my immense gratitude towards my family members, particularly my mother, Silvana, who has been my rock and believed in my dreams and myself, even when I had doubts or struggled. Lastly, I am greatly thankful to James, who has lived through this journey with me; thank you for understanding, for your patience, for being present and, for countless herbal teas and meals.

This degree was carried out following the award of a Tertiary Education Scholarships Scheme Scholarship.

ABSTRACT

The solid-state behaviour of an active pharmaceutical ingredient has a significant influence on its function and effectiveness in the body. Due to such importance, a crystallographic study was conducted to analyse and predict such behaviour in three prominent antivirals, namely ganciclovir, famciclovir and valacyclovir (hydrochloride form), all of which act as nucleoside analogue polymerase inhibitors against herpesviruses. This investigation incorporates the complementary use of multiple computational approaches, novel or otherwise, which offer a direct comparison of results and provide a more holistic understanding.

A new scoring system for the quantification of molecular flexibility was devised, given the proven lack of appropriateness of the number of rotatable bonds parameter to represent such molecular feature exhaustively. This was followed by conformational analysis, which provided the parameters of the accessible conformational space of each molecule. In addition, this investigation entailed a detailed exploration of each crystalline antiviral, particularly the administered form, from intramolecular and intermolecular levels, including full interaction maps and energy frameworks. This analysis provided an insight of how features at these levels can manifest their impacts on supramolecular traits, such as the conformational polymorphism between GCV form I and II. The crystallographic behaviour was also characterised through the use of variable temperature powder X-ray diffraction and simultaneous thermal analysis. Both intermolecular analysis and experimental work supported the thermodynamic stability of GCV and FCV form I, with minimal risk of polymorphism, as opposed to VCV·HCl. This study displayed how polymorphism is governed by the present degree of stability, which in return is influenced by the HB donors to acceptors ratio, HB coordination number per group and the strength of other non-bonding contacts, beyond HB.

Different polymorph and co-crystal formation prediction techniques were validated against parallel experimental work and examined to understand better the extent of their applicability. Results demonstrated how the HB propensity model approach is not adequate to classify the thermodynamic stability of forms whose non-bonding network is heavily composed of non-conventional HB interactions and other contacts beyond HB, as GCV and FCV form I. The prediction of co-crystallisation was conducted through four

distinct approaches, which were accompanied by the testing of multiple settings so as to improve the quality of outcomes. These considerations were related to the inputted molecular conformations and the selection of interactions. The ranking based on the excess enthalpy evaluated through the COSMO approach, resulted to be an adequate indicator of the suitability of coformers, especially for FCV co-crystals. The highest overall success rate was associated with the HB energy approach at 74.47%, followed by molecular complementarity (68.97%) and HB propensity model (65.38%) procedures, all of which performed better for GCV trials.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

LIST OF ABBREVIATIONS

ACV	Acyclovir
AIC	Akaike Information Criterion
API	Active Pharmaceutical Ingredient
ASC	Apparent Surface Charge
B3LYP	Becke, 3-parameter, Lee-Yang-Parr
BFDH	Bravais, Friedel, Donnay and Harker
BP	Becke’s exchange and Perdew’s correlation functionals
CCDC	Cambridge Crystallographic Data Centre
CMV	Cytomegalovirus
COSMO	Conductor-like Screening Model
COSMO-RS	Conductor-like Screening Model for Realistic Solvation
CSD	Crystallographic Structural Database
CSM	Continuum Solvation Model
CSP	Crystal Structure Prediction
DFT	Density Functional Theory
DM	Dipole moment
DNP	Double numeric plus polarisation
FCV	Famciclovir
FNO	Fraction of Nitrogen and oxygen
GAMESS	General Atomic and Molecular Electronic Structure System
GCV	Ganciclovir
GRACE	Generation Ranking and Characterisation Engine
GRAS	Generally Recognised as Safe
HB	Hydrogen Bonding / Hydrogen Bond / Hydrogen bonds
HBE	Hydrogen Bonding Energy
HBP	Hydrogen Bonding Propensity
HF	Hartree Fock

LIST OF ABBREVIATIONS (continued)

HCMV	Human Cytomegalovirus
MBD	Many-body dispersion
MC	Molecular Complementarity
MD	Multi-Differential
MM	Molecular Mechanics
MOPAC	Molecular Orbital Package
MP	Mollet-Plesset
LAG	Liquid assisted grinding
LHP	Logit Hydrogen Bond Propensity
PBC	Periodic boundary conditions
PBE	Perdew-Burke-Ernzerhof
PCA	Principal Component Analysis
QC	Quantum chemical
RMS	Root Mean Squared
ROC	Receiver Operating Characteristic
SD	Single-Differential
SP	Single point energy calculation
TK	Thymidine kinase
TNP	Triple numeric plus polarisation
TS	Tkatchenko-Scheffler
TZVP	Ahlrich’s Triple- Zeta Valence Polarisation basis sets
VCV	Valacyclovir
VCV·HCl	Valacyclovir monohydrochloride
vdW	van der Waals

Chapter 1

INTRODUCTION

1.1 Context of Research

The process for a molecule with a promising potential to become a commercially available pharmaceutical drug entails several laborious stages from laboratory research to clinical trials and rigorous screening prior to possible licensing approvals.[1] It is therefore comprehensible that drug development is characterised by long duration and also high costs. A recent study, based on 18% of novel drugs approved by the US Food and Drug Administration over a period of 10 years, showed that the approximated median expenditure for research and development for each product, was $ 985 million.[2] These characteristics have motivated several investments, in attempt to develop sustainable chemical technologies and maximise the efficiency of this process.[3] The urgency of creating shorter, less expensive methodologies for drug development, without compromising the quality of the commercialized pharmaceutical drug, was even more highlighted with the present coronavirus disease pandemic, with over 678 million cases till February 2023 worldwide, of which around 1% led to death. Such extraordinary, but not unprecedented situation demonstrated the importance for the optimization of such process, whose quality and efficiency affect so many lives. Furthermore, during these times, it was evident how the speed and results of drug development had a direct impact also on the world's economy, children's education and day-to-day lifestyle of the majority of the world's population.

One of the most effective methods applied to minimise the duration of early development stage of an active pharmaceutical ingredient (API) is through the use of computational techniques. The prediction of crystal structure behaviour assists laboratory research by providing theoretically-based suggestions which can reduce the number of experimental trials to those with potentially favourable outcomes, thereby minimising resources and increasing the chance of success.[4-5] Inferences extracted from such procedures can also provide additional assurance towards experimental results, such as the exhaustiveness of polymorph screening. Furthermore, computational advancements also have the ability to facilitate the process of optimising the properties of materials, particularly enhance the bioavailability of pharmaceutical drugs.

1.2 Statement of Scope

The main objective of this study is to investigate and predict solid-state behaviour through the simultaneous use of different computational techniques, novel or otherwise (already present in literature). The complementary use of multiple approaches offers a direct comparison of results and also the possibility of analysing the same subject from different point of views, which might ultimately provide a more holistic understanding. Such approaches involve the utilisation of molecular modelling and energy calculations, as well as combination of statistics and experimental data. The robustness of computationally deduced hypotheses is tested through experimental trials.

This research project focuses on pharmaceuticals, whose thermodynamic profile is commonly under great scrutiny throughout all the stages of development and beyond, given that any change could potentially affect their properties and effectiveness.[1] The methodology is applied to three closely related antivirals, namely ganciclovir, famciclovir and valacyclovir (monohydrochloride form), all of which act as nucleoside analogue polymerase inhibitors against herpesviruses.[6] This investigation entails a detailed exploration of each crystalline antiviral from intramolecular and intermolecular levels, providing an insight of the nature of the molecule and how features at these levels can manifest their impacts on supramolecular traits. This analysis would act as the prologue for polymorph screening, as well as the evaluation of any potential co-crystal formation.

Chapter 2

BACKGROUND THEORY AND LITERATURE REVIEW

2.1 Crystallography in the pharmaceutical industry

Solid-state forms are very present in the pharmaceutical industry, especially since in most cases, a drug substance is handled in its solid form at some stages of its manufacturing process.[1] Such presence is increased by the popularity of administering oral dosage forms as tablets or capsules, as well as inhalation dosage forms as dry powders, together constituting more than 80% of the current formulated drugs. Even though crystallisation processes have been present in this industry for many decades, there was a very recent exponential increase in requirements to ensure the safety and energetic stability of API. This is due to the potential impacts brought around by the unplanned emergence of another form. A change in form could potentially lead to variations in mechanical, surface and physical properties of the drug as dissolution rate, solubility and hardness. These changes would not only impact the manufacturing process of the pharmaceutical drug, but also its bioavailability and mechanism of action in the patient. The dependence of such properties on its crystal structure is a fundamental part for drug development. Examples as ritonavir and rotigotine provide evidence of the possible ethical, health and financial consequences of such occurrences.[7-8] A recent review by Gao *et al.* details new advancements in the crystallisation process of pharmaceuticals, focusing primarily on the nucleation stage of crystal growth, in attempt to gain more control and assurance of the final product.[9]

Information regarding the solid form of a pharmaceutical drug is essential, even if the drug is projected to be commercialised as a solution, due to the different physical properties associated with distant forms. Even though the feature of having different properties imposes the requirement to perform exhaustive studies of the API, it also creates the opportunity to tune the solid-state drug so as to have the optimal characteristics. This opportunity embodied in crystal engineering is highly critical, especially when taking into account that around 70% of the currently formulated drugs have low solubility.[10] An enhancement in physicochemical properties (solubility permeability, dissolution rate, physical stability) would lead to the optimisation of bioavailability, hence lowering the patient's exposure to probable toxic effects and potentially increasing the therapeutic efficiency.[1] The pursuit of the optimal pharmaceutical drug has to ensure that each modification does not jeopardise its ex vivo and in vivo stability. Modulation of APIs can occur through diverse methods, as portrayed in figure 2.1, including chemical modification as salt formation, intermolecular bonding alternation as polymorphism and physical change as particle size reduction. In addition, all the possible formations also have their corresponding potential polymorphs.

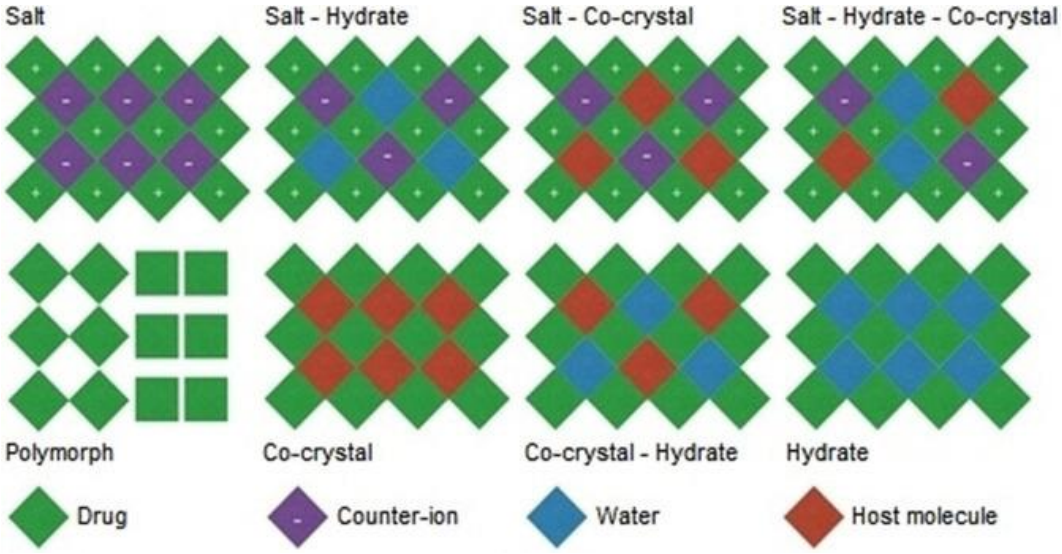


Figure 2.1: A pharmaceutical landscape of possible formations of solid-states.[10]

2.1.1 Crystallography

Given the presence of solid-states in such industry, it is essential to understand the fundamentals of crystallisation in order to be able to control, predict and engineer solid-state pharmaceutical ingredients. Crystallisation can entail a liquid-solid separation, governed by thermodynamic and kinetic factors, where a mass is transferred out of solution, forming a solid with a degree of 3D periodic arrangements.[11] Alternatively, crystallisation can occur through other routes including sublimation, from melt and grinding (dry or liquid-assisted). The repeated units are held together by different types of non-covalent interactions primarily van der Waals (VdW) forces and hydrogen bonding (H-bonds) contacts. The range of order determines the phase type of the entity, which can vary from crystalline (long-range order) to amorphous (short-range order / overall disordered). Some amorphous materials can retain their disorder while others can overcome their associated kinetic barriers and transform into crystalline structures.[12] Further subdivision of the solid-phase class was proposed after taking into consideration the extent of order and condensation in conjunction with large-amplitude motion.[13] In fact, it was debated that the term “solid” should not be used as a scientific term, given that it omits the presence of substances with different degree of solidity, including liquid crystals and plastic crystals.[13]

2.1.1.1 Introduction to the Crystallisation Process

Crystallisation is a process which primarily constitutes 2 stages: nucleation and crystal growth. Nucleation is initiated by bimolecular collisions in supersaturated solutions, which proceed to form aggregates of the dissolved material, some of which might form stable clusters.[14] Nucleation is thermodynamically driven by relative supersaturation, which is the discrepancy between the chemical potential of a molecule at equilibrium and at a supersaturated state, at given physical conditions. Relative supersaturation, σ or supersaturation ratio, S can also be expressed in terms of concentrations (eq. 2.1), where c_{eq} is the concentration at equilibrium and c_{ss} is a concentration greater than saturation.[14]

$$\sigma = \ln(S) = \, ln\left(\frac{c_{ss}}{c_{eq}}\right) \approx \frac{c_{ss}-c_{eq}}{c_{eq}} \qquad (2.1)$$

Evaporation and controlled cooling are two commonly used methods to form supersaturated solutions for crystallisation, taking into account that the rate of each technique will affect not only the size of the crystals formed, but also potentially favour the emergence of a polymorph over others. The creation of a supersaturated solution relies on the solubility of that particular solute in solvent, which in return depends on the quantity and quality of solute- solvent, solvent-solvent and solute-solute interactions. Figure 2.2 illustrates a typical solubility-supersaturation graph, with three zones; the stable zone where the solution is undersaturated and no nucleation occurs, the metastable zone where spontaneous nucleation is very unlikely, but seeded crystal growth may take place (secondary nucleation) and the supersaturated area where spontaneous nucleation, followed by crystal growth occurs.

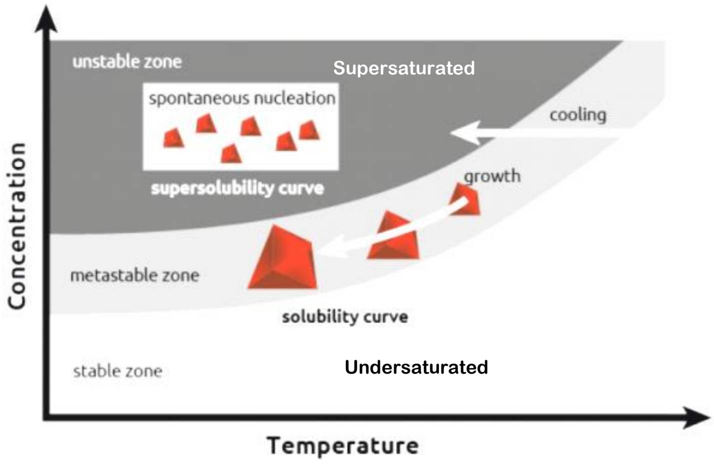


Figure 2.2: Solubility- supersaturation schematic graph, showing the corresponding solubility curves and defined zones.[15]

Given that nucleation occurs at a molecular level, this process has been investigated and modelled in terms of changes in free energy. The Classical Nucleation Theory, proposed by Volmer *et al.* is the most commonly used theoretical model to explain nucleation in a homogenous system.[16] It states that the change in Gibbs free energy during nucleation, ∆G is a result of an interplay between the change in surface free energy and volume or bulk free energy, both of which are dependent on the radius of the nucleus. This relationship is illustrated graphically in figure 2.3 and defined in equation 2.2, where Ω is the specific volume of a growth unit, *r* denotes the cluster size and γ is the solid-liquid interfacial energy.[17]

$$\Delta G = -\left(\frac{4\pi r^3}{3\Omega}\right) kTln(S) + 4\pi r^3\gamma \qquad (2.2)$$

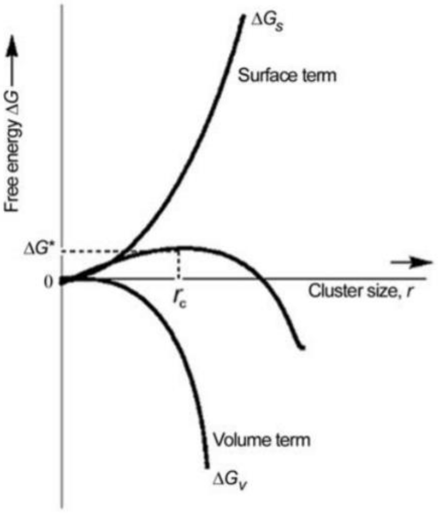


Figure 2.3: A graphic representation of the change in free energy of nucleation with respect to the cluster size, as a resultant effect of both surface and volume free energies.[12]

It is evident that the surface free energy dominates at small values of r, while the bulk free energy is prevalent beyond the critical nucleus size, r_c, which is the radius at the maximum free energy.[17] The probability of nucleus formation is at a minimum at r_c and as the cluster size increases, the free energy decreases, hence making nucleation even more probable. The emergence of a particular form over any other possible metastable structure depends on thermodynamic and kinetic factors, both of which might favour different polymorphs. With respect to kinetics, the nucleation rate, J can be defined using the Arrhenius equation (eq. 2.3), incorporating A which is a factor derived from kinetic considerations, k is the Boltzmann constant, ΔG^* is the free energy at r_c and T denotes the temperature.

$$J = Aexp\left(\frac{-\Delta G^*}{kT}\right) \qquad (2.3)$$

Throughout decades, other theories were proposed to model crystallisation with respect to thermodynamics and kinetics, dealing with both homogenous and heterogeneous nucleations.[18-20] These theories were tested against experimental data and generalised even further as described by several publications, such as the incorporation of the deviation in bulk properties between critical clusters and macroscopic phases within the classical Gibbs theory.[14, 17, 21-23] One of the most stated mechanisms is the Ostwald's rule of stages which claims that when a system departs from a metastable phase, it will go to a more stable one with the closest Gibbs free energy, rather than directly to the most stable form.

After nucleation, particular critical nuclei proceed with further growth through the addition of more solute particles; a process whose discriminative nature governs the assembly of growth units within a crystal. The process of crystal growth is highly influenced by both the internal crystal structure controlled by the intermolecular interactions, some of which are listed in table 2.1, as well as by external conditions including temperature, pressure and type of solvent used. The periodic arrangement of crystals is built by the repetitive translation of a unit, through a specific symmetry. One of the simplest crystal growth models dictates how the incorporation of these units onto the surface of the crystal can take place at three distinct sites; ledges (one location of intermolecular contact), step sites (two locations of intermolecular contacts) and kink sites (three locations of intermolecular contacts).[24] Given that units incorporated at kink sites have the highest number of possible locations for intermolecular interactions, the inclusion at these sites is the most energetically favourable. Further in-depth reviews about crystal growth mechanisms can be found in literature.[24-25] Along the progression of crystal growth, the rate with which each façade grows will determine the final external shape, known as morphology.

Table 2.1: Some of the different types of interactions that govern the mechanism of constructing crystals. The energy values only provide an estimate typical range.[26-27]

Contact type	Strength of interaction	Comment
Van der Waals	Weak (< 2 kJ/mol)	Strength can vary because it depends on the size and shape of the molecules
Dipole – dipole	Weak (1 – 3 kJ/mol)	Strength is influenced by the nature and flexibility of the interacting species
H-bonds	Strong (10 – 30 kJ/mol)	Directionality depends on the HB contributors and steric hindrance
Ion - dipole	Strong (> 30 kJ/mol)	Depends on the ion charge and polarity of molecule

The unit cell (repeated unit) of each crystal contains all the required information to construct the crystal itself by translation, described by the lattice vectors and symmetry. The different correlations between the angles; α, β, γ and dimensions; a, b, c defining the lattice vectors create seven crystal systems. These systems are dependent on the progression of crystal growth, for example in case that a crystal grows uniformly in all directions, the crystal system will be cubic with $\alpha = \beta = \gamma = 90\,°$ and $a = b = c$. Apart from the primitive seven crystal systems, one can also take in account extra lattice points in face- or body-centred locations, all of which together yield the 14 unique Bravais lattice systems. The combination of translational and point symmetry elements with the 14 Bravais lattice systems yield 230 different space groups, which can used to describe any existent crystal structure.

2.1.2 Polymorphs

The phenomenon of polymorphism has been known since the late 1700's, when it was observed that diamond and charcoal only differ in the way that their structures are crystallised, and such difference led to significant variation in their physical properties.[28] M.L. Frankeheim was one of the first researchers to define some principles of polymorphism, based on his observations on mercuric iodide polymorphs.[29] Despite the abundance of polymorphs in literature in both organic and inorganic substances, there is still an absence of a standard definition. However, there seems to be a consensus whereby polymorphism describes the ability of a crystal to exist in more than one phase, each of which is characterised by a different molecular arrangement within its lattice. This means that the internal structure of each phase is distinct due to either conformational differences or dissimilar packing arrangement, despite being composed of the same chemical species. As seen in figure 2.1, crystal forms can either be composed of single or multiple-components, but all of which can have polymorphs, such as the two conformational polymorphs of ethenzamide : ethylmalonic acid co-crystal.[30] Therefore the above description can also be extended to co-crystals and salts polymorphs.

This definition was first proposed by McCrone in 1965, whose work was pivotal with respect to the understanding of the complex nature of polymorphism.[31] Another instrumental publication with respect to polymorphism, is the review written by McCrone and Haleblian in 1969, concerning its crucial role in the pharmaceutical industry.[32] Unfortunately, these two papers were only appropriately acknowledged around the time when the Zantac litigation arose, a case which revolutionised the approach towards polymorphism in such industry. [33]

Within his work, McCrone also introduced the term *conformational polymorphism*, which refers to cases where a molecule is flexible enough to exist in distinct conformational orientations, hence modifying the non-bonding network which possibly leads to a unique packing arrangement within the crystal structure.[31] This implies that in general, for conformational polymorphism to occur, the molecule should either have relatively lower torsional energy barriers or very strong intermolecular network to compensate for a strained conformation. Given the impact of molecular flexibility, conformational analysis through molecular mechanics simulations is considered to be a vital step prior to polymorph screening, especially when dealing with pharmaceutical ingredients as acyclovir or paracetamol.[34-35] Nangia *et al.* provided an in-depth analysis of several examples of conformational polymorphism in fuchsones, and confirmed that their likelihood of exhibiting polymorphism is correlated to the conformational diversity of the *exo*-methylene phenyl rings.[36] Such mechanism is also observed amongst the 5-Methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile polymorphs, also known as the ROY (red, orange, yellow colours) forms due to their characteristic colours, mainly as a result of the flexibility of the thiophene torsion angle.[37] In cases of having distinct forms due to isomerism / tautomerism, the term *configurational polymorphism* is more applicable.[29]

Packing polymorphism refers to those examples where the molecules are packed in a distinct manner within the crystal lattice, while maintaining the same molecular conformation. This type of polymorphism is exhibited between two forms of salicyamide – oxalic acid co-crystal, whereby single crystal X-ray diffraction proved how they only differed in their packing arrangement.[38] Two anhydrous polymorphs of acyclovir are displayed in figure 2.4, visualising the difference in the hydrogen bonding which leads to a variation in the packing systems.

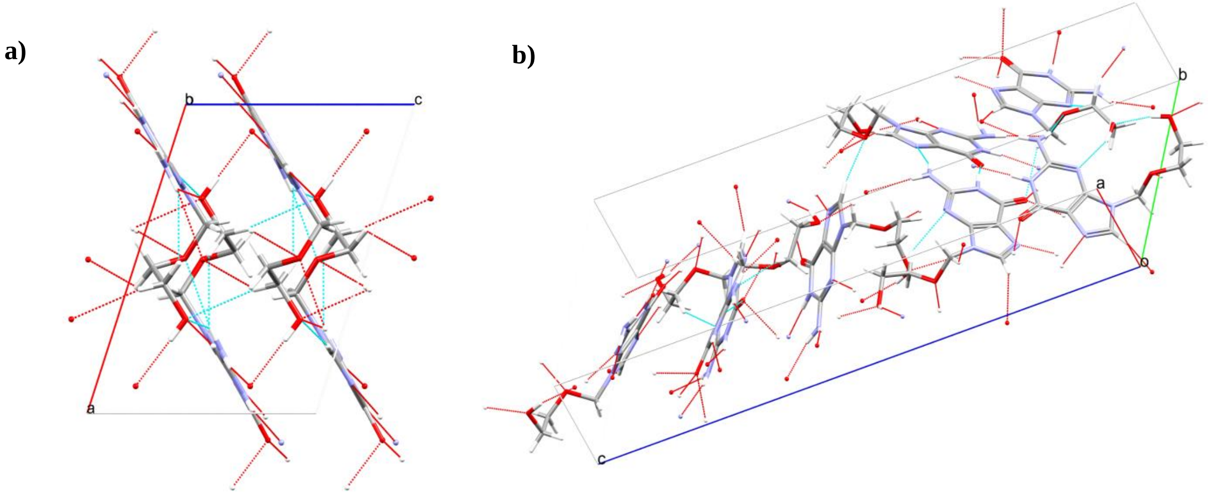


Figure 2.4: Two acyclovir anhydrous polymorphs; a) form II (CSD refcode: MECWIC) and b) form I (CSD refcode: MECWIC03), with hanging H-bonds in red and satisfied H-bonds in turquoise.

Despite having such classes of polymorphism, there is a large population of polymorphs which embody both definitions because difference in conformation is very likely to impose a change in the packing arrangement and vice versa.[29] Another important note is that a change in the crystal morphology does not necessarily implicate the formation of another polymorph, as seen in one of naphthazarin forms which exhibits two distinct crystal habits.[29]

Figure 2.5: Polymorph C of naphthazarin has the ability to exhibit two distinct crystal habits, while retaining the same packing arrangement.[29]

2.1.2.1 Thermodynamics of polymorphs

Thermodynamics dictates the stability of a particular crystal form, hence determining whether other potential forms can actually exist or no at particular physical conditions. Therefore, the relationship between polymorphs has to be examined from a thermodynamic and kinetic point of view, in attempt to have a better insight of the underlying mechanism. The difference in thermodynamic properties within polymorphs, including the internal energy, U and the molar heat capacity, C_m, arises from the variation in their intermolecular network, which highly influences the way by which molecules within the crystal lattice dissipate heat.[39] The uniqueness of such network is derived from the particular combination of its contributors, such as H-bonds and vdW forces, and their associated properties.

Polymorphic transformation is driven by Gibbs free energy, whereby a less energetically stable form will transform into the most stable form, which is characterised by the lowest free energy at those conditions, assuming that such transformation is kinetically viable. This thermodynamic relationship between polymorphs was suggested by Buerger and Bloom, where they specified how with an increase in temperature, the most stable would be the one with the lowest free energy, even if that form might not have the lowest configuration energy.[40] They also argued that only specific arrangements of increasing configurational energy can form while increasing the temperature, without the formation of any intermediate phases.

Equation 2.4 describes the relationship between the change in enthalpy, $\Delta H_{II\to I}^{T_1}$, in entropy, $\Delta S_{II\to I}^{T_1}$ and in Gibbs free energy, $\Delta G_{II\to I}^{T_1}$, accompanying the polymorph transformation from form II to I at temperature, T_1 in a dimorphic system, where form I is the most stable polymorph at absolute zero.[29, 39]

$$\Delta G_{II\to I}^{T_1} = \Delta H_{II\to I}^{T_1} - T_1 \Delta S_{II\to I}^{T_1} \qquad (2.4)$$

This description only holds if taking into accounting the following conditions: a) The conformational difference between forms I and II at 0 K is at a minimal, therefore $\Delta H_{II\to I}^{0}$ can be approximated to zero, b) the forms are perfectly ordered at 0 K, hence $\Delta S_{II\to I}^{0}$ can be approximated to zero and c) there are no phase transitions over the temperature range of 0 K to $T_{1.}$.

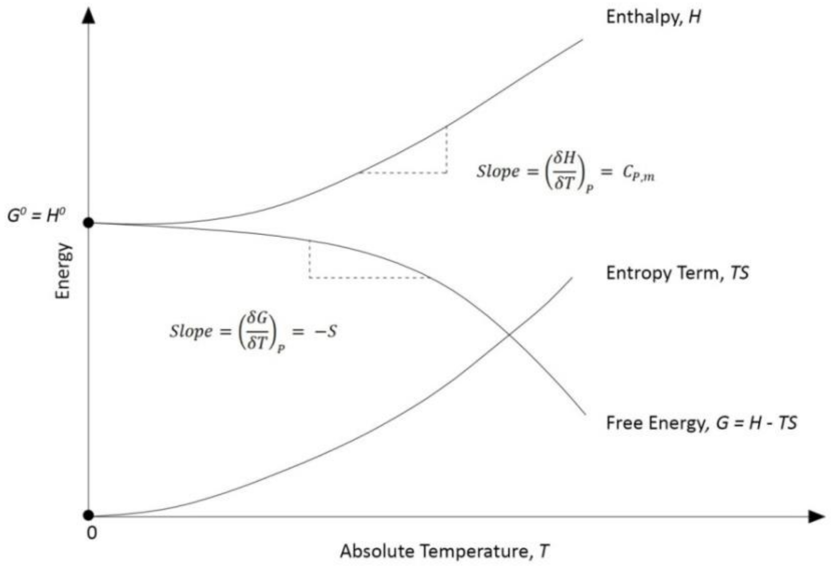


Figure 2.6: Graph showing energy (arbitrary units), against temperature (K), describing the behaviour of free energy, entropy and enthalpy with temperature in polymorphs.[39]

The thermodynamic profile of polymorphs, as expressed in equation 4, can be illustrated through graphs as shown in figure 2.6, offering an interpretable summary of polymorphic behaviour with temperature. Prior to such version, the Helmholtz free energy was used instead of the Gibbs free energy, but the latter is more commonly utilised due to its readily accessible experimental data, especially using differential scanning calorimetry or heat capacity measurements.[29] The curve for the enthalpy can be determined using the heat capacity, whose positive values are demonstrated through the gradient of such isobar. The curve mapping the entropy term illustrates an increase in value (always positive) with an increase in temperature due to the Third law of thermodynamics. Gibbs free energy is a constantly decreasing function because its partial derivative with respect to temperature is the negative value of the entropy.

2.1.2.1.1 Polymorphic relationships

The thermodynamic relationship between polymorphs can be described by the terms *enantiotropy* or *monotropy*, which revolve around the temperatures at which polymorph transition occurs in relation to each respective melting point. More specifically, in an enantiotropic system such transition occurs below the lower melting point, and is thermodynamically reversible. On the other hand, monotropy refers to a system in which the transition is irreversible, and does not depend on the temperature below the two melting points.[39] Despite being the general rule, there are cases where the transition between enantiotropic polymorphs is not kinetically reversible, e.g. carbamazepine (III/I) and sulfamerazine (II/I). [41] In such cases, the irreversible transition does not occur at the thermodynamic transition temperature, possibly because of the high energy barrier to change into the initial form. The dominance of kinetic factors is also experienced at high pressure conditions, as in the case of tolazamide and spironolactone, leading to deviations from the expected behavior described by the above standard definitions.[42-43]

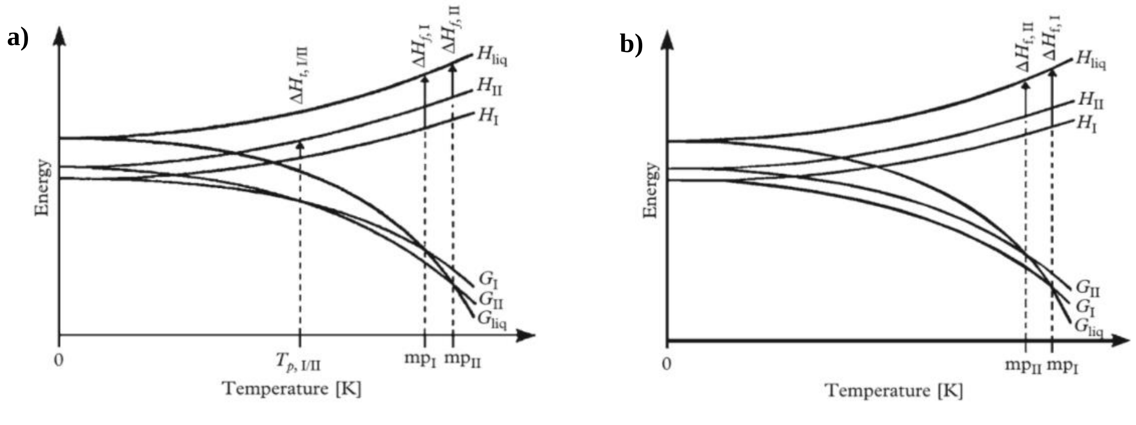
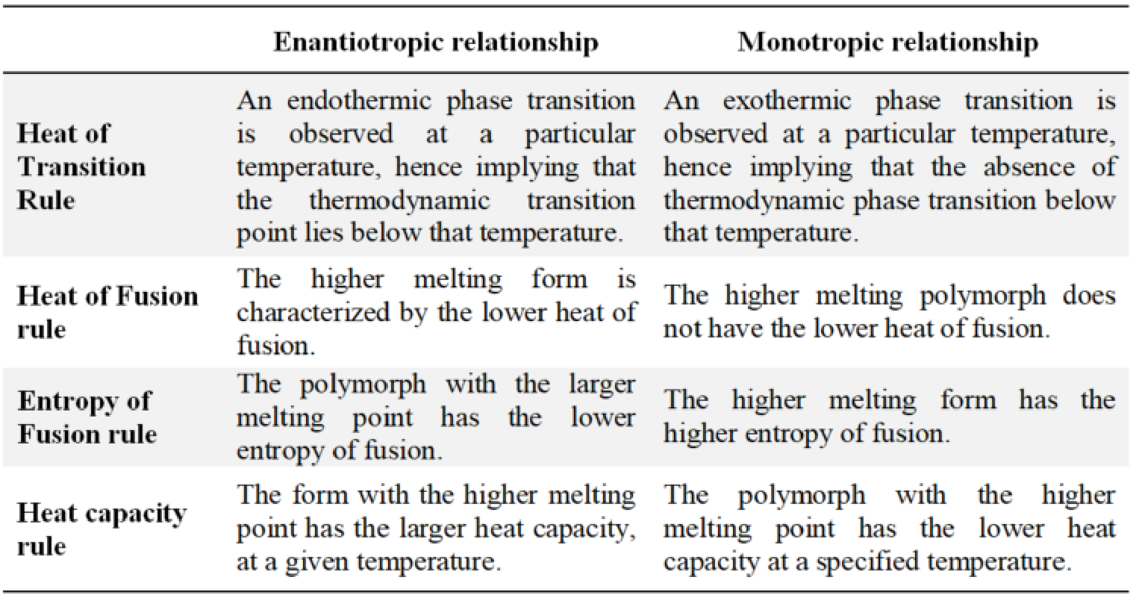


Figure 2.7: Graphs showing energy (arbitrary units) against temperature (K) of a dimorphic system at constant pressure, with form I is stable at room temperature. The thermodynamic relationship between the polymorphs is a) enantiotropic and b) monotropic.[29]

The graphs presented in figure 2.7 are typically utilized to map the thermodynamic relationships between polymorphs in a dimorphic system.[29, 39] As observed in figure 2.7a, when forms are enantiotropically related, their free energies, G_I and G_{II} are equivalent at the point of transition, $T_{p,I/II}$, implying that both forms are at equilibrium. In addition, this transition is accompanied by an endothermic change because the enthalpy of form I, H_I is lower than that of form II, H_{II}. It was shown how the transition temperature of enantiotropic polymorphic pairs can be estimated experimentally through solubilities or heat of solution, as demonstrated for carbamazepine and urapidil.[44] The thermodynamic relationship of monotropically related forms is illustrated in figure 2.7b. The curve associated with the free energy of form I is lower than the one of form II over all the temperature range, particularly also below the melting points, mp_I and mp_{II}. Pressure against temperature graphs can also be used to visualize the thermodynamic behaviour of polymorphic pairs, but using boiling points instead of melting points.

Table 2.2: Some of the standard rules proposed by Burger and Ramberger to predict the type of thermodynamic relationship between polymorphs in a dimorphic system.[45]



	Enantiotropic relationship	**Monotropic relationship**
Heat of Transition Rule	An endothermic phase transition is observed at a particular temperature, hence implying that the thermodynamic transition point lies below that temperature.	An exothermic phase transition is observed at a particular temperature, hence implying that the absence of thermodynamic phase transition below that temperature.
Heat of Fusion rule	The higher melting form is characterized by the lower heat of fusion.	The higher melting polymorph does not have the lower heat of fusion.
Entropy of Fusion rule	The polymorph with the larger melting point has the lower entropy of fusion.	The higher melting form has the higher entropy of fusion.
Heat capacity rule	The form with the higher melting point has the larger heat capacity, at a given temperature.	The polymorph with the higher melting point has the lower heat capacity at a specified temperature.

Burger and Ramberger formulated a set of rules, summarized in table 2.2, which can predict the thermodynamic relationship between polymorphs based on specific properties, such as the entropy of fusion or heat capacity.[45] These rules were tested and extended in various studies, such as the one presented by Qi *et al.* concerning the investigation of the thermodynamic stability relationship of sofosbuvir polymorphs.[46-47] In general these rules are known to produce correct predictions, excluding for specific cases such as having a substantial variation in the conformations of the polymorphs. Deviations from the heat of fusion rule predictions can arise either in examples with large difference in melting points (> 30 K) or if the enthalpy curves differ significantly from those portrayed in figure 2.6. On the contrary, the applicability of the heat capacity rule diminishes when the variance between the heat capacities of the forms is too minimal.

The relative thermodynamic stability of polymorphs in a dimorphic system, where form I is more stable than form II at constant pressure, can also be expressed in terms of their chemical potential. Moreover, if both forms are placed in contact with their respective saturated solution in any given solvent, the chemical potentials of the molecule in the liquid (μ_l) and solid (μ_s) phases become equivalent at equilibrium (μ_{eq}).[12] Therefore, assuming ideal solution conditions and combining equations 2.5 and 2.6 would result in an inequality in terms of solubility, x_{eq}. This inequality shows how form I is less soluble than form II, hence reflecting the relative thermodynamic stability, whereby the form with the lowest solubility is also the most stable polymorph.[12]

$$\mu_s(I) = \mu_l(I) = \mu_{eq}(I) = \mu^o + RTln(x_{eq}(I)) \qquad (2.5)$$

$$\mu_s(II) = \mu_l(II) = \mu_{eq}(II) = \mu^o + RTln(x_{eq}(II)) \qquad (2.6)$$

$$x_{eq}(I) < x_{eq}(II) \qquad (2.7)$$

This feature can be depicted in the graphs presented in figure 2.8, which show cases of dimorphic systems, whose thermodynamic behaviour can be extracted from such solubility graphs. An enantiotropic relationship is observed between 2,3,5-trimethyl-1,4-diacetoxybenzene (TMHQ-DA), with form A being the most stable at room temperature (figure 2.8a).[48] The solubility curves of paracetamol form I and II never intersect as illustrated in figure 2.8b, thus implying the presence of a monotropic system, with form I being the most stable polymorph over all the range of temperature.[49] However, this method of ranking has some disadvantages mainly due to the difficulty of determining the exact solubility, specifically at early developmental stage and also because the solubility ratio between polymorphic commonly ranges between 1 and 2.[41] This implies that such information would not always be enough to rank stability with high certainty.

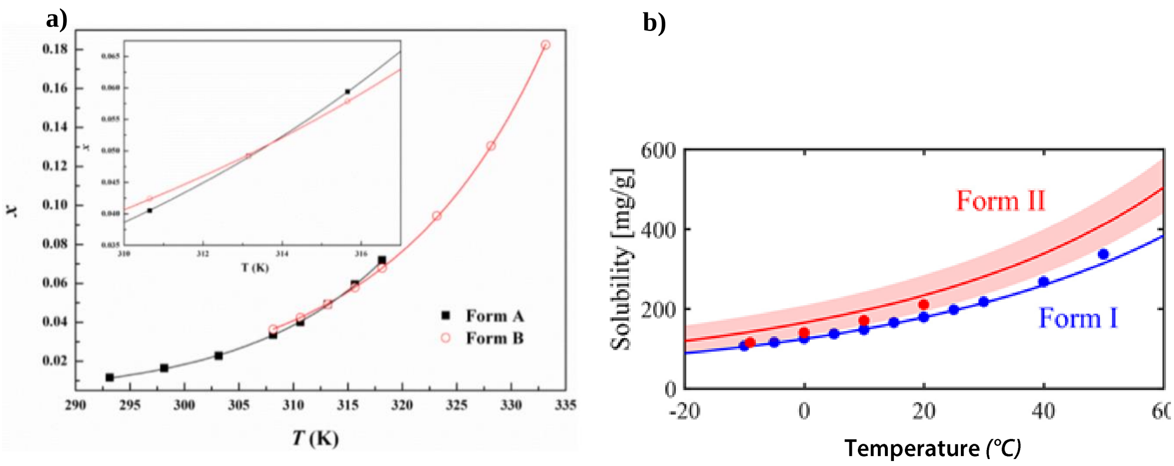


Figure 2.8: a) The experimental mole fraction of solubility against temperature (K) of TMHQ-DA in ethanol, exhibiting an enantiotropic relationship. b) Solubility of two polymorphs of paracetamol in ethanol, which are monotropically related, with the shaded red areas corresponding to the confidence interval.[48-49]

2.1.2.2 Characteristics of polymorphs

As mentioned previously, polymorphs typically have distinguishable intermolecular networks, hence implying that all the physical and chemical properties which depend on such networks are also different. Small intrinsic variations in the orientation or strength of particular contacts can be enough to produce a prominent alternation in the physiochemical characteristics listed in table 2.3. This phenomenon is exploited through crystal engineering, as one strives to create the polymorph with the optimal characteristics, without compromising its stability. However, given the difficulty to obtain such form, one might opt to find a balance between the two factors, while ensuring the safety of patients in case of APIs.

Table 2.3: Some of the properities which might be altered due to a change in solid form.[4]

Class	Properties
Chemical	Chemical stability, photochemical reactivity
Physical	Stability, Hygroscopicity, Morphology, Density, Colour
Mechanical	Compactability, Hardness, Powder flow, Tensile strength
Surface	Surface free energy, Stickness, Interfacial tensions
Thermodynamic	Solubility, Free energy of fusion, melting point, vapor pressure
Kinetic	Dissolution rate, Crystal Growth rates, solid- state reaction rate

Despite being associated with chemical instability, amorphous drugs tend to have better mechanical and physical properties, namely solubility and dissolution rates, due to their relatively higher free energy. For example, formulated tablets of metastable form I of metroprolol tartrate are characterised by lower porosity, leading to higher elasticity and compressibility than those of more stable enantiotropic polymorph II.[50] Better tabletability is also associated with the orthorhombic form II of paracetamol compared to the most stable monoclinic form I.[51] Research studies as the one presented by Capece *et al.* on acetaminophen, showed how promising techniques can assist such amorphous entities to retain their stability by inhibiting surface crystallisation.[52] Such studies provide hope towards a higher applicability of metastable forms, which can elevate the efficacy of several current pharmaceutical drugs.

2.1.2.3 The mystery of polymorphs

Given that there are various methodologies to obtain potential polymorphs, it is difficult to ensure the exhaustiveness of polymorph screening. Several intrinsic details of such transformations are still to be completely understood, hence decreasing the ability to estimate correctly the number of possible polymorphs of crystalline material and identify them. The competition between kinetic and thermodynamic factors, together with their dependence on crystallization conditions, adds additional layers of complexity. The multitude of such issue was already suggested by McCrone when noting, *"in general the number of forms known for a given compound is proportional to the time and money spent in research on that compound,"* and *"most chemists are almost completely unaware of the nature of polymorphism."* [31] Such sentiments were echoed by other researchers through time, provoking intense research work on polymorphism. Substances as axitinib, ROY polymorphs, galunisertib and aripiprazole lay out a perfect example of the possible extensivity of polymorphic landscapes.[37, 53]

Literature shows how efforts are made to control the outcome of crystallization, particularly to design ways to obtain the targeted forms. New methods are being devised in order to create routes to obtain elusive polymorphs, such as through the introduction of impurities during crystallization, as suggested by Kras *et al.* for the formation of benzamide form III in the presence of nicotinamide.[54] The monitoring of polymorphic transformations *in situ*, such as by using Raman hyperspectral imaging, can provide a better insight.[55] Furthermore, experimental techniques as templating via sublimation or crystallization under an external magnetic field, are being applied to navigate uncharted territories of polymorph landscapes.[56-57] Despite such efforts, cases of disappearing polymorphs are still encountered, and they cause doubts regarding the reproducibility of crystallization. A detailed review on disappearing polymorphs was published by Bučar *et al.*, discussing recent examples of ritanovir, paroxetine hydrochloride, rotigotine and sulfathiazole.[58] Such phenomenon of becoming almost unobtainable experimentally can also be observed in co-crystals polymorphs, as the case of caffeine-citric acid co-crystal.[59]

2.1.3 Multi-component systems: Co-crystals

Co-crystals and solvates are crystalline structures that contain a minimum of two distinct components, all of which remain uncharged. The main distinction between solvates and co-crystals concerns the physical state of the separate pure components at ambient conditions prior to the formation of the new multicomponent system. If one of the constituents is a liquid at room temperature, then the crystalline material is classified as a solvate, whereas for co-crystals, all components are in a solid phase at such conditions. Such difference causes several variations in their chemical and thermodynamic properties, as well as in the approaches taken to synthesize them. Both co-crystals and solvates are encountered in the pharmaceutical industry and represent an alternative way of how to engineer better APIs, but given the objectives of this project, the next section will focus on the former category.

Co-crystallisation represents an example where crystal engineering and covalent synthesis are linked together, through the formation of supramolecular synthons, composed of diverse contacts. Despite the lack of a standard definition, the USFDA proposed that co-crystals are "*solids that are crystalline material composed of two or more molecules in the same crystal lattice.*"[60] Molecular interactions between at least two components form the foundations of co-crystallisation, embodying the definition of a supramolecular synthon, formed by an extensive network of non-bonding contacts.[61] These contacts can range from strong H-bonds, characterised by a high extent of directionality, to relatively weaker ones as $\pi\cdots\pi$ or C-H$\cdots\pi$ (energy values are included in table 2.1). Despite their comparative weakness, the latter types of interactions might still have a collective influence on the self-assembly mechanism of the molecules, and on the adapted geometrical conformation.[61] Halogen bonding is also present in some co-crystals and it was noted how it aids the structure to extend at supramolecular level in all directions.[62,63]

In the pharmaceutical industry, co-crystals constitute the API (referred to as the host), together with one or more chemical entities, at a specific stoichiometric ratio. The introduced coformers need to be classified as safe for human consumption, as defined by the GRAS (Generally Regarded As Safe) list. A detailed review regarding co-crystallisation was published by Karimi-Jafari *et al.*, where they define, compare and discuss various co-crystal preparation mechanisms, in attempt to standardise such methods to increase their reproducibility.[64] Traditional methods are classified under solution-based (ex. solvent evaporation, slurry conversion method and anti-solvent addition) or solid-based (ex. neat grinding), but currently, novel procedures as spray drying technique are slowly emerging.

Even though one can find several publications concerning co-crystals, their lack of ease of production on an industrial level, together with the requirement of ensuring the constant robustness of the drug’s quality, still hinder their presence in commercialized pharmaceuticals. Scaling up definitely presents one of biggest challenges when it comes to the industrialization of co-crystal drugs, especially due to the difference in the nucleation and agglomeration occurring at lab-scale.[9] Despite such issues, the drive for co-crystal formation intensified through the last decade due to the discovery of benefits other than solubility manipulation, including mechanical property improvement and taste masking.[64] For example, an enhancement in the compaction behaviour was observed for paracetamol : oxalic acid co-crystals and resveratrol : 4-aminobenzamide co-crystals, when compared to the properties of the pure API.[65-66] One can also find cases where co-crystal systems are characterized by a greater extent of stability relative to the pure counterpart, such as indomethacin-saccharin co-crystals which exhibited lower water sorption during humidity studies.[67] Such potential advantages motivate research towards the optimization of computational methods to produce more accurate predictions, in order to guide lab-work towards investing in API-coformer combinations with the most suitable properties.

2.1.3.1 Stability of Co-Crystals

The formation and stability of co-crystals are based on the probabilistic nature and energetic viability of the supramolecular synthons created to build the crystal structure. In return, such synthons are dependent on the molecular conformation of both host and coformer, which can be influenced by the chemical environment of the synthesis process, particularly the polarity and pH value of the solvent used. Consequently, this leads to an extensive increase in the structural and energetic landscape surrounding a co-crystallisation event. The synthons can be composed of either homo- contacts which are located between two identical functional groups, or hetero- interactions where the contributors are distinct groups.

Table 2.4: Labels used to describe the location of hydrogen bond in a co-crystal.

Interactions	Active Component, form I (A)	Coformer (C)
Active Component, form I (A)	Pure	Impure
Coformer (C)	Impure	Pure

The process of co-crystallisation is rather complex, with energetic and entropic processes occurring in all three directions. However, for simplification purposes, it can be considered as a product of the competition of the energetic favourability of interactions between those within each individual participant (denoted as pure interactions) and those between the two components (defined as impure). The study of such favourability is the centre of several approaches focused on predicting or understanding co-crystallisation. The resultant energy change accompanying the breakdown of the non-bonding network within each contributor and the build-up of the new one between them, determines the relative stability of a co-crystal. This change might be characterised by high variations in enthalpy and entropy, which can in return compensate for each, and still result in a minimal change in Gibbs free energy. Due to such effect, very often only the change in Gibbs free energy is used to determine realistically the stability of any product of co-crystallisation, $A_aC_{c\,solid}$, where the first component (ex. API) is denoted by A, while the coformer is defined by C.[68]

$$A_aC_{C\,solid} \rightleftharpoons aA_{solution} + cC_{solution} \tag{2.8}$$

Given the above equilibrium, the standard free energy change, ΔG^{o} can be given by the following:[68]

$$\Delta G^{o} = -RT\, ln\left(\frac{(a\prime_{A})^{a}(a\prime_{C})^{c}}{K_{sp}}\right) \approx -RT\, ln\left(\frac{(S_{A})^{a}(S_{C})^{c}}{K_{sp}}\right) \qquad (2.9)$$

Where a'_{A} and a'_{C} denote the respective thermodynamic activity of the participants, while K_{sp} is the solubility product. This representation can be approximated by using the solubility of pure components A, S_{A} and C, S_{C}. Thermodynamic expressions become more complex when introducing a tertiary component (solvent or solid- state compound) which can potentially form part of the crystal structure of either pure components or the co-crystal itself. Ternary isothermal phase diagrams of these systems are utilised to understand the broadness of such landscape, as well as to identify the conditions which favour the required product.[69] Specific strategies as the one presented by Dubey and Desiraju for the preparation of ternary co-crystals of quercetin, illustrate viable ways how to navigate through extensive co-crystal landscapes to obtain the wanted solid form.[70]

2.1.3.2 Salt or co-crystal – Proton position

In the pharmaceutical context, a salt is a multicomponent system composed of an ionisable API and counterion (atomic or molecular), with a charge balance and subsequently, a definite stoichiometry.[71] Therefore, in the presence of basic and acidic functional groups located on both components, as in the case of antivirals under study, one can speculate on whether the formation of a new product is a salt or a co-crystal. Despite of only differing with respect to the position of a proton (within a molecular context), this difference is highly probable to cause discrepancies in the crystal lattice and properties.[72-73] Nonetheless, the degree of such discrepancy is dependent on the extent of proton transfer during crystallisation, with possible crystal structures also able to exist along the salt-co-crystal continuum (not necessarily complete or absent proton transfer).[74] The extent of proton transfer is affected by the acidic and basic group strengths, as well as the chemical environment during crystallisation. In general, the product of an acid-base reaction is expected to form a salt if ΔpK_a (pK_a (base) - pK_a (acid)) is greater than three, while a negative value will propose a co-crystal formation.[74-75] However, it is not possible to make any deductions if this parameter falls within the in-between range of values.

The most accurate way to distinguish between co-crystal and salt is by identifying the location of the proton involved. Single crystal XRD remains the most useful tool for locating atomic positions, especially at low temperatures, however this can be more challenging for hydrogen atoms due to their weak diffracting ability.[76] Therefore, additional characterisation techniques as vibrational spectroscopy, and solid-state NMR are very frequently employed to provide further information.[76] This process is often assisted by molecular modelling and by comparing the geometric features (bond length and angle) of the moieties involved with structural information obtained from experimentally determined crystal structures. For example, the C-N-C angle in a protonated pyridine group is likely to be above 120°, whereas its deprotonated analogue is more inclined to have a bond angle below 120°.[72] In case of carboxylic/carboxylate, it has been demonstrated that due to the delocalisation of the negative charge, C-O bond lengths in the carboxylate ion have a higher within group similarity relative to those in carboxylic group.[76]

2.2 Computational approaches for prediction

The extensive broadness of crystal energy landscapes (single or multi-component) can be perceived as a major challenge with respect to form stability assurance, but also an opportunity for crystal engineering. The laborious demands of both endeavours, together with imminent need of vaccines against the rapid progression of diseases, caused a paradigm shift within the approaches taken towards drug and dosage form development. This shift was accompanied by an increase in the contribution of computational techniques, whose advancements aid predictions to converge towards reality. Indeed, the main objective of computational methodologies is to create realistic simulations, ideally embodying all the resultant characteristics of a collection of simultaneous factors that work with or against each other. This target remains the main driving force behind computational developments which continue to add on further refinements and considerations to crude representations, while taking into account time and costs.

2.2.1 Crystal Structure Prediction

Crystal structure prediction (CSP) incorporates any computational methodology which strives to predict the atomic fractional coordinates, cell dimensions and space group of a crystal structure, should this be synthesised and solved by diffraction.[77] The complexity of this task increases with the risk of polymorphism, since it broadens the multitude of variable parameters within the problem, especially in the presence of highly flexible organic compounds or multi-component systems. Beyond the prediction of the crystal structure parameters (and maybe more importantly for this project), this methodology provides a unique insight of the thermodynamic profile of a solid-state species, which is crucial for its potential application as discussed above. This approach for prediction stands upon the theory that the ranking of all potential structures based on their respective lattice energy at 0 K, can illustrate their probability of existence, with the experimentally observed crystal structure coinciding with predicted structure having the global minimum lattice energy.[78] This implication demands the accurate evaluation of the thermodynamic stability of all potential crystal structures, especially when the area above the global minimum is populated by multiple local minima, as observed in highly polymorphic substances, as the notable ROY polymorphs. Given the importance of crystal structure prediction, and its broad applications, one can find various theoretical approaches and techniques in literature, together with blind-tests and benchmark studies which try to order their accuracy and efficiency.[79-82]

2.2.1.1 Overview of CSP methodology

The central engine in CSP is the evaluation of the lattice energy, E_{latt} which depends on both the intramolecular and intermolecular energies within the crystal structure, and can be estimated through the following expression. This definition represents a static model, hence neglecting any kinetic contributions.

$$E_{latt} = U_{inter} + \Delta E_{intra} \qquad (2.10)$$

The intermolecular energy, U_{inter} is the dominant contributor towards the total lattice energy, and a proper quantitative evaluation of such effect is crucial. This is because the lattice energy is defined as the energy of formation of a crystal, starting from gas-phase molecules, and hence implying that the construction process is done through intermolecular bonding.[83] The second component of the lattice energy is the molecular energy conformation penalty or relaxation energy, without which one would be overestimating the final value. The intramolecular energy penalty, ΔE_{intra} is the energy difference between the conformations in the crystal structure and in the ideal gas-phase, and it absorbs any molecular distortion caused by packing forces (commonly calculated through DFT methods). Despite typically having a small magnitude especially for rigid molecules, its influence increases in cases where the molecule is able to form intramolecular H-bonds. When determining such energy difference of known structures, the molecular conformation from the crystal lattice is sometimes allowed to relax partially before proceeding with the energy calculation. This precaution is performed due to the presence of probable inaccuracies in the structure solution, especially related to the position of hydrogen atoms, which might lead to an overestimation of the actual energy value.[84] The precise determination of the intramolecular conformational energies, preferably calculated with high-level electronic structure schemes, is considered a vital key factor for the correct ranking of polymorph stability, as shown in figure 2.9 for axitinib polymorphs.[85]

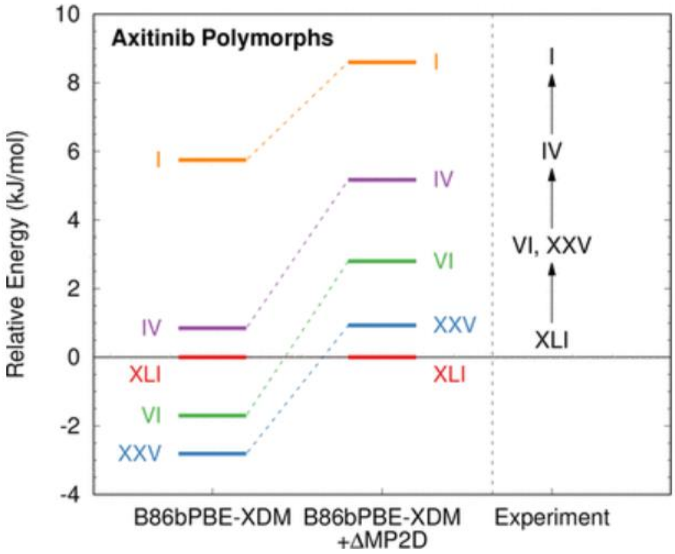


Figure 2.9: The re-ranking of the relative 0 K lattice energies for axitinib polymorphs, in comparison to the stability order deduced from experimental data. The increase in compatibility between the predicted and experimental stability order is due to the MP2D correction based on accurate conformational energy evaluations.[85]

Given the computational cost and multitude of generated hypothetical crystal structures, the level of accuracy for the determination of the lattice energy is not equivalent in every stage. The underlying factor of both accuracy and computational time is the approach encompassing the molecular modelling of the system, which can range from molecular mechanics (MM) based on the classical Newtonian principles, in conjunction with empirical potential functions to *ab initio* methods as Hartree-Fock (HF) self-consistent field theory or density functional theory (DFT). The selection of the type of scheme for molecular modelling at each stage, and the route taken for structure generation are the strongest contributors towards the variation between methodologies for CSP. However, even though each computational algorithm, such as GRACE, MOLPAK, CrystalPredictor and Polymorph Predictor, is characterised with its own set of assumptions and methodology, the core stages of CSP remain fundamentally fixed.[86-88]

A. Appropriate sampling of the conformational space

The initial molecular conformation used as an input for CSP is crucial because it dictates the range of possible variations in the shape and packing capabilities, hence introducing a significant bias with respect to the generation of potential crystal structures.[78] In order to allow the algorithm to be guided without any experimental bias, an isolated molecular conformation is utilised as an input, instead of a conformer in a known crystal lattice. In case of having flexible molecules, as galunisertib in figure 2.10, the rotatable bonds present need to be added as search variables, so as to locate all possible low-energy conformers and use each one as a separate input for CSP.[85]

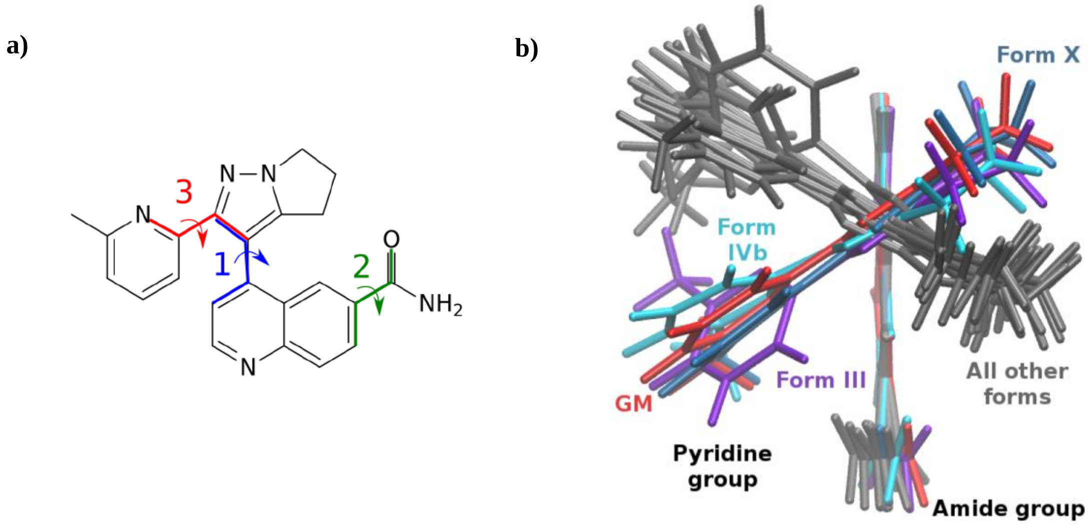


Figure 2.10: Galunisertib molecular structure; a) 2D illustration, highlighting 3 rotatable bonds and b) an overlay of monomers present in different forms, aligning them onto the quinoline ring.[85]

The identification of flexible bonds, in conjunction with their accessible conformational space can be achieved through potential energy scans along each torsion angle, while keeping the rest of structure rigid. It is crucial to note that not every rotatable bond exhibits the same degree of flexibility. Such torsion angle ranges can then be used as constraints for geometry optimisation in order to locate the lowest-energy conformer from each accessible conformational region. Another possible route for the mapping of the molecular flexibility is through the use of databases as the CSD, which allows for the measurement of the targeted torsion angle in experimentally determined structures with chemically equivalent environments. More complex schemes, as the one present by Iuzzolino *et al.* integrate both approaches; the use of information from CSD torsion-angle distributions with *ab initio* calculations so as to reduce the number of conformers needed to represent all the accessible conformational space, while also taking into account conformers which might not necessarily have the lowest energy.[89] This consideration is very important because the molecular conformation in the crystal structure is very often distorted by the crystal packing forces, hence leading to higher energy values.[90]

The initial energy calculation of numerous potential conformers is very often conducted through Molecular Mechanics (MM) due to the relatively fast computation. The central variation in classical approaches revolves around the functional forms chosen to describe the different types of potential energies, together with the specific set of parameters according to the atom types.[91] However, given the importance of the starting conformation, this is refined at higher level of accuracy, usually through DFT calculations, with dispersion corrections (DFT-D).[92-93] The dispersion corrections developed by Grimme and Tkatchenko-Sheffler are widely used in current research DFT studies.[94-95]

B. Appropriate sampling of the crystallographic space

The number of molecules in the asymmetric unit, Z' and the selection of space groups, are parameters which influence the extent of crystallographic space probed during crystal structure generation. In case of taking into account Z'>1 or even further disordered structures would increase the amount of hypothetical structures exponentially, leading to a practically impossible task to solve.[78] The completeness of the search is very often restricted due to computational costs associated with the calculation of the lattice energy of each hypothetical crystal structure. The choice of space groups can be limited based on the chemical nature of the molecule (only a few space groups can be adopted by a chiral molecule) and on the frequency of space groups adopted by chemically relevant molecules in databases as the CSD.[78]

As stated previously, every computational algorithm has its own specific method for structure generation and lattice energy evaluation. For example, Polymorph Predictor, available in Material Studio, utilises a Monte Carlo packing simulation which searches for the global minimum of each chosen space group through simulated annealing, which reduces the probability of becoming trapped in a local minimum by allowing the temperature to be relatively high before cooling.[87, 96] In principle, this will provide enough energy to the crystal structure to overcome any energy barriers that might prevent it from reaching the global minimum during cooling. The generated crystal structures are accepted or rejected based on the Metropolis criterion. Consecutive clustering based on crystal packing similarity and geometry optimisations are conducted, until obtaining the final set of potential crystal structures, ranked according to their lattice energy. Another algorithm which makes use of a Monte Carlo packing simulation, but with different settings and energy evaluations, is the GRACE computational package, developed by Neumann and co-workers.[88] This algorithm resulted to be the most successful route for CSP during the latest complete sixth blind-test organised by the CCDC (the seventh blind-test is currently underway).[81]

The reduction of packing arrangements can be based on specific cut-offs for densities and/ or lattice energies, as implemented in the CrystalPredictor algorithm, which initially generates potential crystal structures by using random values for cell parameters within the parameters permissible by space group symmetry constraints.[86] The reduction of duplicate structures can also be conducted by an automated comparison of simulated powder diffraction patterns, as described by Taylor *et al.* during their CSP for isoniazid and iproniazid.[92]

C. Modelling Lattice energy and Ranking

The lattice energy calculation of the hypothetical crystal structures is a multistage process of successively higher-accuracy level of theory methods, possibly ranging from traditional force fields to expensive periodic DFT calculations. During the first stage in CSP, which compromises a large amount of potential structures, classical simulations are usually employed, using atom-atom methodology to model intermolecular forces. This approach represents the molecular charge distribution through isotropic point charges. For example, the atom-atom *6-exp-1* potential, eq. 2.11 combines the Buckingham potential which accounts for the repulsion and dispersion forces, together with an additional Coulombic term to represent the electrostatic potential.[91]

$$U_{inter} = \sum_{i>j} U_{ij} = \sum_{i>j} A\exp(-BR_{ij}) - CR_{ij}^{-6} + \frac{q_i q_j}{4\pi\varepsilon_0 R_{ij}} \qquad (2.11)$$

where R_{ij} refers to the intermolecular distance between nuclear positions i and j, together with their respective atomic charge parameters q_i and q_j. Coefficients A, B and C are empirical parameters derived particularly from heats of sublimation and evaporation, whereas ε_0 is the vacuum permittivity.[91] Given that the number of interactions grows rapidly, especially when including many-body dispersion (MBD) interactions, cut-offs are introduced to speed up computations. Alternative schemes to simple radial cut-offs concern the inclusion of long-range effects explicitly using the Ewald Summation method or the more advanced mesh-based Ewald methods, which are particularly useful in the presence of periodic systems.[97]

In case of Polymorph Predictor, the energy calculation workforce is based on force field technology, providing options as COMPASS 3.0 force field which utilises quantum mechanically derived parameters for the valence energy computation and applies a modified version of the atom-atom Lennard-Jones potential for the non-bonding energy term, eq. 2.12 (the inverse 9th power to represent the repulsion forces present instead of the usual 12th power).[98-99]

$$U_{inter} = \sum_{i>j}\left[D_{0,ij}\left(2\left(\frac{R_{0,ij}}{R_{ij}}\right)^9 - 3\left(\frac{R_{0,ij}}{R_{ij}}\right)^6\right) + \frac{q_i q_j}{4\pi\varepsilon_0 R_{ij}}\right] \qquad (2.12)$$

where $D_{0,ij}$ and $R_{0,ij}$ are derived from atom-based parameters as defined by R.L.C Akkermans *et al.*[99] The superior performance of COMPASS force field relative to other force fields was demonstrated by Marchese Robinson *et al.*, where they tested their ability to calculate the lattice energy of a large dataset of pharmaceutically-relevant compounds through different protocols.[79]

As an improvement from classical empirical force fields, the GRACE algorithm incorporates a tailor-made force field (TMFF) for the first step in CSP, using parameters derived from DFT-D calculations on several molecular conformations, hence eliminating the possible issues of transferability.[100] The construction of this force field on molecule-by-molecule basis allows for a more accurate evaluation of the crystal structure energy, hence leading to a reduction in the number of hypothetical packing arrangements present within the energy window around the global minimum, especially when dealing with flexible molecules. This approach was able to predict correctly the experimental structure of cyclohexane-1,4-dione, even without requiring further refinements in terms of energy minimisation.[100]

Alternatively, several CSP cases adopt a hybrid method where the conformational geometry and energy are estimated using electronic structure analysis by DFT calculations, together with an intermolecular atom-atom model potential.[101-102] Furthermore, molecular charge density can also be represented through an expansion of multipole moments, enabling the model to reflect the influence of anisotropy in the electron charge density on the intermolecular forces.[103] In this approach, usually noted by distributed multipole analysis, point charges, dipoles, quadrupoles and higher multipoles are scattered throughout the molecule, usually located at the atomic positions. Typically, multipoles are calculated from quantum mechanics wavefunctions, using Gaussian basis functions. The improvement in accuracy through the use of distributed multipoles was found to be more pronounced for structures with a minimal degree of flexibility, indicating that it is more effective to describe intermolecular rather than intramolecular interactions.[104]

The multitude of studies carried out on the comparison and evolution of interatomic potentials offer a selection of force fields for a large variety of atom types, such as W99 force field and FIT force field.[80, 105-106] There are various versions of such force fields using the empirically parameterised Buckingham potential for the non-electrostatic interactions, but differing in the way that they describe the molecular charge density (ex. multipoles or molecular electrostatic potential derived atomic charges). As shown in the benchmark assessment conducted by Nyman *et al.*, hybrid force fields employing anisotropic atom-atom multipole systems can provide the same accuracy as commonly used DFT-D as displayed in figure 2.11, with an error margin two to three times larger than the most sophisticated DFT-D techniques.[107] The success of algorithms as DMACRYS, which also involves anisotropic repulsion potentials, encourages further implementation and optimisation of this technique.[108-109]

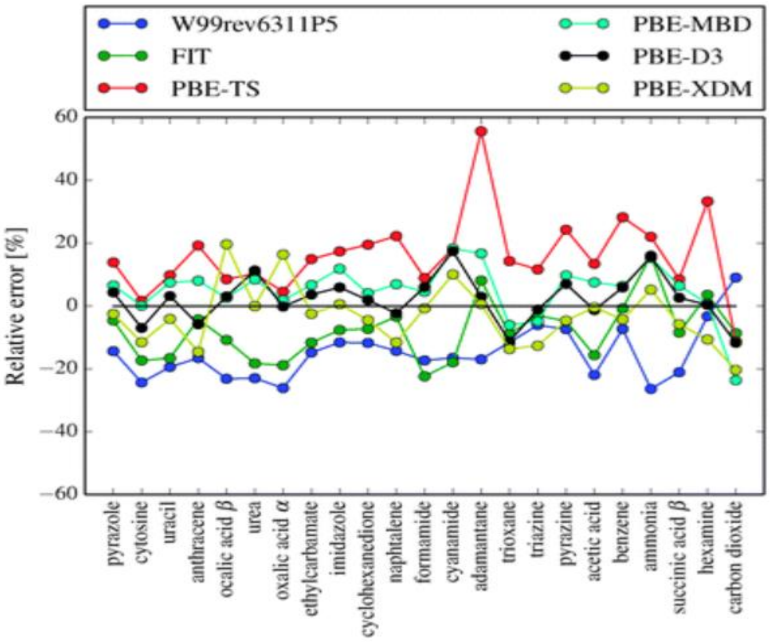


Figure 2.11: Relative error in lattice energies evaluated by dispersion correction PBE calculations and two hybrid force fields, as presented in the benchmark study by Nyman *et al.*[107]

Extrapolating the atom-atom concept but deviating from the sole dependence on experimental data, the intermolecular energy in a system of molecules can also be expressed as the summation of all pairwise intermolecular potential energies. One of the most established hybrid methods is the PIXEL scheme, proposed by Gavezzotti, using a semi-empirical forcefield.[110] In this approach, the molecular electron density is dissected into a number of pixels, each with their corresponding density and charge, calculated through Mulliken population analysis.

$$E_{total} = E_{Coulombic} + E_{Polarisation} + E_{Repulsion} + E_{Dispersion} \qquad (2.13)$$

This methodology enables the separate evaluation of different contributors of intermolecular forces (eq. 2.13), each of which is a result of the collective effect of pairwise pixel-pixel, nucleus-nucleus and pixel-nucleus interactions between distinct molecules. The PIXEL scheme was proved to be successful in reproducing enthalpy of sublimation values and crystal structures of considerable amounts of crystal structures, as well as accurately predicting the heats of evaporation of common organic liquids, hence validating its role in both static and evolutionary molecular simulations.[110-111]

Similar to the formalism present in the PIXEL method, Spackman and co-workers developed a model (CE-model) for the evaluation of intermolecular interactions through energy partitioning via *CrystalExplorer* package.[112-113] Apart from the energy decomposition, the two methods also have a comparable approach for the determination of non-empirical Coulombic energy, E'_{col} as well as the inclusion of adjustable parameters, k. Sajesh *et al.* presented a study aimed to investigate the lattice energies and conformational energy differences of the ROY polymorphs using this approach, including ΔE_{intra}.[84] The results displayed in table 2.5 show the relative lattice energies of seven ROY polymorphs, produced by different computational methods including the PIXEL scheme and the CE-B3LYP model, in comparison with experimental data. It is evident that while the CE-B3LYP model shows improved ranking ability with respect to other approaches, the absence of vibrational energy term considerations still had an impact on the accuracy of the relative lattice energies of such flexible polymorphs.

Table 2.5: Lattice energies (kJ mol^{-1}) of seven ROY polymorphs relative to form Y which is the most thermodynamically stable form, presented in different studies.

	Y	YT04	R	OP	ON	YN	ORP
Dunitz and Gavezzotti [110]	0.0	–	7.0	16.0	8.0	18.0	22.0
$E_{lattice}$ from Li *et al.*[114]	0.0	9.5	2.7	3.6	-4.1	7.8	-7.7
Vasileiadis *et al.*[115]	0.0	2.3	2.1	3.3	9.3	-1.0	7.6
Habgood *et al.*[116]	0.0	2.4	-0.9	0.6	2.1	-0.6	2.0
Sajesh *et al.*[84]	0.0	5.2	9.1	9.4	12.9	8.9	12.2
Experimental enthalpies[37, 117]	0.0	0.9	1.4	1.9	2.6	3.0	4.1

Lastly, due to their success in the latest blind-tests, periodic DFT calculations are being widely used in recent CSP studies, especially during the final refinement stage, because they provide improved energetic ranking compared to forcefields, albeit at a much greater computational cost.[81-82, 93, 118] Indeed, periodic boundary conditions are applied so as to be able to take into account intermolecular forces at longer ranges by using only unit cell and space group information, rather than actually performing calculations at much higher dimensions. The final optimization through this scheme also allows for the relaxation of the intramolecular degrees of freedom, together with the unit cells parameters. Other electronic structure techniques apart from periodic DFT methods, entail periodic second-order Moller-Plesset perbutation theory, diffusion Monte Carlo and fragment-based electronic structure schemes.[119]

D. Re-ranking due to thermal contributions

In recent years, in conjunction with the lattice energy, studies are incorporating also vibrational contributions dominated by entropic effects present at ambient conditions as the last stage for refinement.[92-93, 107, 120-122] Even though such considerations have a small magnitude, they can still be enough to cause re-ranking especially when taking into account that on average, the energy difference between polymorphic pairs is less than 2 kJ/mol.[93] Furthermore, given that the experimental value for the enthalpy of sublimation, ΔH_{subl} is used as a direct measure of the lattice energy, the zero-point vibrational energy difference between the ideal solid and ideal gas, ΔE_{ZPE} and thermal corrections need to be included to compensate for the factors acting at a finite temperature.[79, 120, 123] The thermal corrections can be estimated by determining the difference in heat capacities, ΔC_p between the gas and crystal, respectively. A simplified version of eq. 2.14 would approximate ΔE_{ZPE} to zero and estimate ΔC_p by – *2RT*, where R is the molar gas constant.[123]

$$\Delta H_{subl} = -E_{latt} + \Delta E_{ZPE} + \int_{0\,K}^{T} \Delta C_p \, dt \qquad (2.14)$$

Furthermore the vibrational contributions, $F_{vib}(T)$ can also be incorporated so as to obtain the Helmholtz free energy, *A(T)* as follows:

$$A(T) = E_{latt} + F_{vib}(T) \qquad (2.15)$$

The evaluation of the vibrational contributions through periodic electronic structure calculations can be executed by applying the harmonic approximation and integrating over the Brillouin zone at specific temperatures.[122] The accuracy of the evaluation of these contributions is dependent on the quality of the calculated vibrational frequencies and on the appropriate sampling of the reciprocal space. The convergence of the phonon density of states in crystal structures would require the inclusion of several **k**-points in the calculations, in a way to sample the first Brillouin zone effectively, often using supercell expansion in different directions to capture the anisotropic characteristic of phonon dispersion.[107]

The traditional reason for the neglect of entropic contributions is based on the fact that usually during CSP, one is interested in the relative crystal energies, and hence assuming that such contributions will cancel out. Moreover, lattice energy values estimated from empirically derived force fields would have already partially accounted for thermal effects because the potential parameters used were originally optimized by fitting experimental data.[101] Nevertheless, the neglect of such considerations might still cause incoherencies with experimental ranking, especially when the error in the lattice energies is greater than the differences between the relative energy values.[93]

2.2.1.2 Multi-component CSP

A route for the prediction of multi-component systems is by comparing the thermodynamic stability of the hypothetical crystal structure with respect to its single-components through CSP. This methodology is based on the assumption that co-crystallisation would only occur if the predicted co-crystal is more thermodynamically stable than its pure counter-parts.[124] The introduction of a foreign molecule within the structure search makes CSP of multi-component systems very demanding and complex due to the exponential increase in possible structures for the exhaustive probing of the conformational space and also the persistent requirement for accurate energy evaluation.

The reliability of using the thermodynamic stability as a predictive tool for the formation of multi-component systems, was demonstrated by Chan *et al*, who investigated 102 salts and co-crystals of isonicotinamide, nicotinamide and picolinamide.[125] They showed how lattice energy calculations performed through the use of first-principles plane-wave based DFT method, in conjunction to PW91 exchange-correlation functional can predict the formation of a co-crystal or salt with a success rate of 97 %. A more robust study later re-confirmed the predictive ability of this approach by a comparative periodic DFT study based on 350 organic co-crystals, which showed how in general they are more stable than their respective pure components with an average of 8 kJ mol^{-1}, while less than 5 % of the cases stated otherwise.[126] This study also highlighted how simple qualitative descriptors as the changes in packing efficiency or count of hydrogen bonds do not necessary fully capture the driving force behind the mechanism of co-crystallisation.

2.2.2 Solvation models: Conductor-like Screening Models

One of the challenges in drug development revolves around low solubility, which would result in poor absorption and limited bioavailability, as in the case of ganciclovir.

Therefore, an imperative task for computational techniques is to simulate and study solvated systems in attempt to provide accurate predictions of solubility and thermodynamic properties for drug development.[127] Such simulations are very often based on solvation models, which seek to capture the actual chemical environment of solvation, using descriptions and assumptions which vary according to the employed model. A particular category of solvation models, known as continuum models, constitutes in using continuous distribution functions to define solvent molecules as a continuous medium, rather than treating each one of them explicitly. Such conceptual framework aims to realistically represent the electrostatic screening of the solute within the dielectric continuum portraying the solvent molecules.[128] This is done by embedding the solute molecule in a virtual finite space, noted as cavity, derived from Onsager's model, and surrounding it by such dieletric continuum.[127, 129] The shape of the cavity is usually expected to represent the outer shape of the molecule, so as to avoid the presence of anomalies in the charge distribution after solvent polarisation. In general, the dimensions of the cavity are adjusted according to either a specific electronic isodensity level or to the vdW atomic radii of the solute, as illustrated in figure 2.12.

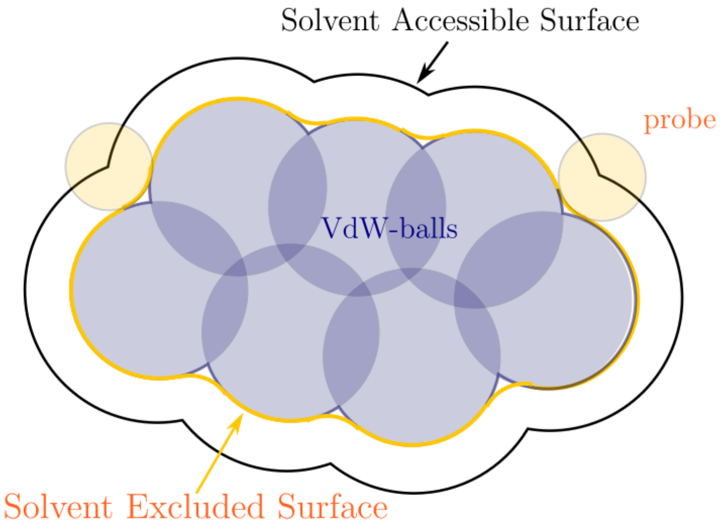


Figure 2.12: The solvent accessible surface is depicted along the centre of the probe representing the solvent molecule nearest to the solute, without overlapping.[127, 130]

The COSMO adopted the concept of representing the solvent molecules as a dielectric continuum, but restricting the source of the reaction potential to a closed surface, as in typical apparent surface charge (ASC) approach.[131] These models mirror several characteristics which define ASC methodologies, particularly in the iterative approach taken to obtain geometry convergence. In COSMO, given that the solute molecules are screened in a conducting continuum, a polarisation charge density, σ is induced on the molecular surface in contact with the solvent medium. As a consequence of such induced charges, a more polarised charge density is built across the solute molecule, and this iterative quantum chemical process proceeds until the solute has the optimal energy with respect to the charge electron density.[132] In order to simulate the continuous conducting medium, the dielectric permittivity, ε is set to infinity. The effect of a finite value for permittivity is re-introduced by scaling the ideal unscreened charge density, using ε of the particular solvent, in conjunction to a constant associated with the cavity shape of the solute.[131] The mapping of the induced opposite charge density, also referred to as surface screening charges, can be visualised as a σ-surface, where these charges are mapped onto the molecular cavity, as depicted in figure 2.13.

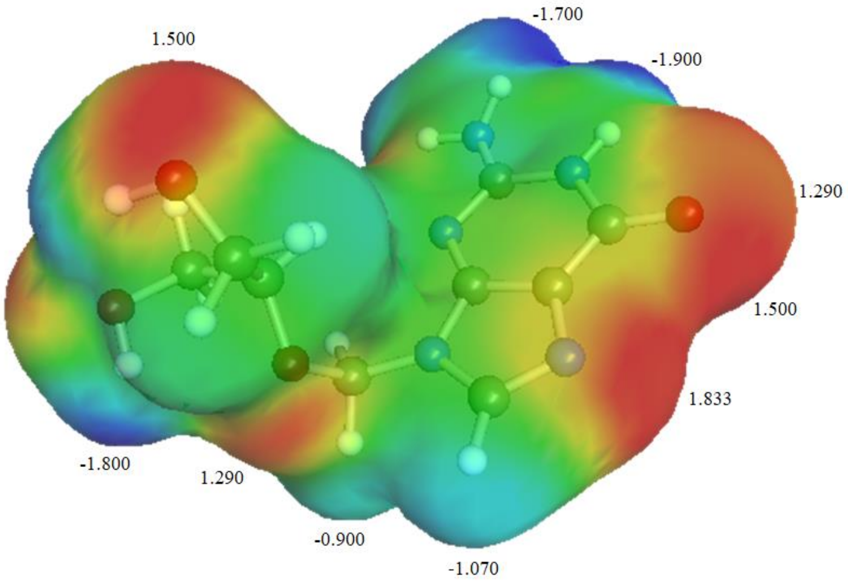


Figure 2.13: The σ – surface of ganciclovir, calculated by COSMO*quick*, together with some σ-charge density values in e/nm^2. The red areas illustrate positive COSMO charge densities, with the underlying negative molecular charges, while the blue segments portray negative COSMO charge densities which are caused by underlying positive molecular charges.[133]

2.2.2.1 Conductor-like Screening Models for Realistic Solutions

Despite the improved results obtained by COSMO, especially when dealing with water as a solvent, it has some limitations concerning its inability to differentiate between solvents with similar dielectric constants but distinct chemical environments, such as benzene and cyclohexane.[134] Furthermore, the assumption of linear polarizability, represented collectively by a static dielectric constant, highly discounts the actual fluctuations in the polarizability of polar solvents, and its effects on the solute molecules. Such restraints led to the development of COSMO for Realistic Solutions, which incorporates the basics of dielectric continuum models with a statistical thermodynamic treatment for interacting surface segments.[135]

In COSMO-RS theory, the reference state is defined as an ensemble of closely packed molecules, each of which is still perfectly screened by the conducting continuum.[134] The cavity of each solute molecule is composed of a set of surface segments which can be defined by their own position, area and surface charge, σ induced by its surrounding environment (which in case of the reference state is a virtual conductor). This leads to the dissemination of all possible contacts between molecules into pairwise interactions, occurring at each contact area, where each σ has its σ' counterpart associated with its interacting segment on the neighbouring molecule. For COSMO for real solutions, the virtual conductor is removed, hence giving rise to electrostatic energy, ΔE_{misfit} whenever there is a net screening charge density ($\sigma \neq \sigma'$).[132] The misfit term is only negligible in case of having pair segments with the same charge density, $\sigma = -\sigma'$, because it implies that they are being screened ideally as the reference state. [134] The computation of ΔE_{misfit} and the hydrogen bonding energy contribution, ΔE_{HB} requires the effective surface contact area, a_{eff}, together with constants c_{misfit} and c_{HB}, respectively.[134]

$$\Delta E_{misfit} = a_{eff} c_{misfit} (\sigma + \sigma')^2 \qquad (2.16)$$

$$\Delta E_{HB} = a_{eff} c_{HB} min(0, \sigma\sigma' - \sigma_{HB}^2) \qquad (2.17)$$

$$E_{total} \cong \sum_k a_{eff} \left[e_{misfit}(\sigma_k, \sigma_k') + e_{HB}(\sigma_k, \sigma_k') \right] \qquad (2.18)$$

The construction of the definitions of the energy terms in equations 2.16 and 2.17 proves how the total interaction energy in solvation is a product of a collection of surface contacts, k characterised by the local charge densities of their respective pairing segments (eq. 2.18). This implies that macroscopic thermodynamic properties can be described by averaging the energy of the permanently fluctuating ensemble of microscopic pairwise interactions through statistical thermodynamics.[136]

The screening charges of the solute molecule X viewed on a σ-surface, as in figure 2.13, can be plotted graphically using a probability distribution function to construct a σ-profile, $p^X(\sigma)$ (eq. 2.19). The notations $n_x(\sigma)$ and $A_x(\sigma)$ describe the number of segments with a specific charge density, σ and their total surface area, respectively.[137] In presence of mixtures, the resultant profile would be a weighted average of the separate σ-profiles of each pure component, based on the mole fractions. The σ-profile of solute molecule X, and the σ-potential, $\mu_S(\sigma)$ (visualised in figure 2.14) which offers a description of the affinity of solvent, S for the specific solute molecular surface, can be utilised to evaluate the σ-potential of solute X in solvent S, μ_S^X. This is executed by integrating eq. 2.20, which is a function of charge density, concentration and temperature, over all the surface of the solute. The dependence of μ_S^X on its mole fractions and the influence of relative sizes of solvent and solute molecules are taken into consideration and corrected by the logarithmic term, referred to as the combinatorial term.[134]

$$p^X(\sigma) = \frac{n_x(\sigma)}{n_x} = \frac{A_x(\sigma)}{A_x} \tag{2.19}$$

$$\mu_S^X = \int d\sigma\, p^X(\sigma)\mu_S(\sigma) + kTln\{x + \gamma^X_{comb,S}\} \tag{2.20}$$

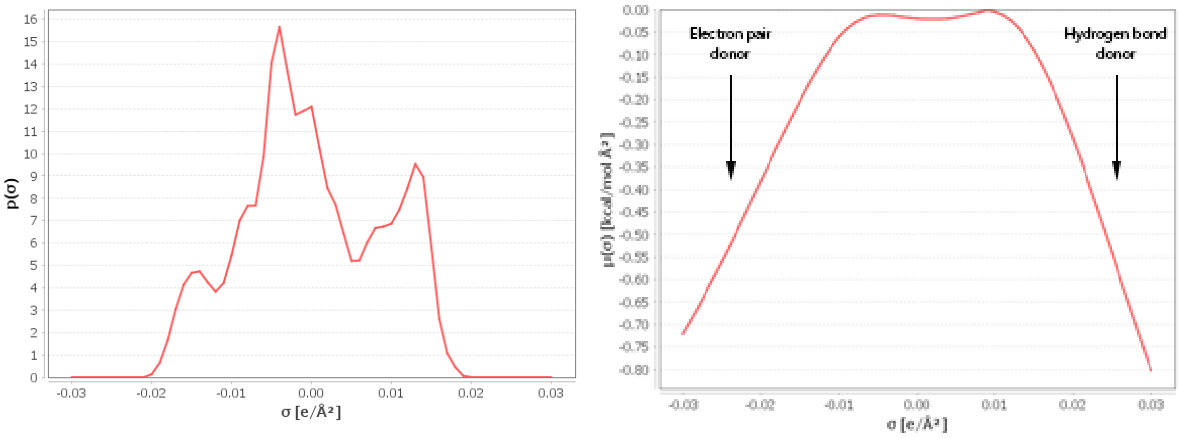


Figure 2.14: The σ-profile (left) and σ-potential (right) plots of ganciclovir as constructed by *COSMOquick*, through the use of fragments in *COSMOfrag*, mirroring the σ-surface in figure 2.13.

The determination of the chemical potential of a solute in solvent provides access to the entire spectrum of solution based thermodynamic properties, involving any combination of chemical entities, without the necessity of experimental data. Mixtures of multiple solutes or solvents can also be investigated through this framework, because they are treated at the same level of theory, without any requirement for the additional parameterisation of solvents.[134] Furthermore, the realm of vapour-liquid equilibria can also be explored if quantum chemical gas phase energies are supplied.

Since its inception, the theory of COSMO-RS alongside its parameters, have been refined and extended further, through the use of DFT calculations, to be applicable for a broader range of chemical environments. The performance of this approach was thoroughly discussed and compared to other solvation models and to classical group contribution methods, at different quantum chemical level of theory.[138-140] Recently, interest increased in the exploration of ionic environments, as reported by Han *et al.* who presented a new parameterisation of COSMO-RS model for ionic liquids.[141] The ability of COSMO-RS to predict infinite dilution activity coefficients of molecular solutes in ionic liquids was evaluated by Paduszynski, by comparing the results with experimental data.[142]

2.2.2.2 Biovia Solvation models and applications

The application of COSMO-RS theory is very often carried out through the Biovia Solvation Models package, primarily *COSMOtherm*, which computes thermo-physical properties as the prediction of solubility, free energy of solvation, activity and partition coefficients, interfacial tension and the construction of phase diagrams, among others.[143] This program also handles the presence of multiple conformers, particularly in cases where the molecular conformation in solution is not equivalent to the one present in the solid state. It has been noted that the initial molecular conformation used for the calculation has a major influence on the quality of the prediction.[144] The weight of each conformation is determined through the Boltzmann distribution, based on the free energy of that specific geometrical orientation in solvent, *S*. Even though conformer generation can be conducted in *COSMOtherm*, one can find much more options in *COSMOconf*, which is equipped with several algorithms having various conformer generators, clustering techniques and levels of QC calculations.[145] Several functions available in *COSMOtherm* can be performed quicker in *COSMOquick*, but at a reduced level of accuracy as the latter constructs the σ-profile of the molecule from different fragments sourced from a database.[146]

The work engine behind *COSMOtherm* is *TURBOMOLE,* which is a quantum chemical suite that accommodates all the required calculations, including the preparation of the appropriate input files for *COSMOtherm.*[147-148] The accuracy of COSMO-RS is highly influenced by the underlying quality of the quantum chemical calculations employed, due to its dependence on the appropriate definition of the actual molecular polarity and electron density distribution. In fact, some of the known limitations of the COSMO-RS methodology are caused by inaccurate predicted molecular polarities such as for conjugated systems characterized by high charge separation.[139] The highest level of theory employed in *COSMOtherm* utilizes the BP functional in conjunction to the def2-TZVPD basis set, at a fine level for the cavity construction. A recent benchmark study concerning the application of different QC levels with COSMO-RS, including pure DFT functionals as PBE, hybrid functionals as Mo6-2x and wave function methods as MP2, concluded that their performance was slightly worse or comparable to the BP/def2-TZVPD settings.[139]

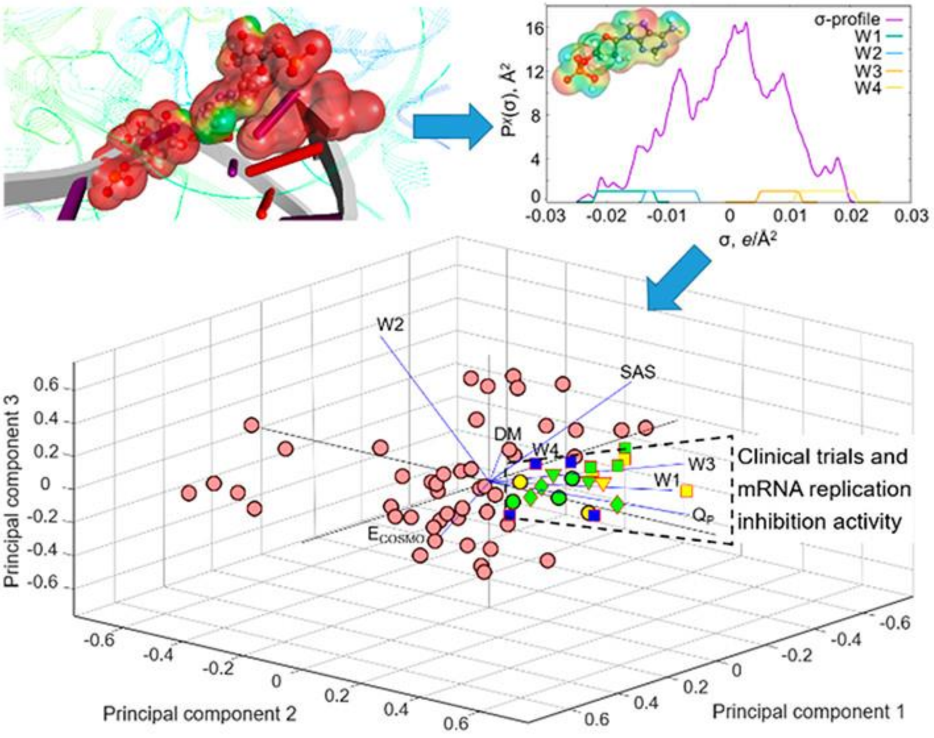


Figure 2.15: During drug development, chemical similarity approaches can be employed by correlating COSMO-RS molecular descriptors of drugs with those of a physiological counterpart or phenotype. This can help in accelerating the development of treatment for emerging diseases as COVID-19.[149]

This framework was employed in various studies having a multitude of objectives including solvate and co-crystal screening, improvement of solubility of solid and gaseous chemicals (ex. atmospherically relevant molecules) and solubility screening.[132, 136, 144, 150-151] This approach was also applied to perform a screening in order to predict which ionic liquid/s would be suitable for the removal of phenolic compounds, through liquid-liquid phase extraction and also for the absorption of CO_2.[152-153] Furthermore, the remarkable list of molecular descriptors provided by *COSMOtherm*, mainly based on the σ-surface, is also very useful for other objectives, beyond the realm of COSMO-RS itself, as seen in figure 2.15.[149, 151]

The possibility of using several conformations as an input provides an alternative route for prediction methods which are in general restricted to a single conformation. This is particularly true for *in-silico* co-crystallisation procedures, which so far are often structured to include only one conformation as an input, even when the experimental process incorporates any amount of solvent. Despite such ability, there is an absence in literature concerning the assessment of co-crystal formation prediction through this framework with a multi-conformation input. This might be relevant also because the success rate of co-crystal formation prediction seems to be influenced by the chemical nature and size of the target molecules.[150, 154-155]

2.2.3 Statistical techniques – Cambridge Crystallographic Data Centre

The Cambridge Structural Database (CSD) was initiated by Dr.Olga Kennard, with valuable support of JD Bernal in 1965, who formed part of the initial members of the CCDC.[156] Despite the initial lack of computational support and the huge endeavour of collecting data from researchers around the world, the primary vision for the CSD has now been exceeded with over one million structures available (V.5.42). The exponential growth of the database and the high quality of the data present is mainly due to technological advancements which ease the process of data collection and ensure that there is a substantial extent of compatibility between experimental and structural data for each entry.[156] This is highly important because the reliability of the results obtained from analysis of the entries within the database directly relies of the quality of the inputted information. In order to sustain such quality, each submitted entry goes through a validation process before approval, together with additional checks required in special cases such as in complex metal structures.[157]

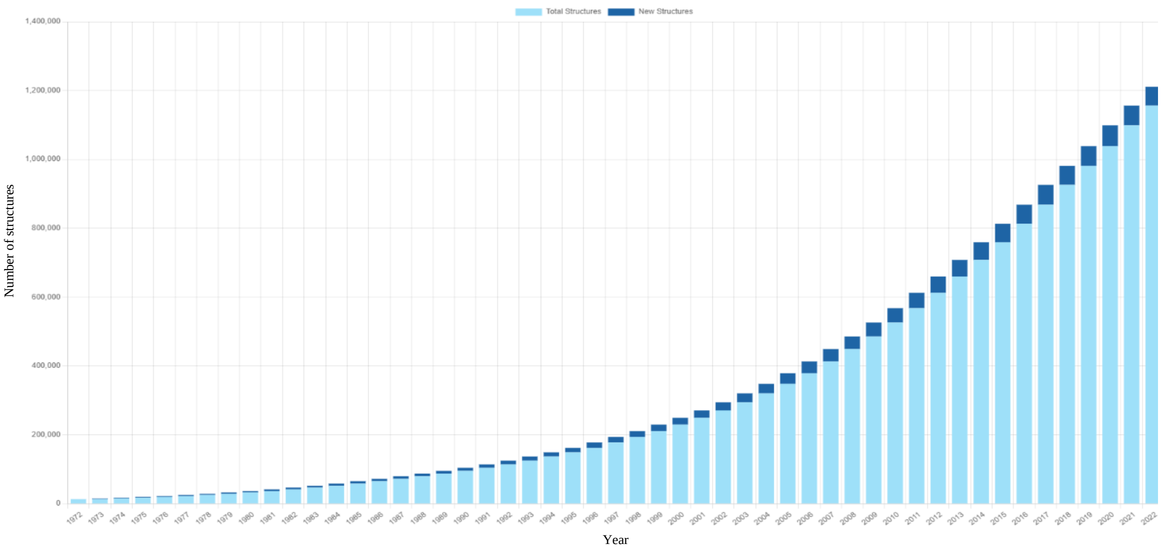


Figure 2.16: A histogram showing the growth of the CSD along the decades, with over 1.2 million structures present currently. As seen from the top blue bars, the amount of structures submitted yearly increases through time.[158]

The complexity of the submitted entries, particularly drugs is escalating through time, namely proved by the increase in their respective molecular weight and number of rotatable bonds.[159] A statistical study on the drug subset in CSD showed how despite having a higher level of complexity, the number of hydrogen bond donors present in each drug is retained low, as opposed to the increase in the number of hydrogen bond acceptors, especially in industrially available pharmaceuticals.[159] Such ratio has direct implication in the degree of solubility exhibited by the API, and certainly further research can be performed to optimise the drug’s efficacy using such information. This study portrays one of the most advantageous features of the CSD which revolves around the ability to examine a collection of data, and extracting information that go beyond isolated cases. Statistical analysis on the multitude of molecular and crystallographic data present in CSD can be used to extract patterns, trends and correlations, which otherwise would have been lost.

2.2.3.1 Applications

In order to perform such analysis, the CCDC provides a toolkit of different software, all of which present the opportunity to construct knowledge-based methods to examine chemical environments of crystal structures. This toolkit enables the holistic study of crystal structures through different modes of analysis at different levels, starting from the individual molecular structure, proceeding to the intermolecular analysis and then to a supramolecular system. One can find various case studies, under the realm of Solid Form Informatics, which apply such approach.[160-162] A detailed review, presented by Robin Taylor and Peter Wood, published on the occasion of exceeding the millionth structure mark, explores the wide range of tasks which can be conducted through the CCDC toolkit.[163] They also included details about the limitations of CSD-based research and explained how these weaknesses can affect the relevance of the inferred results.

Intramolecular data including bond lengths, ring geometries and torsion angles, all of which define the conformation of a molecule, were used to assist theoretical methods in understanding underlying factors in crystal structures. For example, such tools were used to examine the role of C-H···π interactions in stabilising alkyl-aromatic bonds and axial alkylcyclohexanone orientations.[164] Similar methodology was applied by Galek *et al.* to expose the stabilising effect of an intramolecular bond within rings of 5, 6 or 7 atoms.[165] Conformational geometries can be validated using the library-based program, *Mogul*, which is able to construct distributions for each geometrical parameter within the molecule by retrieving information from the database.[166] Moreover, these distributions can be used to assist the structure solution from powder data, to optimise molecular geometries and to navigate the conformational space of a molecule, in attempt to assess its flexibility.[167]

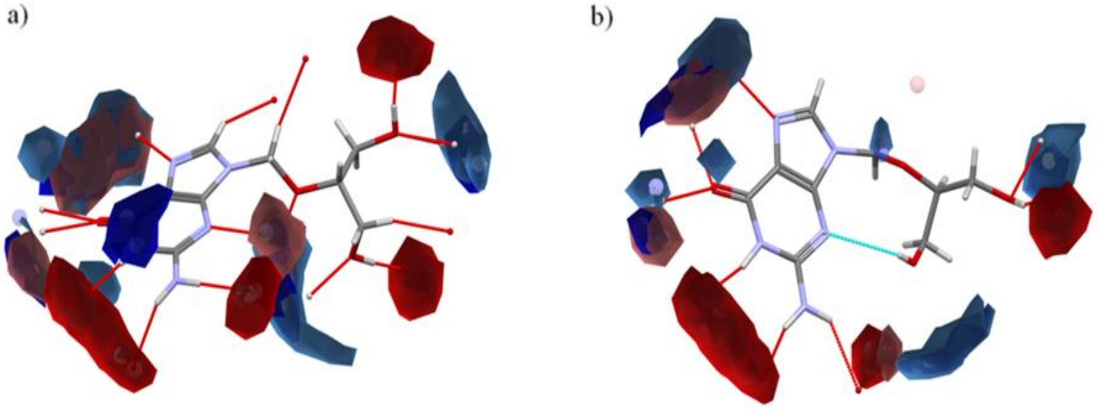


Figure 2.17: Full interaction map of GCV (a) form I and (b) form II, with uncharged NH nitrogen (blue), RNH3 nitrogen (light blue), alcohol oxygen (light red) and carbonyl oxygen (red), at level 6.0. The maps also contain the hydrogen bonds predicted by *Mercury*.[162]

Through the use of *IsoStar*, one can predict the tendency of a functional group to participate in hydrogen bonding by analysing the chemical environments of that specific group in other chemically relevant structures.[168] This technique can be extrapolated to examine all the possible sites within the whole molecule for interactions through the use of full interaction maps in *Mercury,* as viewed in figure 2.17.[169] The study of intermolecular interactions also comprises of the analysis of systematic crystalline assemblies, namely motifs and synthons. This exploration provides an informative insight with respect to the packing arrangement and morphology of crystal structures, as well as their stability. In addition, *Mercury* enables the determination of molecular and crystal structure similarity, both of which can be useful in applications as clustering during conformer generation or crystal structure prediction process.

2.2.3.1.1 Polymorph and co-crystal screening

Due to such large depository of data, this toolkit provides several routes towards the prediction of crystal structure characteristics, particularly those dependant on the HB network, by deriving trends from a multitude of chemically relevant structures.[170] The concept of prediction is based on the underlying assumption that the frequency of a particular parameter or interaction is an indication of its contribution towards the overall stability of a molecule or crystal structure. For example, the stability of hypothetical polymorphs and co-crystals can be predicted by analysing the frequency of specific supramolecular synthons which might be in competition during the potential synthesis.[171-172] The ability to extract trends from sets can be also employed to construct H-bond Propensity (HBP) models which categorise the stability of a polymorph based on the probability associated with its intermolecular H-bonds.[161, 173] The data is utilised to construct a Logit HBP model which produces propensity values associated with potential pairwise interactions within the reference structure. Through this method, Nauha *et al.* evaluated the risk of exhibiting polymorphism of several compounds in the European Pharmacopoeia.[174] HBP models are also utilised for co-crystal prediction by comparing the probability values of the most favourable pure and impure interactions.[175]

HB coordination (HBC) and Molecular Complementarity (MC) methods are another two procedures available in *Mercury* for the assessment of the likelihood of co-crystal formation.[170, 176] Similar to HBP, HBC employs a training set to deduce probability of every possible coordination number of each HB participant in both API and coformer. A positive prediction for co-crystallisation would result from having the donor-acceptor pair with the highest coordination number located in different components.[177] Costa *et al.* modified the HBC approach to predict the likelihood of co-crystallisation on the basis of the relative degree of mismatch in the ability to donate and accept HB, between the potential co-crystal and individual components.[155] Meanwhile, MC tool is built on a method developed by Fábián, whereby the chance of co-crystallisation depends on the extent of molecular geometric compatibility of the pure components.[175, 178]

Due to the difficulty in predicting co-crystallisation, the results of these distinct approaches are very often either integrated together or accompanied by outcomes of other prediction methods.[155, 175, 177, 179-180] Their success rates are considerably high when employed for relatively small, rigid molecules, with the performance of the HBC route being the least reliable, even after modifications. It was displayed how the success rates decrease significantly when larger, flexible molecules are under study, with the success rate of HBP reduced from 87% to 53%.[177, 179] Further research is required to test whether taking into consideration multiple interactions for HBP, beyond the most probable pair, would lead to an improvement when flexible molecules are being studied. Similar approach was adopted for hydrate formation prediction.[181] Another criteria which requires investigation revolves around the role of carbon as HB donor, which is also excluded from any approach. The major drawback with respect to MC revolves around the limits of the acceptance criteria associated with each geometric descriptor, which are so far set as a standard, irrelative of the chemical nature of the molecules under investigation.[155, 179] Furthermore, the fact that the molecular descriptors of MC are dependent on the conformation, and hence can yield a different prediction, is very often disregarded.[177-178] The couple of instances where different molecular conformations were used as inputs, there were no clear indications of which route of conformer selection provided the best prediction.[155, 179]

2.3 Antivirals

Ganciclovir (GCV), famciclovir (FCV) and valacyclovir (VCV, in monohydrochloride form) are the chosen candidates in this project, primarily due to the lack of crystallographic information and solid-state behaviour profiling available in the literature, even despite their importance as antiviral agents. FCV and VCV are pro-drugs of penciclovir and acyclovir (ACV), respectively, and together with GCV, they have been the recommended medication for the treatment or prevention of the herpes simplex virus (HSV) type 1 and 2, and varicella zoster virus (VZV), for two decades.[182-183] The presence of HSV is widely distributed amongst the human population, with an estimate of

67% of people under 50 years being infected with HSV-1.[6, 184] Despite being also affective against HSV, the administration of ganciclovir is very often restricted as an active agent against the human cytomegalovirus (CMV), which is a particularly complex human virus when it comes to gene diversity and contains additional splicing and codons, which increase the variety of viral proteins even further.[185,186] HSV and VZV belong to the *Alphaherpesvirinae* subfamily, while the human CMV belongs to the *Betaherpesvirinae* subfamily.[182] The severity of symptoms suffered due to such viral infections is markedly high, and can be fatal if not treated in immunocompromised hosts, transplant recipients and neonates.[6]

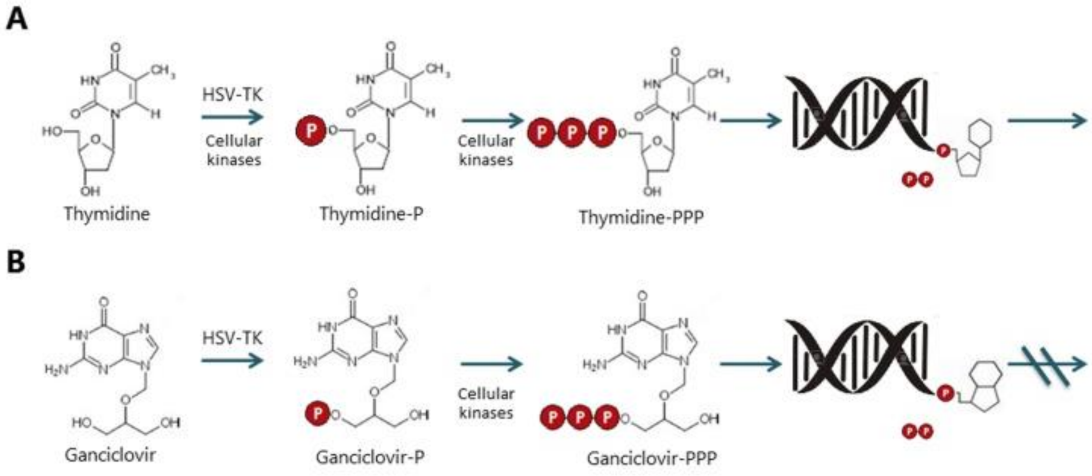


Figure 2.18: The comparison of the metabolism of thymidine (A) and GCV (B) by HSV thymidine kinase (TK) and cellular kinases, where consecutive phosphorylations take place prior to the integration into the DNA. Unlike the human TK, HSV-TK is able to catalyse the initial phosphorylation of GCV. [187]

These antivirals are guanosine nucleoside analogues, which act as inhibitors for viral DNA synthesis, thereby causing a reduction the viral population within the host. After the insertion within the host cell, the antiviral agent undergoes a series of modifications, primarily consisting in the consecutive addition of phosphate groups, as illustrated in figure 2.18. The resultant deoxyguanosine triphosphate analogue competitively inhibits naturally occurring nucleosides or nucleotides from incorporation by viral DNA polymerase to transcribe the viral DNA chain. The integration of GCV-triphosphate leads to the termination of the construction of the viral DNA as depicted in figure 2.21b. Apart from treating the viral infection, there is evidence that some antivirals, specifically FCV have the ability to minimise the potential for latency.[183] Furthermore, systems as HSV-TK in the presence of GCV are being tested in pre-clinical and clinical studies against various types of cancer as brain, bladder and liver, through suicide gene therapy.[188-191]

Besides their essential role against viral infections, they are selected due to their resemblance in structure, with the elongation and composition of the branch being the main difference amongst them (figure 2.19). Such characteristic adds a layer of comparability within the research itself, since it enables the analysis of the effect of having longer branched chains on the solid-state behaviour of each antiviral, with respect to polymorphism and ability to form co-crystals.

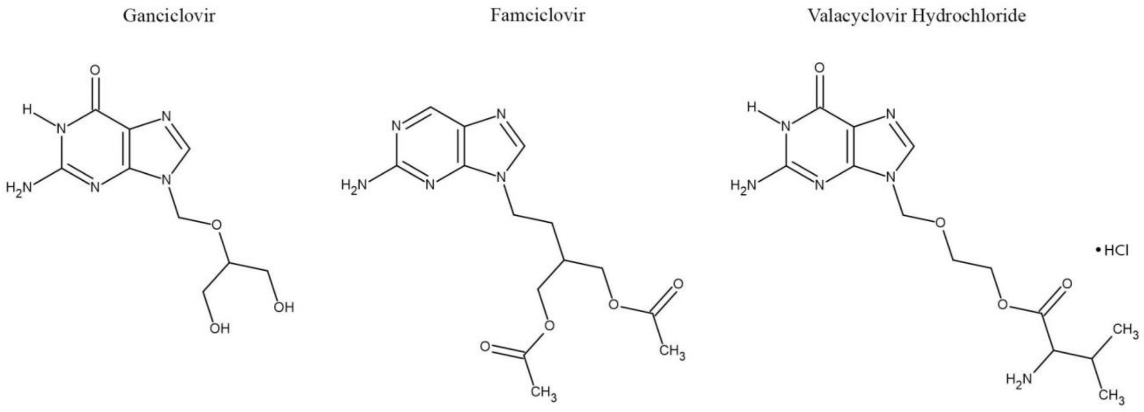


Figure 2.19: Schematic structure of the targeted antivirals.

Despite the importance of GCV, its physicochemical properties are still not optimised, due to the presence of low solubility and permeability, all of which restrict its oral bioavailability (5-7%), causing its administration to be frequent and in high dosages.[192-194] In fact, GCV is usually formulated as a sodium salt and administered through an intravenous route (Cytovene®) or as an ophthalmic gel which contains GCV in its pure form, with a very small percentage of water (Virgan®).[195-196] One can find four structurally determined crystal forms of GCV in the literature: anhydrous form I (United States Pharmacopeia reference standard – CSD refcode: UGIVAI01)[197], anhydrous form II (CSD refcode: UGIVAI)[198], hydrous form III (CSD refcode: ENEBUW)[199], hydrous form IV (CSD refcode: ENECAD)[199] and its HCl salt (CSD refcode: ENEBOQ).[199] These forms were thoroughly studied experimentally, defining their stability with respect to temperature and humidity, as well as the relationship between themselves. They were characterised by Fourier transform infrared spectroscopy, mass spectroscopy, elemental analysis, as well as by discrete and simultaneous measurements of X-ray diffraction with differential scanning calorimetry.[197, 199-201]

FCV is the diacetyl, 6-deoxy, ester prodrug of penciclovir, and it was developed due to the poor bioavailability of oral penciclovir, which was even lower than that of ACV (by 10 to 25% lower).[202-204] The oral administration of FCV amplified the bioavailability of penciclovir to 60%, thereby reducing the requirement of frequent drug intake.[202] The ability of FCV to manifest polymorphism has been demonstrated through the number of anhydrous forms discussed in the literature, but only one of which has been determined structurally (CSD refcode: IMADEI).[205-206] This structurally determined polymorph, referred to as form I, is the one utilized to manufacture the commercially available API. Dolitzky *et al.* explained various ways of obtaining forms I, II and III (isolated or mixture) through the use of recrystallization with several solvents or heating processes.[205] In addition, the newly identified form IV was reported to be produced after a consecutive process of heating form I to 110°C and cooling to -50°C.[206] The multitude of methods described for the transformation of one polymorph to another, coupled with the possibility of obtaining mixtures, portray the complex task of unveiling the thermodynamic relationship between them. Furthermore, FCV has a monohydrate form (CSD refcode: YACHAL) which is commonly either used as a precursor for the formation of the anhydrous forms or present within a resultant mixture of products.[205, 207]

VCV is a L-valyl ester prodrug of acyclovir, primarily synthesized to overcome the poor solubility characteristic of acyclovir, which highly restricts its bioavailability. Such improvement in bioavailability led to the cessation of using intravenous ACV and to the reduction of administration frequency.[208] Nonetheless, despite being considered as a BCS class III drug, some of its dissolution results remained inconclusive. VCV is easily metabolized by a hydrolase located in the brush lining of the gastrointestinal tract and the liver, producing ACV and valine.[194] Given that the most commonly administered form of this antiviral is the monohydrochloride form (VCV·HCl), marketed as VALTREX®, this was set as the main subject of interest during analysis. Despite the advantageous use of VCV·HCl, there is a lack of crystallographic information in literature (particularly computational), with no structurally determined form to date, probably due to its inability to diffract X-rays sufficiently well for crystal structure determination. One should note that there are two anhydrates and numerous hydrate forms of VCV·HCl reported thus far, highlighting the need of an in-depth investigation, especially with regards to thermal stability and risk of polymorphism.[209-214]

2.4 Project objectives

The primary objective is to characterise the crystallographic behaviour of the three antivirals – Ganciclovir (GCV), Famciclovir (FCV) and Valacyclovir (VCV, in monohydrochloride form), due to the lack of information in literature, using a range of experimental and computational methods. Such *in-silico* techniques will involve a combination of published (standard), modified and novel methodologies, which in return will be assessed against experimental results. Several routes are employed to probe at different angles, so as to study such structures more holistically. In particular, the study will aim to:

i. quantify adequately the molecular flexibility, beyond the number of rotatable bonds and obtain the necessary geometric parameters for the mapping of accessible conformational space,
ii. analyse the crystal structure of anhydrous forms to derive insights on their thermal stability, thereby complementing experimental work,
iii. investigate the risk of polymorphism with respect to anhydrous solid-state forms,
iv. predict the viability of co-crystal formation, using the commonly administered form, and test the effect of using different molecular conformations as inputs and accounting for alkyl protons as HB participants.

Success in such objectives could lead to better pharmacokinetic properties of these antivirals and improved methods for crystallographic behaviour assessment and prediction.

The following chapters relate to each of the above individual objectives, each containing its respective methodology and results. This setup was chosen to illustrate how information at a molecular level can be utilised to guide analysis and explain observations at a supramolecular level.

It is important to note that every chemical scenario dealt with during this project does not concern any possible process which happens *in vivo.* The presented characterisation of these antivirals was restricted to their chemical stability and properties prior to entering the body.

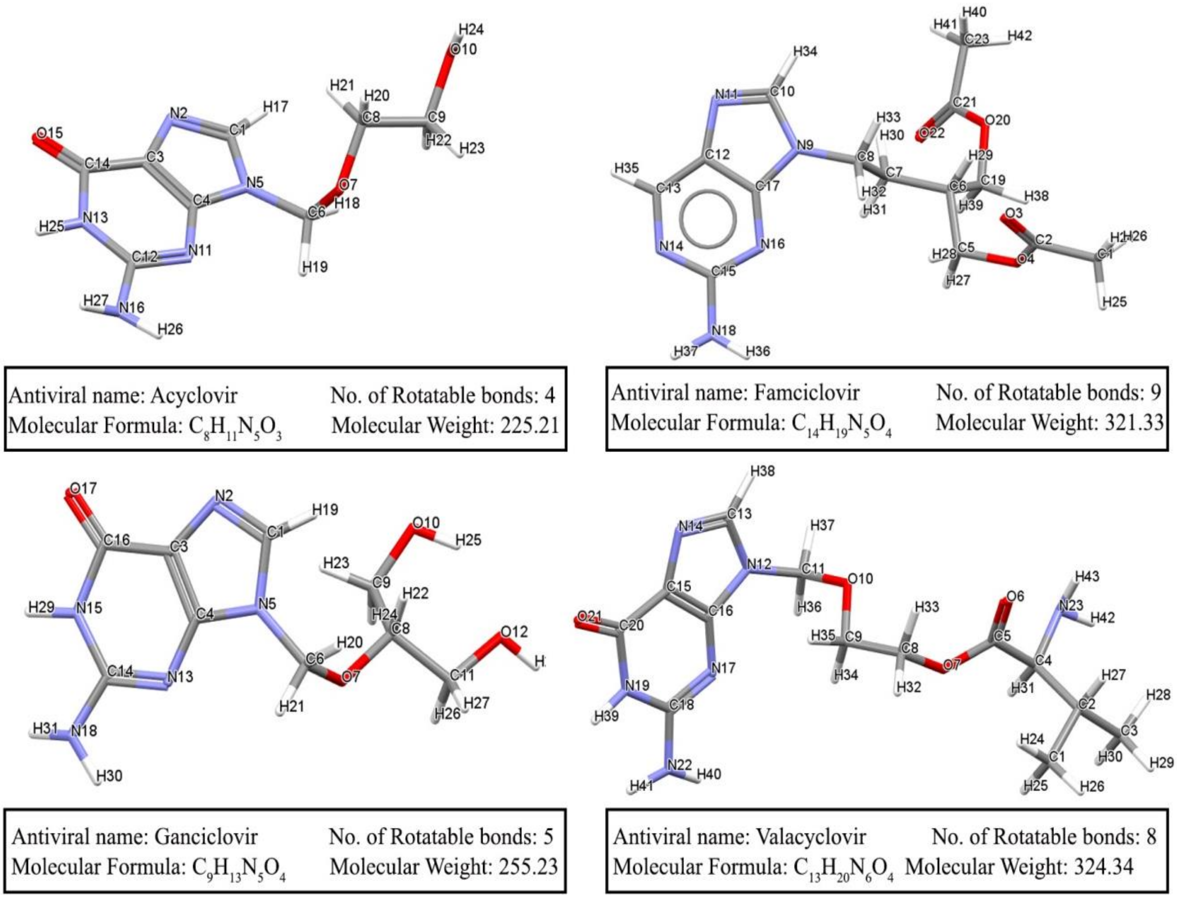
Chapter 3

EXPLORING FLEXIBILITY

3.1 Introduction

The exhaustive definition of the accessible conformational space of a molecule is known to be a determinant factor to compound classification and crystal structure prediction.[77, 89, 215-216] The extent of molecular flexibility is at the centre of such attribute because it defines the extensivity of the conformational space, as well as the ease of crystal behaviour prediction.[89] Conformational variations may cause a change in the intermolecular network, which in return would lead to a shift in supramolecular characteristics, and thereby altering the physicochemical properties of the chemical species. It has been shown how highly flexible molecules are associated with either reduced tendency of crystallisation or else increased probability of conformational polymorphism due to the presence of multiple configurations which might provide enough stabilisation.[217] A large degree of flexibility implies that the molecule itself is able to be energetically stable in a larger number of orientations when compared to a molecule whose rotations are accompanied by high energy barriers.

The most common measure of flexibility is the number of rotatable bonds, NRot within the molecular structure, thereby counting "*any single bond, not in a ring, bound to a nonterminal heavy atom*." [218] Such property, explicitly measured using this crude definition, has been used to reflect flexibility in multiple investigations such as in relation to oral bioavailability of pharmaceuticals and prediction of co-crystallisation.[218-221] Despite its extensive usage for flexibility representation, NRot has several issues including its discrete nature, its lack of ring flexibility consideration and inability to distinguish between isomers.[220] Other less popular, but better methods include Kier molecular flexibility index and global flexibility index, both of which rely on topological descriptors.[222] However, none of these methods provide the individual contribution of each rotatable bond towards the molecular flexibility.



Antiviral name: Acyclovir No. of Rotatable bonds: 4
Molecular Formula: $C_8H_{11}N_5O_3$ Molecular Weight: 225.21

Antiviral name: Famciclovir No. of Rotatable bonds: 9
Molecular Formula: $C_{14}H_{19}N_5O_4$ Molecular Weight: 321.33

Antiviral name: Ganciclovir No. of Rotatable bonds: 5
Molecular Formula: $C_9H_{13}N_5O_4$ Molecular Weight: 255.23

Antiviral name: Valacyclovir No. of Rotatable bonds: 8
Molecular Formula: $C_{13}H_{20}N_6O_4$ Molecular Weight: 324.34

Figure 3.1: Molecular conformation of the antivirals used as an input for conformer generation, displayed in *Mercury*, with the numbering system used in this chapter.

Due to its crucial implications, the objective was to define and quantify the extent of accessible conformational space of the four antivirals illustrated in figure 3.1. In this chapter, the VCV molecule was considered in its neutral state (not HCl salt), for comparison and simplicity purposes. An alternative methodology was devised to achieve such objective, beyond the use of NRot. In particular, this procedure was performed through two novel approaches; the statistical analysis of several variables which describe the unique Hirshfeld surface constructed between the backbone and straight chain of each theoretically generated conformer of each antiviral, and the calculation of a flexibility score for every molecule based on torsional data from the CSD. These approaches combine the use of theoretical techniques with experimental data from determined crystal structures. In order to examine the potential relationship between the length and composition of the branches and molecular flexibility, this procedure was also performed for acyclovir (ACV).

3.2 Methodology

3.2.1 Exploring flexibility through torsion angle analysis

CSD Python API was used to retrieve the hits associated with each torsion angle, as defined by SMARTS code, from the CSD V5.41 (November 2019).[156] Given that the likelihood of rotation is not always symmetrical, this technique of search ensured that every measurement was accurately represented in histograms with full range from 0° to 360°.[223] Each torsion angle was defined by describing the chemical nature of all the immediate neighbouring atoms constituting that particular angle (see Appendix). Some definitions were altered to be less restrictive in cases with very few hits, while still ensuring the relevance of the results. Disordered hits were excluded. *Origin 2019b* was used for the illustrations of the torsion angle histograms.[224]

3.2.1.1 Flexibility Score

The information required to quantify the flexibility of a rotatable bond had to be extracted from the data collected in the previous stage. Given the disadvantages of histograms, especially related to the subjectivity of the bandwidth, kernel density estimation for circular data was employed using the von Mises density *(3.1)* with parameters as specified by McCabe *et al.*[223, 225]

$$\hat{f}(\theta;\mathrm{v}) = \frac{1}{n(2\pi)I_0(\mathrm{v})}\sum_{i=1}^{n}\exp[\mathrm{v}\cos(\theta-\theta_i)] \qquad (3.1)$$

where θ_i represents the torsion angles, $I_r(\mathrm{v})$ is a modified Bessel function of order r, and concentration parameter, v which is the inverse of the smoothing parameter.[226] Such smoothing criterion was estimated by a robust version of the method proposed by Taylor for sample sizes less than 40 hits, and using an alternative procedure based on cross-validation criteria for larger datasets.[226-227] The peaks characterising each estimated density were identified so as to enable the calculation of the relative frequency, r_i of the hits within ±15° range of the peak itself. It was ensured that the ranges which were taken into account did not overlap, because this would have caused double counting. All data analysis was conducted in *R*.[228-229]

The flexibility score, *F* of a rotatable bond:

$$F = 1 - \sum_{i=peak}^{n} r_i w_i \qquad (3.2)$$

where *n* is the total number of peaks and w_i is the height of the respective peak divided by the summation of the heights of all identified peaks. The second term can be considered as the weighted average of the relative frequency, using the ratio of the peaks' heights as weight. This term was constructed to quantify the rigidity of the rotatable bond, whereas its complement defines the flexibility, with a possible range from 0 to 1. The total flexibility of each antiviral was estimated by summing all the individual scores of its constituent rotatable bonds.

3.2.1.2 Conformational energy graphs

The molecule was drawn in *Material Studio* 2021 and its geometry was optimised by employing dispersion-corrected density functional theory, using PBE functional and double numeric plus polarisation (DNP) basis set, at a fine level through the DMol^3

module.[230-232] The resultant molecule was used for the conformational sampling, by employing a systematic grid search and restraining one torsion angle every degree at a range from -180° to 180°. This procedure was performed for each torsion angle present in molecule. Each energy calculation was performed using COMPASSIII (V1.1) forcefield and forcefield-assigned charges, with an atom-based and cubic spline methods applied for summation (electrostatic and van der Waals) and truncation purposes respectively.[99] The utilisation of ESP derived charges produced similar results. This method was performed for GCV torsion angles only as a compatibility check with torsional angle histograms. All convergence criteria can be found in Appendix.

3.2.2 Exploring flexibility through conformer generation

The geometry optimisation of the starting molecule in gas-phase for approaches A and B was performed in *Turbomole* V7.2 through *TMoleX* V4.4.1.[147-148] The calculation was conducted using the hybrid functional PW6B95-D3 with RI approximation, which was recommended by Goerigk *et al.* in their benchmark analysis, in conjunction to def2-TZVPD basis set and BJ damping function for convergence.[233-236] Full details of the convergence criteria are included in appendix. It is noteworthy to add that the proceeding methodology for the generation of conformers did not aim to produce the final candidates for polymorph screening, but to generate theoretic conformers with reasonable chemical environment which can represent every conformational region of each antiviral, in order to understand the extent of flexibility. This was achieved through the use of three routes, summarised in figure 3.2 and detailed in the next section.

3.2.2.1 Conformer Generation methods

Set A – Thermodynamic approach

The energy file of the energetically minimised molecule was used as an input in COSMO*conf* V4.2, and the default settings of BP-TZVP-FINE-GAS template were retained (see appendix).[145] Such template employs the Balloon conformer generator which applies a multi-objective genetic algorithm to create an ensemble of conformers within specified criteria.[237] Such algorithm is suitable to handle flexible molecules, and it produces a set a conformers from a random search after a series of evaluation of solutions, energy calculations, mutation operations, clustering and optimisation with respect to specific criteria.[237] The final 15 geometrically optimised conformers (BP86/def2-TZVP), within an energy window of 6 kcal/mol were considered to form set A.[230, 235, 238]

Set B – Hybrid approach

COSMO*build* V17.0 was utilised to open the energy file and exhibit the 3D structure of each molecule, so as to add any necessary double bonds and convert it into a .mol format.[239] The in-built "Conformer Generator" function in *CSD Python API* was employed to generate a maximum of 500 molecules, using a combination of energy calculations and statistical information derived from the extensive rotamer libraries within the CSD.[240] All default settings related to probability scores, molecular clustering and number of unusual torsion angles, were retained. The inputted molecule was not further energetically minimised prior to the initiation of the procedure. The first most probable 15 generated conformers were selected to form set B.

Set C – Clustering

Set A and B represented potential conformers with minimised energies, however, it was possible that they did not represent every possible conformational region which is chemically viable. Therefore, in order to attempt obtaining such representation, further analysis on the complete set of conformers generated by *CSD Python API* was carried out. Set C was formed from such list of conformers due to the large sample size provided in minimal computational time by the *CSD* conformer generator. Each conformer was defined in the dataset by means of its measured torsion angles. Multivariate analysis was conducted on the torsion angles by applying GeoPCA algorithm which performs principal component geodesic analysis, appropriate for circular data rather than data points in the Euclidean space (using *Clustangles* package in *Python 2.7*).[241-243] The first two principal components were calculated for each conformer based on their torsion angles, and clustered.[241] The members of each cluster were sorted in an order which placed the conformer with the highest probability (score provided by CSD conformer generator) and lowest number of unusual torsion angles on top. The first conformer of each cluster was selected to form set C, while removing any chosen conformers which were already present in set B.

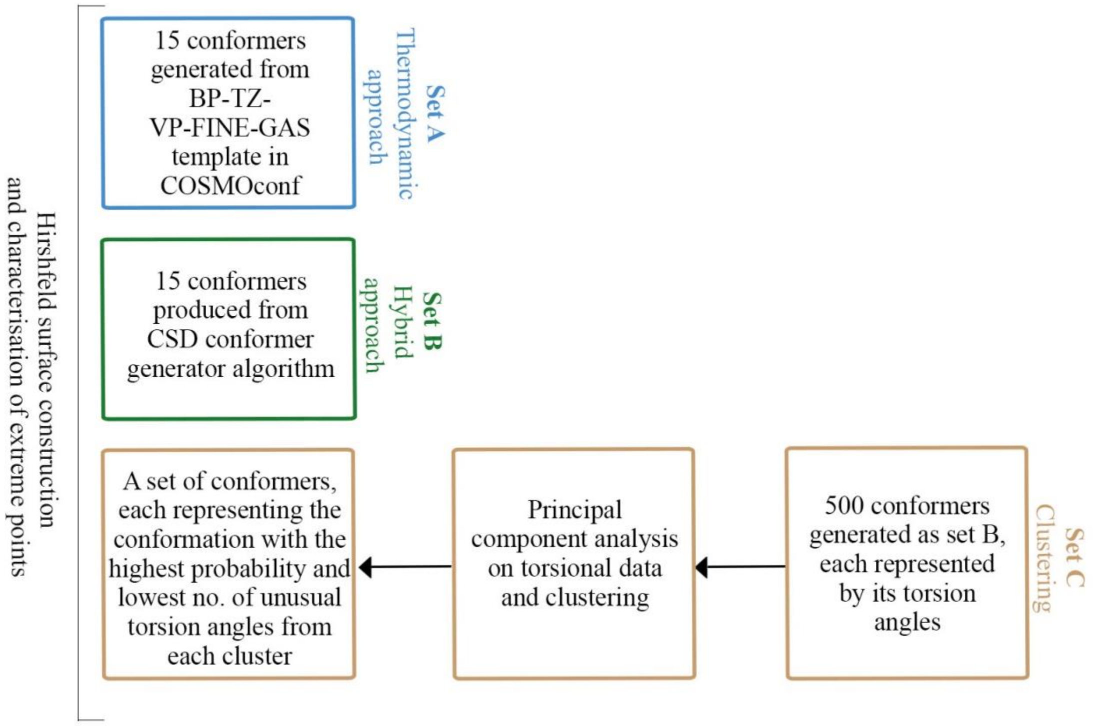


Figure 3.2: Summary of the different routes for conformer set selection and analysis.

3.2.2.2 Generation of Hirshfeld surfaces

Before being used to define the space of a molecule in a crystal, Hirshfeld surfaces were originally aimed to describe atoms in molecule, combining Bader's quantum theory of atoms in molecules (QTAIM) with F.L Hirshfeld's partitioning scheme.[244-245] The weight function for an atom, with respect to its molecule, $w_a(r)$ is expressed in equation 2, where $\rho_i^{at}(r)$ represents the averaged electron densities of an atomic fragment,i.[246]

$$w_a(r) = \frac{\rho_a^{at}(r)}{\sum_{i\in mol} \rho_i^{at}(r)} \qquad (3.3)$$

$$d_{norm} = \frac{d_i - r_i^{vdW}}{r_i^{vdW}} + \frac{d_e - r_e^{vdW}}{r_e^{vdW}} \qquad (3.4)$$

Multiwfn V3.7 was used to construct a Hirshfeld surface between the aromatic system and the straight chain, as shown in figure 3.3, for the conformers in all three sets of the four antivirals. The Hirshfeld surface of a fragment is constructed at points where the summation of $w_a(r)$ ∀ atoms ∊ fragment is equal to 0.5.[246] Given that the wavefunction for each conformer was not available (.mol input format), built-in spherically averaged atomic densities were utilised in the process, as suggested by literature (see appendix).[247] More details about the procedure can be found in the software manual, chapter 3.[248] The partition of the molecule was chosen primarily because the profile of the surface can be used to reflect the degree of folding, thereby having a unique surface for each conformer. The normalised contact distance, d_{norm} (eq. 3.4), was the property selected as the mapped function of the surface.[246] Its expression involves the vdW radii, r^{vdW}, and d_e and d_i which denote the distance from a point on the surface to the closest nucleus outside and inside the surface, respectively.[246] The extreme points were characterised by geometric coordinates and d_{norm} value, all of which were noted for further analysis. All extremes on the periphery were excluded from the dataset. In order to define the extent of folding even further, measurements of areas (Å^2) with specific d_{norm} ranges were also included. *VMD* V1.9.3 was employed for visualisation purposes.[249]

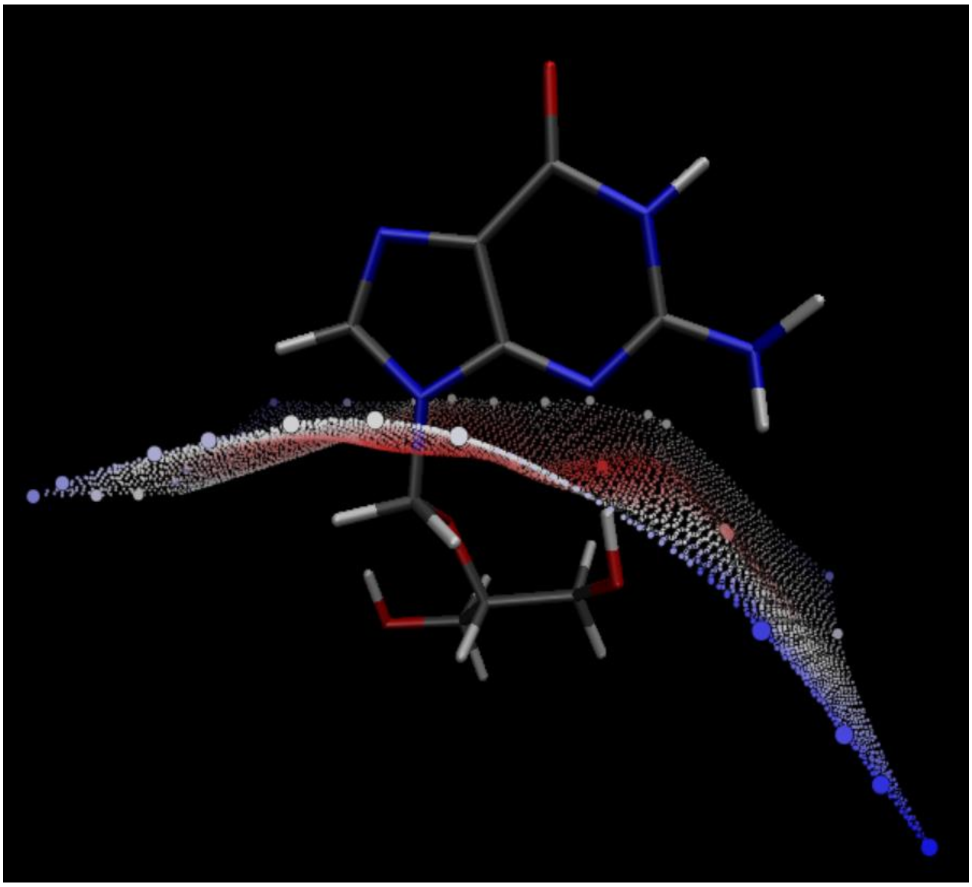


Figure 3.3: The most energetically stable conformer produced by COSMO*conf* for GCV, visualised using *VMD*.[249] The colour scheme reflects the d_{norm} values, with (-) red to white to blue (+) shades, whereas larger points represent the extreme points.

3.2.2.3 Data Analysis

The level of similarity between each pair of conformers in all sets was quantified by a RMSD value using CSD Python API, which also took the positions of the hydrogen atoms into account.[156] Descriptive statistics was employed to investigate trends within the RMSD values, as well as in other descriptors obtained from the Hirshfeld surfaces. Further techniques included Principal component analysis, clustering, and visualisation after multi-dimensional scaling. Univariate and multivariate tests for homogeneity of variances were conducted to assess whether there were any statistically significant differences between the within-group variations captured by means of the variables collected.

3.2.3 Prediction of conformation propensity

The objective of this method was to build a regression model using descriptors of the Hirshfeld surfaces plotted in section 3.2.2.2, for a group of 100 conformers generated as described for set B. The regression model was constructed using the probability associated with each conformer given by *CSD Python API* as the dependant variable, and the overall average value of electron density corresponding to each atom detected by the Hirshfeld surface, as predictor. The goal was to build a model which can ultimately be able to predict the probability of any conformer (irrespective of how it was generated) to be present in a solid-state form by inputting information concerning the atoms detected by the Hirshfeld surface (sensitive towards the degree of folding within a molecule).

Zero electron density average values of detected atoms were changed to $1x10^{-1}$, so as to distinguish them from atoms which were too far away from the constructed surface for detection (electron density was set to zero), but still small enough to remain truthful. Furthermore, equal probability values were modified (the 11th decimal place) so as to respect the rank given by conformer generator. Highly correlated atomic variables were removed to avoid multicollinearity problems. Given the nature of dependant variable, beta regression was chosen and conducted in *R* V3.6.3.[228, 250-251] Tests were conducted to find the appropriate link functions for precision and mean parameters, as well as to determine which variables and interactions are statistically significant. Model selection was also based on the Akaike Information Criterion (AIC), pseudo-R^2 value and the properties of the produced residuals.[252]

The performance of the model was assessed by processing conformers present within known forms of the antiviral, and ensuring that their associated probability was relatively high. Given that a good number of conformations within structurally determined forms were required for testing, this procedure was only conducted on ACV. Furthermore, this methodology was repeated using local surface area values instead of electron densities, so as to provide a model based on geometric environment.

3.3 Results and discussion

3.3.1 Exploring flexibility through torsion angles analysis

3.3.1.1 Flexibility of individual torsion angles

Torsional analysis through the use of histograms (based on CSD data) was the first step to analyse qualitatively the level of flexibility of the four antivirals which classified as Rule-of-five complaint. [157, 253-256] This procedure provided relevant geometric bounds of the accessible conformational region of each antiviral. Some rotatable bonds were examined through chemically different torsion angles, such as C1N5C6O7 (G1) and C4N5C6O7 (G2), both of which had the same flexibility score. The definitions of torsion angles, together with their histogram are displayed in the appendix. Each torsion angle has an associated unique code, first letter corresponding to its respective antiviral's name, as shown in table 3.1 (see appendix for full list).

Several histograms exhibited how the rotatable bonds have limited flexibility, most of which are restricted to two to three specific regions as A2 (O7C8C9O10), F3 (N9C8C7C6) and F4 (C8C7C6C5) in figure 3.4. This feature highlights the importance of considering the flexibility of a rotatable bond as a partial number, rather than a discrete quantity. Conversely, a greater extent of rigidity is illustrated through the histograms of C11O10C9C8 (V9) and C5O4C2O3 (F8), whose lack of ease of rotation is caused by the repulsion exerted between the lone pairs of electrons of the oxygen atoms in the ester group. For exhaustiveness purposes, torsion angles constituting in hydroxyl groups, such as C7C8O3H1 (G6), were also analysed, and their histograms demonstrated a considerably high extent of flexibility (see appendix). However, due to their minor influence on the overall structure of molecule, such torsion angles were excluded from the flexibility score calculations. Histograms also demonstrated how data is not exactly symmetrical; implying that the probabilities of having torsion angle of $\theta°$ and $-\theta°$ respectively, are not necessarily equivalent.

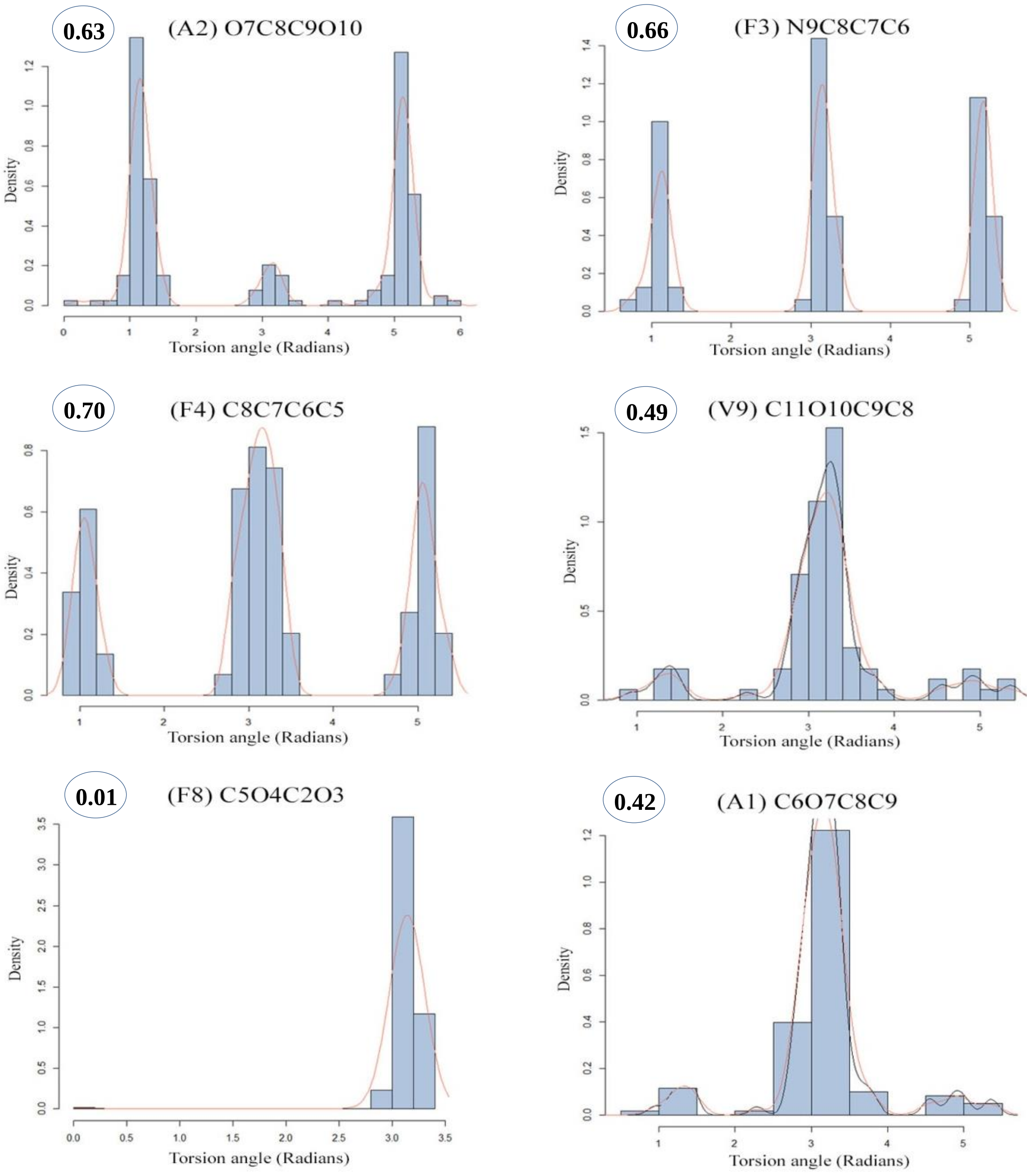


Figure 3.4: Histograms of different torsion angles (radians), together with their respective kernel density approximation and flexibility scores (circled). The histogram of V9 and A1 illustrate the two methods for bandwidth determination for comparison, with bw.cv (coral) and bw.rt (black).

Given the wide range of shapes of torsional distributions, simple cosine dependant functions are insufficient to reproduce such complexity.[223] Kernel density approximations provided a way how to standardise histograms, irrelative of the size of the dataset and reduced the possibility of plotting histograms which presented misleading information due to non-optimal bandwidth selection.[223] However, even though minimised, overestimation might still have been present in instances where a small amount of hits was obtained. The torsion angle C6O7C8C9 is present in both GCV (G4) and ACV (A1), however, due to a different environment on C8, these were defined separately. One can notice from table 3.1 how the flexibility score of G4 was overestimated relative to A1. Almost all the entries in the hitlist of G4 had a similar chemical composition to GCV, with two hydroxyl groups that can adjust quite extensively to form intramolecular bonds, which can provide an extent of stabilisation. This might have contributed towards the higher degree of flexibility. Nonetheless, larger sample sizes would increase the level of confidence with respect to the interpretation of such cases.

The flexibility score was developed to provide a quantity which can reflect the degree of flexibility of a rotatable bond based on the kernel density approximation of the respective histogram. The objective was to capture the rigidity of the torsion angle through the relative frequency of the peaks detected, while taking into consideration their height and the number of peaks. This was achieved by using the weighted average based on the ratio of heights, thereby differentiating cases as A2 and F3 (figure 3.4). Both torsion angle histograms have three peaks, but F3 has almost equal peak heights and hence is less rigid than A2. A high weighted average of relative frequency implies that a large portion of the hits fall within very defined regions, therefore indicating that the probability of having a torsion angle outside such regions is low. In such scenario, the rotatable bond is considered to have a fairly high extent of rigidity. The equation of flexibility score takes the complement of the weighted average of relative frequency, given that latter represents rigidity.

Table 3.1: The quantification of flexibility of each rotatable bond, whereby each row represents a peak in the histrogram. The first three torsion angles are common for ACV, GCV and VCV.

Code	Torsion Angle	Min (°)	Max (°)	Relative Frequency	Weight of heights	Rigidity	Flexibility
A, G, V	C1N5C6O7	80.422	110.422	0.356	0.510	0.181	0.644
		249.578	279.578	0.356	0.490	0.175	
A, G, V	C4N5C6O7	68.855	98.855	0.373	0.505	0.188	0.644
		268.373	298.373	0.339	0.495	0.168	

Table 3.1 (continued)

Code	Torsion Angle	Min (°)	Max (°)	Relative Frequency	Weight of heights	Rigidity	Flexibility
A, G, V	N5C6O7C8	60.181	90.181	0.339	0.452	0.153	0.672
		165.723	195.723	0.068	0.071	0.005	
		269.819	299.819	0.356	0.477	0.170	
A1	C6O7C8C9	61.627	91.627	0.058	0.078	0.004	0.419
		118.012	148.012	0.008	0.015	0.000	
		165.723	195.723	0.694	0.827	0.574	
		266.928	296.928	0.033	0.051	0.002	
		290.060	320.060	0.017	0.030	0.000	
A2	O7C8C9O10	0.000	28.012	0.076	0.008	0.001	0.630
		6.687	36.687	0.010	0.008	0.000	
		51.506	81.506	0.437	0.455	0.198	
		165.723	195.723	0.081	0.085	0.007	
		220.663	250.663	0.005	0.007	0.000	
		278.494	308.494	0.391	0.418	0.164	
		311.747	341.747	0.015	0.017	0.000	
G4	C6O7C8C9	96.325	126.325	0.212	0.526	0.112	0.860
		209.096	239.096	0.061	0.474	0.029	
G7	O10C9C8C11	39.940	69.940	0.190	0.223	0.043	0.693
		118.012	148.012	0.005	0.007	0.000	
		167.169	197.169	0.386	0.420	0.162	
		230.783	260.783	0.005	0.007	0.000	
		287.169	317.169	0.300	0.342	0.103	
V1	O10C9C8O7	52.952	82.952	0.404	0.415	0.168	0.642
		103.554	133.554	0.002	0.003	0.000	
		133.916	163.916	0.002	0.003	0.000	
		165.723	195.723	0.110	0.136	0.015	
		213.434	243.434	0.001	0.002	0.000	
		277.048	307.048	0.407	0.431	0.176	
		342.108	372.108	0.005	0.009	0.000	
V2	C9C8O7C5	70.301	100.301	0.073	0.092	0.007	0.638
		165.723	195.723	0.500	0.683	0.341	
		238.012	268.012	0.064	0.110	0.007	
		262.590	292.590	0.064	0.116	0.007	
V4	C8O7C5C4	164.277	194.277	0.990	0.997	0.987	0.013
		201.867	231.867	0.005	0.003	0.000	
V6	O7C5C4C2	63.072	93.072	0.077	0.153	0.012	0.731
		119.458	149.458	0.103	0.183	0.019	
		245.241	275.241	0.359	0.664	0.238	
V8	C5C4C2C1	39.940	69.940	0.234	0.244	0.057	0.678
		159.940	189.940	0.403	0.449	0.181	
		285.723	315.723	0.274	0.307	0.084	
V9	C11O10C9C8	63.072	93.072	0.071	0.104	0.007	0.488
		168.614	198.614	0.612	0.819	0.501	
		265.482	295.482	0.047	0.077	0.004	

Table 3.1 (continued)

Code	Torsion Angle	Min (°)	Max (°)	Relative Frequency	Weight of heights	Rigidity	Flexibility
F2	C17N9C8C7	65.964	95.964	0.225	0.318	0.072	0.717
		162.831	192.831	0.063	0.087	0.005	
		200.422	230.422	0.025	0.049	0.001	
		271.265	301.265	0.375	0.545	0.205	
F3	N9C8C7C6	50.060	80.060	0.263	0.242	0.064	0.656
		164.277	194.277	0.400	0.393	0.157	
		281.386	311.386	0.338	0.365	0.123	
F4	C8C7C6C5	45.723	75.723	0.216	0.270	0.058	0.704
		165.723	195.723	0.392	0.407	0.159	
		274.157	304.157	0.243	0.324	0.079	
F5	C7C6C5O4	45.723	75.723	0.250	0.258	0.064	0.664
		165.723	195.723	0.339	0.371	0.126	
		285.723	315.723	0.393	0.372	0.146	
F6	C6C5O4C2	99.217	129.217	0.044	0.087	0.004	0.531
		167.169	197.169	0.578	0.780	0.451	
		251.024	281.024	0.111	0.133	0.015	
F8	C5O4C2O3	0.000	20.783	0.003	0.004	0.000	0.007
		164.277	194.277	0.997	0.996	0.993	

The flexibility score of each rotatable bond is shown in table 3.1, together with the respective relative frequency. The figures vary from 0.007 which belongs to C5O4C2O3 (F8) in FCV, mirroring no tolerance towards flexibility, to a maximum of 0.860 that is associated with C6O7C8C9 (G4). This method of quantification performed well because it was also able to differentiate between histograms of similar profiles as F3 (F = 0.656) and F4 (F = 0.704), where the latter has wider peaks which suggest a greater level of flexibility. Such results support how the procedure can adequately quantify what can be perceived qualitatively from torsion angle histograms.

3.3.1.2 Molecular flexibility

The geometric parameters for the potentially energetically accessible conformational space of GCV, FCV and VCV are represented in figure 3.5 through rug plots, each associated with the respective histogram of torsional data. The suggested ranges of angular data are reliable because they were extracted from structurally determined forms with relevant chemical environments. Such reliability was tested through the construction of conformational energy plots for GCV torsion angles in *Material Studio*. The most energetically favourable torsion angles ranges were fairly similar to those suggested by the CSD histograms, with minor variances when dealing with histograms of fewer hits (presented in brackets in figure 3.5). The accessible range for torsion angles as G1 (C1N5C6O7) is more easily predicted due to the clear repulsion present in case of having the C6O7 branch in parallel to the aromatic system. It was noticed how G4 (C6O7C8C9) and G7 (O10C9C8C11) had conformers distributed over a wider range of energy, thus suggesting a greater influence on the molecular geometry.

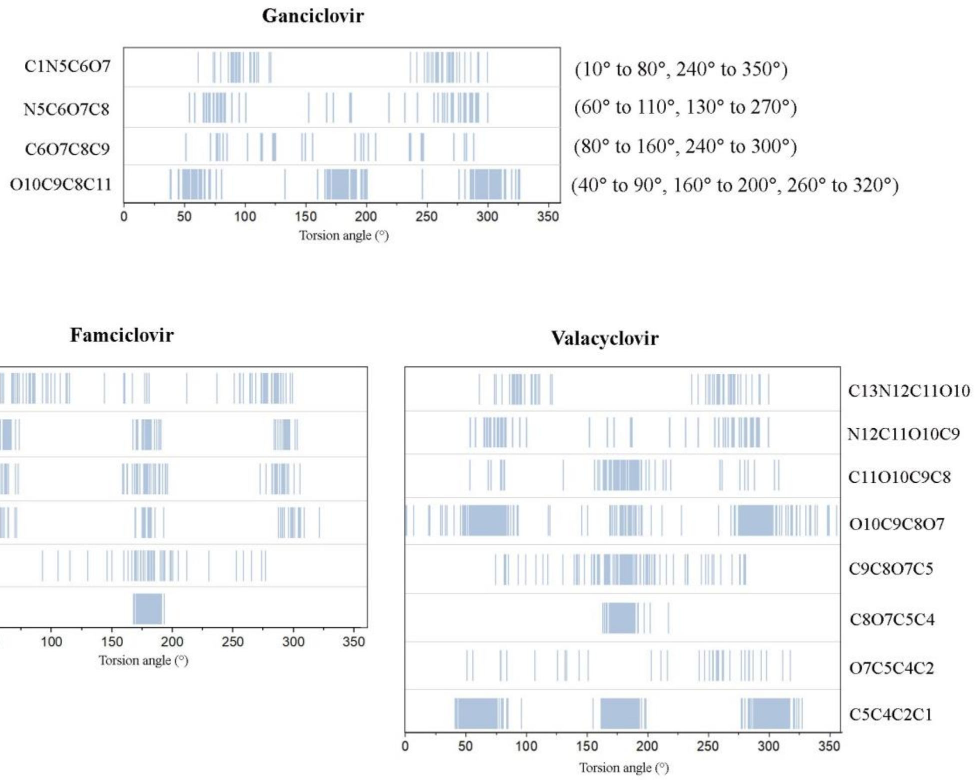


Figure 3.5: Rug plots displaying the torsional data from CSD search for each antiviral, where each value is represented by a line. The ranges of torsion angle values associated to relatively low energy conformers for GCV, calculated in *Material Studio*, are shown in brackets.

The molecular flexibility score was computed by summing the individual rotatable bond scores within the molecule, as presented in table 3.2. ACV and VCV calculations included G1 and N5C6O7C8 (G3) scores because they exist in both molecules with the same long range chemical environment. In cases of having data of different torsion angles of the same rotatable bond, only one was utilised for the flexibility score calculation, given that the datasets contained the same hits with different readings. Ring flexibility was not considered in any of the scores because the structure of the aromatic system has almost no space of rotation, thereby leading to a negligible effect on the overall molecular flexibility. In addition, the scores were used for comparison purposes, amongst four antivirals with a similar aromatic skeleton. However, if required, ring angles can be extracted from the rotamer libraries of the CSD and included within the flexibility score.

Table 3.2: Molecular flexibility scores

Acyclovir			**Ganciclovir**		
G1	C1N5C6O7	0.644	G1	C1N5C6O7	0.644
G3	N5C6O7C8	0.672	G3	N5C6O7C8	0.672
A1	C6O7C8C9	0.419	G4	C6O7C8C9	0.860
A2	O7C8C9O10	0.630	G7	O10C9C8C11	0.693
			G7	O12C11C8C9	0.693
Flexibility Score		**2.366**	Flexibility Score		**3.561**
Famciclovir			**Valacyclovir**		
F2	C17N9C8C7	0.717	G1	C13N12C11O10	0.644
F3	N9C8C7C6	0.656	G3	N12C11O10C9	0.672
F4	C8C7C6C5	0.704	V9	C11O10C9C8	0.488
F5	C7C6C19O20	0.664	V1	O10C9C8O7	0.642
F5	C7C6C5O4	0.664	V2	C9C8O7C5	0.638
F6	C6C19O20C21	0.531	V4	C8O7C5C4	0.013
F6	C6C5O4C2	0.531	V6	O7C5C4C2	0.731
F8	C19O20C21C23	0.007	V8	C5C4C2C1	0.678
F8	C5O4C2O3	0.007			
Flexibility Score		**4.479**	Flexibility Score		**4.505**

When analysing the flexibility scores, one should particularly focus on the differences between such values in relation to the NRot and the molecular size of the targeted molecules. Even though it is expected that ACV has the lowest flexibility score due to its size, it is difficult to predict the influence of each specific rotatable bond on the overall flexibility of the respective molecule. The inclusion of one rotatable bond can provide access to a much larger conformational space, as shown by the comparison of ACV and GCV, even when taking into account the possible overestimation of G4. Such effect was also indicated by the fact that the conformer generator for set B reached the limit of 500 conformers for GCV, whereas it produced only 203 conformers for ACV. Conversely, the disparity between the flexibility scores of GCV and FCV does not reflect the fact that the latter has four more potentially rotatable bonds. This implies that level of flexibility is influenced by the NRot, but such descriptor does not provide an exhaustive representation of it, even for molecules which are complaint with Lipinski's rules and considered small.[253]

The marginal gap between the flexibility scores of FCV and VCV is not a definite indicator of any possible significant difference in their respective flexibility. Results display that both molecules have seven bonds which contribute significantly to different extents, towards their overall flexibility. The scores illustrate how both antivirals have rotatable bonds which are very restricted, causing the NRot to overestimate their level of flexibility, especially when compared to other molecules.

3.3.2 Exploring flexibility through conformer generation

Another route for the classification of flexibility was conducted through the generation of conformers and the subsequent analysis of their Hirshfeld surfaces, characterised by several descriptors. The construction of the Hirshfeld surface was aimed to capture the extent of folding for each conformer. Molecular flexibility of each antiviral was mirrored by the level of variability extracted from the descriptors of its respective conformers. Furthermore, the difference in flexibility between antivirals was reflected through the comparison of the within-group variance values of each antiviral.

The generation of conformers was carried out using two routes because it was required to gather a sample which contained a variate but feasible group of conformers, without having any underlying bias. The use of energy thresholds would automatically exclude high energy conformations, which if characterised by a more extended geometric configuration than their lower energy counterparts, might have a higher chance of forming a solid-state.[90] This limitation is partially compensated when using the conformer generator provided by the CSD, because mutations are based on data of existing forms. However, this process might still produce unfavourable conformers because it follows a fragment based approach, where it can select highly probable torsion angles which interact together in an unlikely manner.[89] The construction of histograms of torsion angles characterising the conformers generated by the statistical approach mirrored the same trends as those presented in the previous section, hence confirming the use of the CSD rotamer libraries in the process. Given such scope, the idea of prioritising geometrically distinct conformations over similar ones, while keeping within reasonable chemical environments, was crucial. This was achieved by generating set C primarily on the geometric conformational domain of a molecule in terms of torsion angles.

3.3.2.1 Molecular flexibility in different sets

The extent of conformational variation within sets was captured through the average of RMSD values calculated from the superimposition of each constituent conformer pairs. Table 3.3 illustrates how in general, the within-set levels of structural dissimilarity for A and B were fairly equivalent, except for GCV. This suggests that both thermodynamic and hybrid routes portrayed a similar degree of flexibility for each antiviral. Conversely, the intended maximisation of the conformational space coverage within set C was verified through the relatively high RMSD averages particularly for ACV (1.657) and FCV (2.392). The highest average of the within-set RMSD values for VCV and GCV belonged to set A and B respectively, with a minimal discrepancy from the averages of sets C.

Table 3.3: The average of the RMSD values (including hydrogen atoms) when comparing pairs of conformers from two sets. (Please see appendix for the 95% confidence intervals associated with each average.)

		Set A	**Set B**	**Set C**
Acyclovir	Set A	1.543	1.521	1.612
	Set B	1.521	1.563	1.583
	Set C	1.612	1.583	1.657
Ganciclovir	Set A	1.273	1.711	1.666
	Set B	1.711	1.728	1.655
	Set C	1.666	1.655	1.692
Famciclovir	Set A	2.199	2.237	2.291
	Set B	2.237	2.126	2.247
	Set C	2.291	2.247	2.392
Valacyclovir	Set A	2.418	2.663	2.598
	Set B	2.663	2.403	2.322
	Set C	2.598	2.322	2.409

Nonetheless, the level of variation within certain torsion angles of conformers selected in set C for VCV and more particularly FCV suggested that it might be the case that two principal components were not enough to capture fully the variability of relatively flexible rotatable bonds in these two antivirals. Due to the circular nature of the chosen technique, it was not possible to evaluate the appropriateness of the generated principal components with respect to the variability present in the descriptors. Furthermore, given the bulkiness of the branches on FCV, simultaneous small alterations in flexible rotatable bonds might cause a considerable change in the overall molecular shape, even if the numeric variance in the individual angles is not large. This effect might not have been detected through the approach employed, given that individual torsion angles were the only descriptors used.

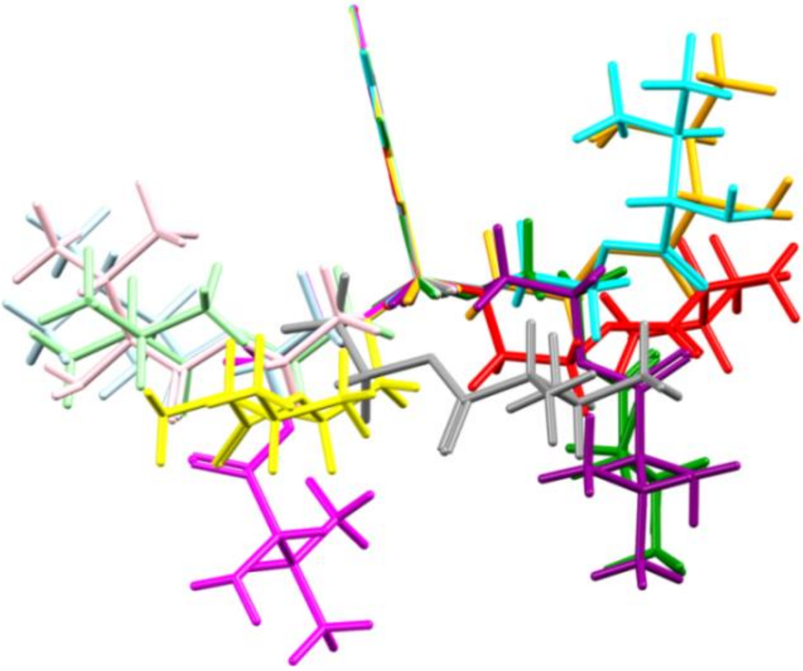


Figure 3.6: Set C for VCV with the aromatic ring systems superimposed on each other in *Mercury*.

The impact of simultaneous minimal changes in torsions was also evident for VCV, probably to a lesser extent than FCV. Even though C13N12C11O10 torsion angles are typically constrained to two narrow ranges, the presence of several consecutive torsion angles still enabled the generation of markedly different straight chain configurations, yielding the highest averages for RMSD values. As illustrated in figure 3.6, the majority of the conformers were folded to varying degrees to either sides of the aromatic rings (ex. cyano conformer), whereas others had an extended configuration (ex. magenta conformer). The most unique conformer (grey) for VCV can be identified by its loose spiral staircase-like branch.

Measurements of areas on Hirshfeld surfaces with discrete ranges of d_{norm} were utilised to represent the progression of the degree of folding for each molecule, and these were analysed through boxplots. On comparison, the conformers of set A were folded to a similar extent (small boxplots) within group for ACV and GCV, whose geometric resemblance is also highlighted through the relatively small averages in table 4.4. Such correlation can also be noted with comparatively high RMSD averages for set C (ACV and FCV), set A (VCV) and set B (GCV). A visible characteristic in the boxplot of VCV is how the maximum area with d_{norm} between 0 Å to 1 Å is much higher for set A when compared to those corresponding to set B and C. The small measure of d_{norm} for this specific range (0 Å to 1 Å) suggests that the conformers of set A for VCV were relatively more folded opposed to the rest of the sample. In case of FCV, the boxplots demonstrate how the conformers in set B were in general more extended compared to those in other groups, due to smaller area values at shorter d_{norm} values (figure 3.7).

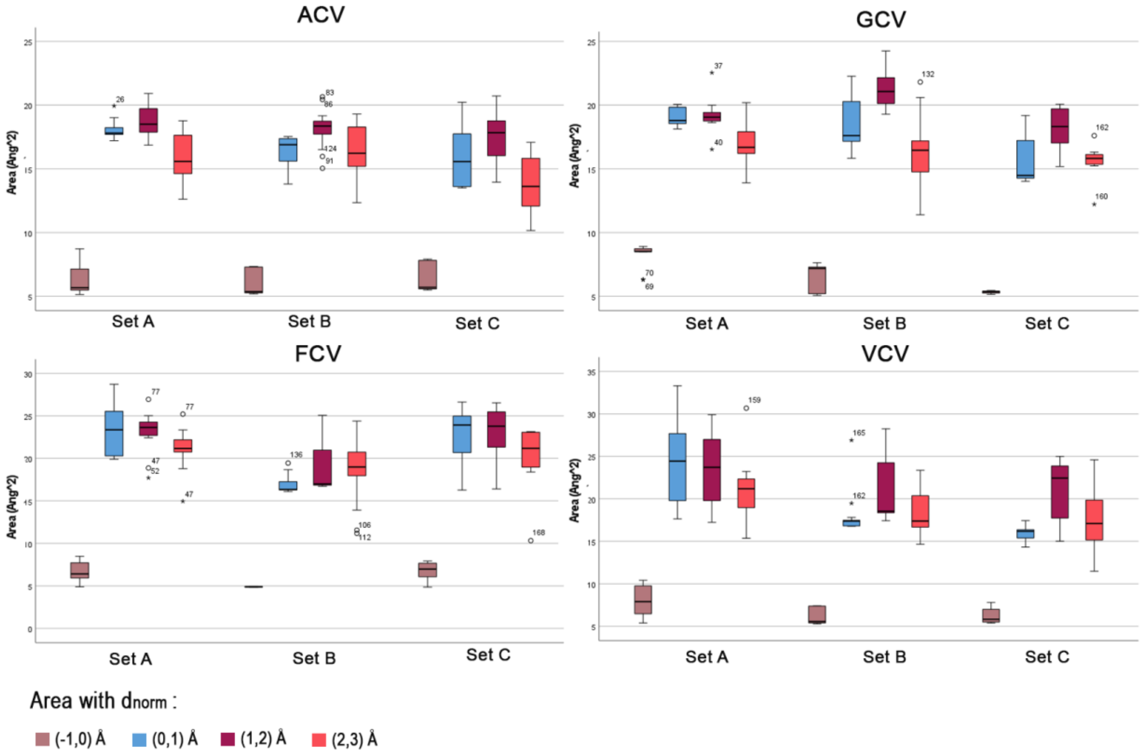


Figure 3.7: Boxplots showing the area (Å^2) on the Hirshfeld surface with different d_{norm} values, for each antiviral. The sign in front of the d_{norm} value is only a reference to the direction relative to the surface.

3.3.2.2 Molecular flexibility amongst antivirals

3.3.2.2.1 Individual descriptors

The construction of Hirshfeld surfaces was conducted on the premise that the change in the configuration between conformers can be expressed to a sufficient degree by the variation of their respective Hirshfeld surfaces, primarily through changes in characteristics as extreme points and area ranges. This hypothesis was tested by analysing visually the distribution and number of extreme points, in relation to the geometric conformation of each conformer using *VMD*.[249] Observations indicated that in general, the number of extremes on a Hirshfeld surface increased with the level of folding (figure 3.8). In addition, this premise was examined through Spearman's correlation test between the number of extreme points and variables describing the area measurements with discrete d_{norm} ranges (see appendix).[257] Relatively high, statistically significant correlation coefficients (0.01 level of significance), which ranged from 0.306 to 0.739, supported the deductions made visually even further.

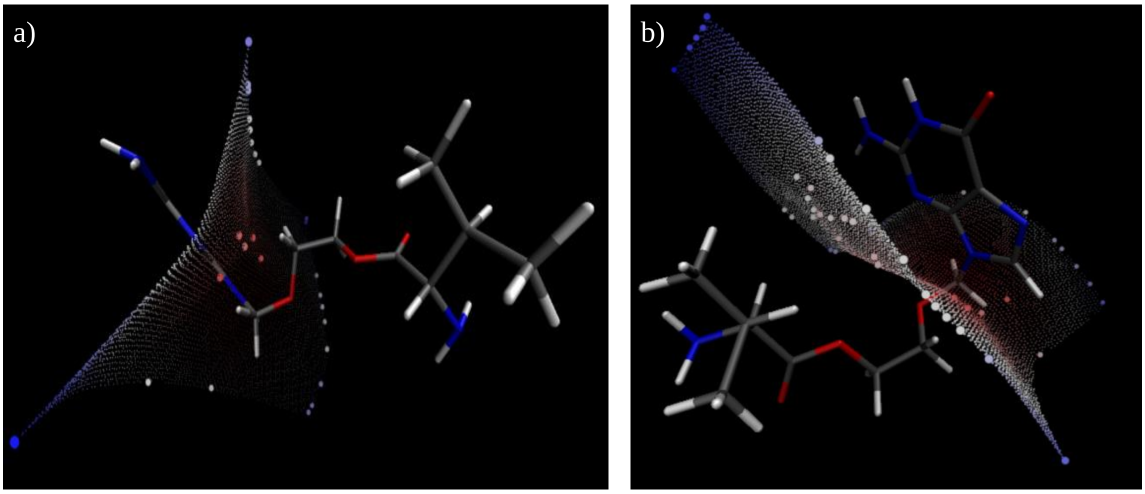


Figure 3.8: Conformers of VCV; a) Extended conformation with 5 extreme points, b) Folded conformation with 17 extreme points (maximum number of points).

The standard deviation of the number of extreme points was the lowest for ACV and the highest for VCV, indicating that the latter had the greatest within group variation, despite having a smaller NRot than FCV (table 3.4). A similar trend was observed for the variable of the number of extremes with a d_{norm} range of (-2, 2 Å) (equivalent to 4 Å), which was introduced to represent specifically higher instances of folding. For both descriptors, FCV had the second highest standard deviation, which emphasises its considerably large accessible conformational region. The same pattern was noted for the standard deviations of the area measurements at different d_{norm} ranges, except for area with d_{norm}(-1,0 Å) where GCV exhibited the second highest value. This hierarchy in flexibility was supported by the averages of RMSD values in table 3.3, where ACV had the lowest averages and standard deviation, depicting how its extent of conformational variation was the smallest relative to the rest of antivirals. Such trend is also illustrated in figure 3.9, where the data associated with FCV and VCV stretches along the majority of the range. In addition, figure 3.9a also portrays the positive relationship between the number of extreme points on the Hirshfeld surface and the area variable, as indicated by the Spearman's correlation test.

Table 3.4: Standard deviations of individual variables. The sign associated with d_{norm} values is relative to the Hirshfeld surface (for direction purposes only).

	ACV	GCV	FCV	VCV
Number of extremes	2.212	2.265	4.754	6.050
No. of extremes with d_{norm} (-2,2 Å)	2.130	2.097	4.666	5.517
Chosen standard deviation	0.631	0.528	0.547	0.419
Distance (Mean and global min in Å)*	0.669	0.627	0.699	0.918
Grouped units - Area with d_{norm}(-2,1 Å)	2.409	3.007	4.911	5.978
Area with d_{norm}(-1,0 Å)	1.198	1.432	1.271	1.601
Area with d_{norm} (0,1 Å)	1.668	2.053	3.792	5.015
Area with d_{norm} (1,2 Å)	1.561	1.887	3.595	4.139
Area with d_{norm} (2,3 Å)	2.218	2.212	3.515	3.570

*Distance between the mean position of all the extreme points used and the global minimum

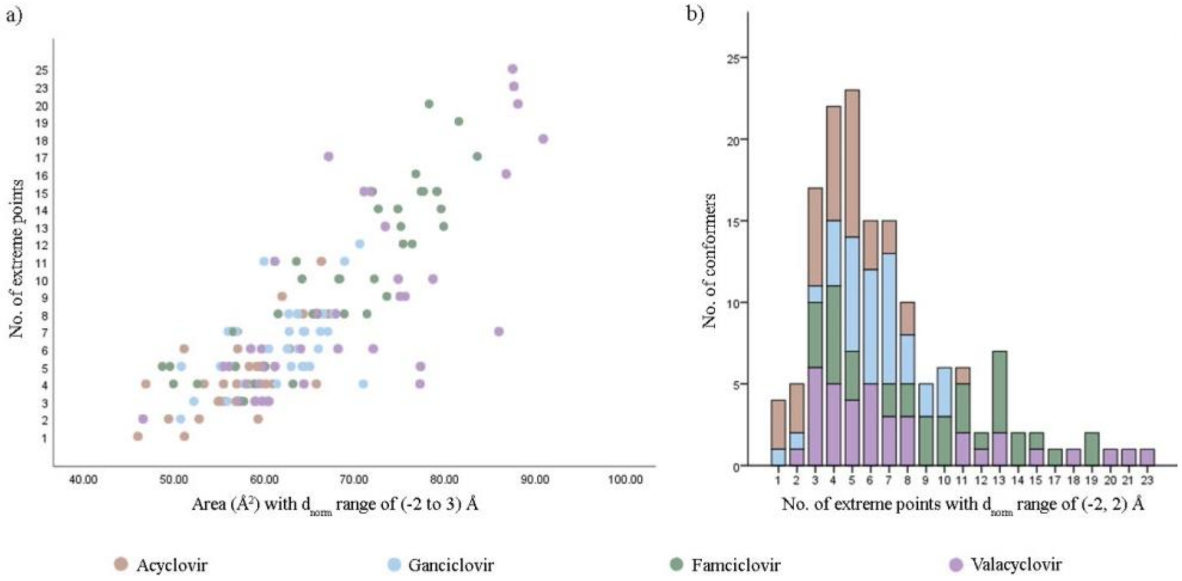


Figure 3.9: a) Scatterplot showing the relationship between the number of extreme points present b) Stacked bar chart showing the number of conformers with specific number of extreme points with a d_{norm} range of (-2,2) Å.

The distribution of the extreme points per conformer was illustrated through the standard deviation of their respective x, y and z-coordinates. However, given that the molecules were centred differently, only the highest standard deviation with respect to the three directions per conformer was used in the analysis. As opposed to previous observations, the minimum standard deviation of this variable (listed as “Chosen standard deviation” in table 3.4) was associated with VCV, whereas the maximum corresponds to ACV, with a relatively small discrepancy. This result could be explained by the fact that in certain cases, even though the straight chain of VCV was changing from one conformer to the other, some parts of it were too distant from the Hirshfeld surface to be detected, thereby producing similar extreme point distribution. This presented to be a relevant limitation of this approach, which was partially compensated by the large variety of properties used and information gathered in the previous sections. It also demonstrates how the success of this methodology depends on the location of the construction of the Hirshfeld surface and the overall shape of the molecule. Given the evidence provided by other descriptors, one can conclude that this variable might have underestimated the variability of the conformers with longer branches, particularly VCV and FCV. Conversely, given the smaller size of ACV, every alternation in the straight chain was mirrored on the surface.

3.3.2.2.2 Multivariate analysis

The majority of the individual descriptors indicated an increase in flexibility following the order of ACV, GCV, FCV and VCV. The presence of unequal degree of group dispersions (grouped by antiviral) was also highlighted through univariate tests of homogeneity of variance for individual descriptors (see appendix). However, it was required to test the equivalence of within-group variances of the antivirals using all descriptors collectively, in attempt to capture the entirety of their flexibility. This analysis was performed using Anderson’s permutation test (PERMDISP2 procedure), which examines the multivariate homogeneity of groups dispersions, without being dependant on parametric theory or influenced by unbalanced designs.[258-259] The spatial median was used as the centroid in the multivariate space, and a correction was applied to compensate for the bias in the presence of small, unequal group sizes.[260]

The objective for the selection of variables was to include the maximum amount of descriptors which can reflect the extent of folding, based on the individual variable analysis, while ensuring that features at different distances from the Hirshfeld surface were being represented effectively. All variables in table 3.4, excluding the extreme points with d_{norm} range (-2,2 Å) and chosen standard deviation, were utilised in the analysis. The null hypothesis of this procedure states that *"the average within-group dispersion (measured by the average distance to group centroid and as defined in the space of the chosen resemblance measure), is equivalent among the groups."* [259] The result (F=11.21, P-value = 0.0001, no. of permutations= 9999) suggests that there was a statistically significant difference between the within-group variability of antivirals, indicating how one or more groups have non-equivalent flexibility. A principal component analysis of the normalised data displayed in figure 3.10 illustrates the dispersion of each antiviral, together with the standard deviational ellipse. This plot shows an apparent increase in dispersion through ACV, GCV, FCV and VCV, thereby reflecting their respective degree of flexibility. This pattern is coherent with the observations derived from multiple descriptors discussed previously. It is noteworthy to mention that ACV (least flexible) has the most structurally determined structures of anhydrous forms, followed by GCV.

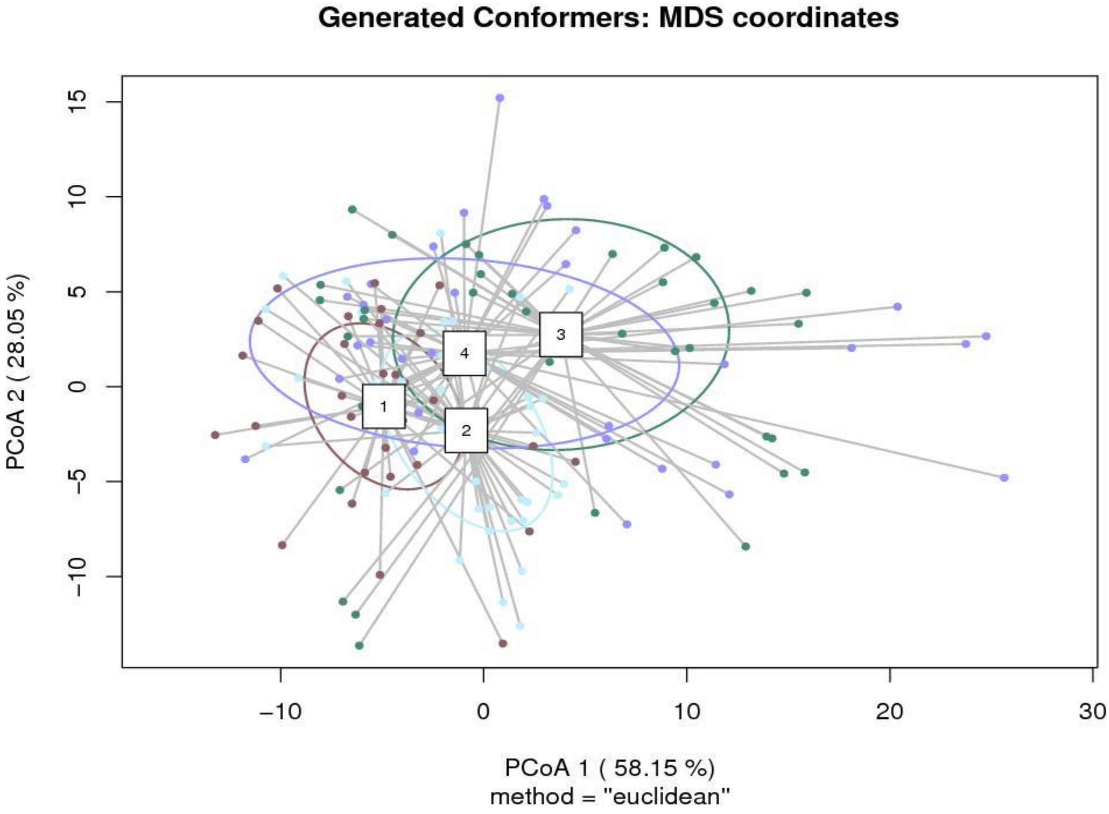


Figure 3.10: Graph showing first and second principal components (Multidimensional scaling), together with the standard deviation ellipse. (1 = ACV, 2 = GCV, 3 = FCV, 4 = VCV).

Pairwise comparisons of group mean dispersions were conducted; yielding permuted p-values < 0.05, hence indicating how the within-group dispersions were significantly different between all, except for the comparison of the dispersions of FCV and VCV, whose difference resulted to be statistically non-significant (see appendix). The similar level of flexibility of both antivirals can be attributed to the fact that their straight chains contain both very flexible and rigid fragments. Even though VCV has less bulky side chains, hence experiencing less hindrance towards movement, significant alteration is required to produce an effective change in the overall conformation. In contrast, FCV can compensate for the presence of more bulky groups, by having a fairly longer chain of carbon atoms, which exert much less resistance for rotation than the ether oxygen atoms present in VCV. These flexibility traits have parallel implications in relation to the respective energy profile of the antiviral. Their relatively high extent of flexibility is also exhibited through the difficulty of solving their crystalline structures, particularly of their anhydrous forms.

3.3.3 Comparison of results

The main objective of this investigation was to explore and quantify adequately the extent of molecular flexibility for each antiviral, beyond the use of the NRot. Despite the influence of the NRot on the degree of flexibility, it does not represent such property faithfully, primarily due to its underlying assumption that every rotatable bond has the same allowance for rotation. Such inadequacy was highlighted in literature for molecules which are not compliant with Lipinski’s rules.[220, 253] However, this investigation proved how NRot is also not an accurate measure of flexibility for molecules which are complaint and considered as small. The analysis illustrated how each rotatable bond has its own individual effect on the conformational space of the molecule, and such influence is highly misrepresented by a discrete number as NRot. This feature could be observed through the flexibility score differences in relation to the respective NRot (table 3.5), as well as through the individual descriptors and pairwise comparisons of variances, where it was shown how FCV does not have a higher extent of flexibility than VCV, as projected by its NRot. Conversely, the difference of only one flexible bond between ACV and GCV extends markedly the flexibility of the latter, as indicated by the additional generation of 297 conformers by the hybrid approach (set B).

Table 3.5: Comparison of different indicators of molecular flexibility

Antiviral	NRot	Kier's index[261]	Average distance to median*	Molecular flexibility score**
Acyclovir	4	2.837	5.697	2.366
Ganciclovir	5	4.194	7.041	3.561
Famciclovir	9	5.716	10.001	4.479
Valacyclovir	8	5.422	10.943	4.505

*Average distance to median in figure 3.10

**Proposed flexibility scoring system - section 3.3.1.2

The qualitative observations made throughout the analysis are mirrored in the molecular flexibility scores outlined in section 3.3.1.2. The scores are also in line with the multivariate analysis conducted on the Hirshfeld surface descriptors, highlighting the statistically non-significant gap between the degree of conformational variation of FCV and VCV, despite having different NRot. The average distance from data points to median per antiviral, in figure 3.10, produced from the multivariate analysis also support this novel scoring method. Lastly, the Kier flexibility index was calculated to test the performance of the flexibility score procedure, and on comparison, results are in close agreement despite methods relying on different descriptors.[222, 261]

The advantage of the proposed scoring system revolves around the integration of the torsion angles of existing crystalline forms, hence making such method more statistically favourable to be representative of the actual molecular flexibility. Furthermore, this method is capable to capture differences between regio-isomers (not stereo-isomers), because the definition of the chemical environment of the rotatable bonds in *CSD Python API* can be composed accordingly, as opposed to the NRot or Kier's index.[261] It also provides a clear representation of the contribution of each rotatable bond towards the cumulative molecular flexibility. This method is also computationally less intensive when compared to $NConf_{20}$, which estimates molecular flexibility by using a conformer generator to explore the conformational space, within specific energy threshold.[262] Nonetheless, it is vital to acknowledge that a larger and more diverse sample is required in order to reassure that such flexibility score is robust enough to become a standard procedure for a wide range of chemically distinct environments, especially for beyond-

Rule of five molecules.[220] The continuous growth of the CSD projects better future results as more precise hits can be obtained, thus reducing bias and increasing certainty.

3.3.4 Prediction of conformer propensity

The objective was to build a regression model that would be able to predict the probability of finding a particular conformation (irrespective of the approach used for its generation) within a solid form by inputting information concerning its atoms detected by the Hirshfeld surface. The model was constructed on the information gathered (electron density or local surface area) through the plotting of the Hirshfeld surfaces on a sample of 100 conformers of ACV, produced by the CSD. The predictor list contained all possible atoms within the molecule which were detected by the Hirshfeld surface in at least one conformation, while the dependant variable was the probability given by *CSD Python API* itself, associated with each conformer.

Several types of modelling approaches were attempted in order to find the optimal fit, however due to the continuous nature of the probability values within interval (0,1) and the possible shortcomings associated with certain transformation techniques, beta regression models were selected.[263] The flexibility of probability density functions for the beta distribution is enabled through its parameterisation in terms of the mean, μ and precision, ϕ, both of which need to be modelled accordingly using specific functions to link them with the explanatory variables. Further technical details regarding the beta regression model and its definition can be found elsewhere.[263-265]

Initial non-parametric correlation test (data diverged significantly from a normal distribution) was applied so as to remove any highly correlated variables to avoid multicollinearity problems. Due to the considerable list of variables and potential interactions, the relationship between the dependent variable and each baseline predictor was explored using univariate beta regression with a loose significance criterion ($\alpha = 0.20$). The variable was only removed if its respective p-values for both mean and precision were larger than 0.20. The resultant list of independant variables, together with their associated bivariate interactions were inputted into both mean and precision model. A one-step backward selection method was employed with $\alpha \leq 0.05$, until both mean and precision models contained only significant contributors.

A couple of final candidate models were obtained using electron density data of ACV conformers, but all of which had relatively low pseudo R-squared values, despite several efforts to maximise it. Boosted and extended beta regression models were also considered, but there were complications with respect to the interpretability of the final model.[250-251] Ultimately, the beta model with the best AIC value (referred to as model A) was selected for testing and will be discussed here. This model had a pseudo R-squared value of 0.3906, and the resultant residuals also did not illustrate a completely random pattern when plotted, exposing the possibility that the fitted model was not capturing all the required information. Similar issues were encountered when using local surface area, but to a much lesser extent. In fact the best fit (model B) for this dataset had an associated pseudo R-squared value of 0.6616. Both models performed better when the precision part was modelled by a function and not a constant. All details regarding both selected models can be found in appendix.

The predictive ability of both models was tested by inputting information of conformers present within known anhydrous forms of ACV (refcodes: MECWIC, MECWIC01, MECWIC03, MECWIC04). Relatively high probability values would have encouraged further refinements and deeper investigation of the performance of the models. However, the resultant probability values ranged from 0 to 0.93794, hence indicating that the chosen models were not capable to produce trustworthy predictions. This failure can be a product of several reasons, including very small probability values associated with the generated conformers, some of which were also identical. There was also a gap in the probability values which made it very difficult to model both parts simultaneously (see appendix). Such issue might have been resolved if the dataset was fitted with two models, each catering for different parts of the data. Furthermore, it would have been ideal to use the data of a much larger sample of conformers within known structurally determined structures of the same antiviral for testing, but this was impossible due to the limited amount of ACV solid forms in the CSD. Lastly, one has to take into consideration that the conformation within a solid form does not necessarily be equivalent to a stable conformation in isolation, especially in cases as MECWIC03 whose asymmetric unit is composed of two distinct conformers (their non-covalent network has a dominant effect on the thermodynamic stability).

Chapter 4

POLYMORPHISM

4.1 Introduction

Polymorph screening is an indispensable task in any industry which deals with crystalline solid-states, but given the extent of potential consequences, the pharmaceutical industry has higher risks, as explained in chapter two. Due to such importance, the understanding of polymorph mechanism and the prediction of the probability of the manifestation of this phenomenon will be at the centre of this chapter. The existence of polymorphism depends on the energetic stability of the respective solid-state form, which in return relies on the degree of intramolecular strain and the robustness of the intermolecular network of non-covalent bonding. Due to such dependence, the investigation of polymorphism requires an analysis of each structurally determined form, at different levels ranging from molecular to supramolecular. The examination of crystal structures provides an insight towards the qualification of the respective level of stability, which consequently has implications on the risk of having other potential unknown polymorphs. In addition, such study also assists towards understanding polymorphism and the influence of specific bonds and substituents on the stability of particular form. This information complements the risk assessment approaches based on the statistical analysis of data of known forms and on thermodynamic energy calculations.

4.2 Methodology

The main subject of analysis consists of the three antivirals (GCV, FCV and VCV·HCl) in their administered form, and hence all work, experimental and computational, revolved around characterising them, their extent of thermodynamic stability and their relationship with other forms produced during solubility screening.

4.2.1 Experimental methods

The objective of the following solubility screening was to perform a routine method, applied widely in research, so as to analyse the stability of the antiviral in various solvents and temperature profiles. Even though desirable so as to ensure exhaustiveness, the aim was not to specifically produce all the published forms and solvates of each antiviral. All information regarding the trials performed, particularly the solvents used and their grade can be found in appendix.

4.2.1.1 Solubility Screening

Manual solubility screening was mainly conducted on VCV·HCl, using solvents with a varying range of polarity indexes, so as to map the solubility tendencies of this antiviral. This method was applied because the solubility properties of VCV·HCl were not clearly characterised in literature. Small samples of around 6 mg were utilised in each trial, but scale-ups were necessary in case of achieving full solubility for characterisation purposes. It was ensured that the solvent was added dropwise so as to have a rough estimate of the solubility, and in cases where this was not achieved, heat was introduced to assist the process. Binary solvent mixtures involving water were also tested.

Solubility screening for GCV, FCV and VCV·HCl was also conducted using Technobis Crystal16$^{®}$ and Crystalline$^{®}$, following their specific workflow to obtain accurate solubility values. A variety of concentrations were used in different solvents, while ensuring that the temperature profile was within the freezing to boiling points range of the respective solvent and API. In some cases, particularly for GCV, very small concentrations were utilised in order to increase the chance of solubility. Before each run, the instrument was tuned in order to calibrate the transmissivity to an instance of complete solubility. Stirring was retained throughout the entirety of each process so as to avoid any instances of full transmissivity which did not correspond to full solubility. Each run was composed of three consecutive heating-cooling cycles, to check the repeatability of clear points. The data were processed in *CrystalClear*, which assisted in the identification of clear points, whose information (concentration and temperature) was then utilised to model the solubility of the antiviral in individual solvents.[266]

In trials where the solute dissolved completely, without any colour change, the sample was extracted for characterisation after re-crystallisation by slow evaporation.

4.2.1.2 Characterisation techniques

Hot-stage microscopy

A Leica Z16 APO microscope equipped with a QImaging QICAM Fast 1394 digital CCD camera, together with a heating and cooling system (Linkam stage), was used to monitor the change in crystal habit with temperature. Slow heating rates were employed so as to identify each respective melting point and possible decomposition. Given that the samples were powders, it was difficult to assess the degree of crystallinity.

Powder X-ray diffraction (PXRD)

PXRD patterns were collected on a STOE Stadi P diffractometer, using Cu K_α radiation (40kV, 40 mA) and a curved germanium monochromator, at room temperature. Unless specified otherwise, measurements were taken over the angular range (2θ) of 2° to 50°, with a scan time of 120 s/step. Due to the large number of samples and the setup available, the first screening for identification purposes, involved packing the samples in flat plates. When necessary and possible, measurements were repeated using 0.5 mm capillaries. Only forms with unique PXRD patterns were characterised further using STA, FTIR spectroscopy and vt-PXRD.

Variable temperature powder X-ray diffraction (vt-PXRD)

The variable temperature measurements were collected on a STOE Stadi MP diffractometer with Cu K_α radiation source (40kV, 40 mA), but using a high temperature attachment for capillaries 0.65.LT. The angular range varied from 2° to 50°, with a step size of 0.015°. Temperature profiles were set up to study the extent of stability of the crystal structure from room temperature till a few degrees beyond its melting point. Temperature ranges were varied according to the sample nature and objective of run, with a heating rate of 5°C/min, unless specified otherwise. Furthermore, the three administered forms were subjected to successive heating and cooling cycles (minimum temperature of -50°C) for exploratory purposes, as a route to screen for other potential polymorphs, similar to the synthesis of FCV form IV.[206] It was ensured that the maximum temperature of the first heating cycle did not exceed the point of decomposition.

Fourier Transform Infrared (FTIR) spectroscopy

FT-IR spectra were collected using a Shimadzu IRAffinity spectrophotometer, over a spectral range of 4000 cm^{-1} to 500 cm^{-1}. The procedure involved the preparation of samples in KBr disks.

Simultaneous Thermal Analysis (STA)

A Netzsch – STA 449 F3 Jupiter ASC was used to perform differential scanning calorimetry (DSC) and thermogravimetric analysis (TG) for various samples, with a steel furnace able to provide a maximum temperature of 1000°C. This instrument is coupled with a Gram Schmidt IR instrument and mass spectrometer QMS 403 Aëolos Quadro to detect the mass for any gaseous species evolved during heating. Approximately 3 mg of the sample was weighed into a concavus Al crucible, with a pierced lid. Both calibrations and experimental runs were performed in a dry Helium atmosphere (50 mL/min purge; 20 mL/min protective gas). All the profiles for STA mirrored the parameters set for vt-PXRD, so as to be able to understand better the changes/ effects observed. Only peaks occurring over a constant heating rate were taken into account for interpretation.

4.2.2 Computational methods

The idea behind the following trail of procedures was to first analyse the administered polymorphs of all the three antivirals, in order to have a better insight of their own extent of stability and to collect information regarding the mechanism of polymorphism. Such information was necessary to comprehend and assess the subsequent polymorph prediction methods.

4.2.2.1 Crystal Structure Analysis

The following selection of tools of analysis includes methods which require the knowledge of the structurally determined crystal structure and others where it is not necessary. This implies that the use of methods was conditioned by the type of information available, ranging from GCV having two determined anhydrates (form I and II) to VCV·HCl having none.

4.2.2.1.1 Non-covalent bonds analysis at different levels

The process of analysing a crystal form at intramolecular, intermolecular and supramolecular levels using the CCDC toolkit was thoroughly described in work performed earlier.[133, 162] Firstly *Mogul* 2021.2 (CSD last updated in May 2021) was utilised to explore the intramolecular geometry, as described by Galek *et al.*, while ensuring the exclusion of organometallic compounds.[160] All parameters were compared with those in fragments of similar chemical environments, increasing the relevance threshold when necessary, in order to check whether the intramolecular geometric features were within likely regions.

IsoStar 2020.3 was employed for the initial examination of non-covalent contacts within the anhydrous form, producing CSD-based scatterplots with nitrogen, oxygen and carbon as separate probes.[168] A default normalisation of the hydrogen atom positions was implemented in order to compensate for the intrinsic underestimation of such bond lengths by X-ray diffraction. This was followed by the construction of full interaction maps in *Mercury* 2021.2.0, which provided an insight of the type and location of potential non-covalent interactions, while taking into account any possible restrain or non-availability caused by the specific molecular conformational geometry. [267-268] In case of VCV·HCl, *Conquest* was used to investigate interactions involving the chloride ion as an acceptor and the competition between potential HB donors in such contacts.

The definition of a HB (hydrogen bond) in *Mercury* was modified to have an angle of more than 110° and the distance range was slightly extended further to be less or equal to the sum of VdW radii + 1.00 Å, to ensure that all potential interactions were included.[269] This was done due to the presence of several possible carbon HB donors, whose contacts are usually characterised by features associated with weak contacts, namely long distances and small angles.[269] This enabled the visualisation of the HB network within the structurally determined anhydrous forms. Intermolecular potentials between the molecule in the asymmetric unit and its neighbours were calculated using the empirical "UNI" force field, whose parameters were published by Gavezzotti.[270,271]

In order to have a better understanding of other types of contacts in structurally determined forms, beyond the conventional HBs, each CIF was also processed by *PLATON* V.30118.[272] Further analysis was performed through the construction of Hirshfeld surfaces on each of the anhydrous forms using *CrystalExplorer* V17.5.[113] The participants of such contacts were identified through fingerprint scatterplots, which map the distance from a point on the surface to the closest nucleus inside the surface, d_i against the distance outside the surface, d_e.

Lastly, the unit cell and packing patterns of each determined anhydrous form were examined to infer observations regarding its supramolecular architecture. The calculation of voids within the structure was conducted using a probe radius of 0.7 Å, with a grid spacing of approximately 0.7 Å. The solvent accessible surface was utilised for the calculation because it displays the volume that can be taken by the full probe, hence estimating the volume that is available for any solvent or guest molecule.[273] The morphology of the crystal structure was simulated using the Bravais, Friedel, Donnay and Harker (BFDH) theoretical method, ensuring that the resultant morphology was in proportion to the unit cell.

4.2.2.1.2 Energy frameworks and lattice energy

Energy frameworks were constructed for each structurally determined anhydrous form, in *CrystalExplorer* using the CE-B3LYP model (basis set: 6-31G(d,p)), taking into consideration only a cluster of molecules with 7.6 Å in diameter.[112] All frameworks were scaled equally so as to facilitate comparison between polymorphs. The energy model for a pair of interacting molecules is decomposed into distinct components (similar to Gavezzotti's PIXEL scheme), each of which can be quantified even visually through the energy frameworks (eq 4.1). The contributors are polarization energy, E'_{pol}, repulsion contribution, E'_{rep}, dispersion energy term, E'_{dis} and electrostatic energy, E'_{ele}, together with their respective constants; k_{ele} = 1.057, k_{pol} = 0.740, k_{dis} = 0.871 and k_{rep} = 0.618.[112]

$$E_{tot} = k_{ele}E'_{ele} + k_{pol}E'_{pol} + k_{dis}E'_{dis} + k_{rep}E'_{rep} \tag{4.1}$$

The lattice energy was estimated by summing these pairwise interactions over a finite number of molecules B interacting with a central molecule A, within a specific radius, R_{AB} until the difference in energy between iterations was less than 1 kJ/mol.[274] The calculation of the lattice energy was conducted based on the methodology provided by Thomas *et al*, who suggested further refinements in addition to the summation of $E_{tot}^{CE-B3LYP}$ in case of dealing with polymorphs (eq. 4.2). These refinements include the cell dipole energy correction, $E_{cell\,dipole}$ (eq. 4.3), where p_{cell} is the magnitude of the dipole moment of the unit cell, Z is the number of molecules within the unit cell and V_{cell} is its volume. The magnitude of the dipole moment was a result of a vector sum of molecular dipole moments present within the unit cell, obtained by a single point energy calculation of each unique molecule. This particular correction was only applied for crystal structures with a polar space group.

$$E_{lat} = \frac{1}{2}\sum_{R_{AB}<R} E_{tot}^{CE-B3LYP} + E_{cell\,dipole} + \Delta E_{intra} \tag{4.2}$$

$$E_{cell\,dipole} = \frac{-2\pi p_{cell}^2}{3ZV_{cell}} \tag{4.3}$$

As referenced in the literature review, an intramolecular energy penalty, ΔE_{intra} is required to avoid the overestimation of the magnitude of the lattice energy. The methodology for calculation of this penalty was based on the work done by Thomas and Spackman on the ROY polymorphs.[84] Geometry optimisation of each molecule at B3LYP-D3(BJ)/def2-TZVPD level of theory in vacuum was followed by a MP2/cc-pVTZ single point energy calculation, the latter being recommended for accurate conformer energies.[235, 275-276] Some geometric parameters were allowed to relax during the geometry optimisation of the conformation in the solid-state, while ensuring that any known intramolecular bonds remained present. Both refinements were calculated using *Turbomole*, using *TMoleX* as interface.[148]

4.2.2.1.3 Lattice energy using periodic boundary conditions

Given the typical small discrepancy between the lattice energies of polymorphs, it was required to utilise another method to ensure the reliability of the results obtained previously. In this approach, the definition of lattice energy is given in eq. 4.4, where E^{iss} represents the energy of an ideal static solid (perfect infinite crystal structure) and E^{isg} is the energy of its related static molecule in a gas phase (no interactions with other molecules), with both conformations at 0 K.

$$E_{lat} = \frac{E^{iss}}{Z} - E^{isg} \tag{4.4}$$

This technique was based on the work published by Buchholz and Stein, where they illustrated the ability of periodic DFT methods to evaluate lattice energies, with a relatively high degree of accuracy.[118, 123] Geometry optimisations were set to use PBE exchange correlation function, in conjunction to the Grimme's D3 dispersion correction and RI approximation.[234] The def2-TZVP set was chosen because it is a medium sized basis set, which provides a good balance between computational time and accuracy.[118, 235] The Γ-point centred mesh of k-points was set to 3 x 3 x 3, so as to take into account interactions with the neighbouring replicates of the unit cell. The sampling on such uniform k-points mesh was suggested to be sufficient to achieve convergence with respect to the lattice energies.[123] The unit cell parameters were not optimised because it was shown how for PBE-D3 calculations, such relaxation in unit cell does not lead to a better agreement with experimental data.[118] Other *Turbomole* default settings were retained (see appendix).

4.2.2.2 Risk assessment of polymorphism

4.2.2.2.1 Hydrogen Bond Propensity model for polymorph prediction

This procedure was performed in *Mercury* (last updated September 2021) following the guidelines described in earlier work.[133, 162] The definition of HB was set according to the results of the analysis performed in section 4.2.2.1.1, so as to make sure that the observed HB can be included within the list of hits. Each molecule was dissected into a set of substructures, which represented faithfully local chemical environments involving each potential HB contributor. Systems with errors or disorder were excluded from the study, together with organometallic compounds and entries with a R- factor > 0.075. It was empirical that each model was constructed on a relatively large training set (> 2000 hits), as well as ensuring that each functional group was represented with an adequate number of hits.

After processing the built-in examination of hits, donor or acceptor candidates with very small true hits were removed to avoid regression failures. A logistic regression model was fitted on the data, and assessed using the area under the ROC curve value, the log likelihood test p-value and the AIC parameter. Any non-statistically significant predictors were omitted in order to obtain an adequate parsimonious model. The HB propensity and coordination tables were computed after achieving a model with a satisfactory level of goodness of fit.

This statistical approach does not depend on the particular molecular conformation, as it dissects the molecule into different functional groups, to produce a model irrespective of the availability of each group to participate in HB. Due to such feature, apart from for structurally determined forms, this procedure was also performed for VCV·HCl, using a geometry optimised conformation at B3LYP/6-311++G** level of theory with dispersion correction.[277] Default settings were retained given that no information about its solid-state was yet known.

4.2.2.2.2 Crystal Structure Prediction

The following procedure was performed for GCV only, due to time and logistics limitations. Further details, particularly regarding quality criteria and settings can be found in appendix.

4.2.2.2.2.1 Conformational search

The geometrically optimised conformation from section 3.2.1.2 was used as an input in *Material Studio.* Conformer generation was parameterised using the feasible torsional angle ranges defined in section 3.3.1.2. Each search was allowed to navigate a specific accessible conformational region, defined by a particular combination of torsional angle boundaries. The settings outlined in section 3.2.1.2 were retained for geometry optimisation (COMPASSIII forcefield) without constraint. The unique lowest energy conformers were re-optimised by DMol^3 using PBE functional with TNP level of theory at a fine level.

4.2.2.2.2.2 Packing and Clustering

Polymorph predictor module was employed to generate different packing arrangements using a unique low energy conformation in the asymmetric unit, with Z' restricted to one. The packing arrangements were generated for the most common space groups according to the statistics extracted from the CSD (January, 2022): $C/2c$, Cc, $P1$, $P\text{-}1$, $P2_1$, $P2_1/c$, $P2_12_12_1$, $Pbca$, $Pna2_1$, $Pnma$.[278] During simulated annealing, the number of steps accepted before cooling was increased to 24 and specific torsion angles were allowed to move. This was done due to the presence of conformations in unit cells which did not coincide with a low energy conformer, as in the case of GCV form I.

The geometry optimisation of each structure was set at ultra-fine quality, using COMPASSIII (V1.1) and forcefield assigned charges, while allowing all geometric parameters to relax. These settings were chosen because they produced the highest crystal superimposition scores when comparing simulated unit cells with the experimental structures of GCV forms. The maximum amount of iterations was increased to 800 so as to ensure that every geometry optimisation converged.

The final step of the Polymorph predictor framework was clustering at an ultra-fine quality, using forcefield type as a clustering factor. For all space groups, each search was restricted to 500 structures and repeated six times to ensure a sufficient degree of convergence with respect to low energy structures. This procedure was performed for each low energy conformer found during the conformational search. Clustering was performed among structures generated from all runs, to remove duplicates or closely-related entries.

4.2.2.2.2.3 Re-ranking

After the final clustering, crystal structures were ranked according to their total energy and only entries within the lowest 60 kJ/mol energy window were retained for finer energy calculations. DMol^3 package was accessed through the QSAR models module so as to perform a single point energy calculation using the PBE functional and *Insulator* settings (see appendix) at a medium level on the remaining list of structures. This was followed by another energy window filter (50 kJ/mol), geometry optimisation (coarse level), single point energy calculation (medium level) and final window filter of 28 kJ/mol. The crystal structures in the final list were geometrically optimised at fine level (PBE/DNP), followed by a single point energy calculation using PBE functional with TNP basis set, dispersion correction and improved **k**-point separation to ensure convergence (0.05 1/Å).

4.3 Results and discussion

The following results section contains a combination of experimental and computational outcomes, which complement each other to provide a holistic view of each antiviral, particularly their administered form.

4.3.1 Ganciclovir

4.3.1.1 Form characterisation

GCV has four structurally determined polymorphs; two anhydrous conformational polymorphs I (USP reference standard) and II, and two hydrates, all of which have their respective PXRD pattern illustrated in figure 4.1. In addition, research also shows the presence of a structurally unsolved hydrate (form V), which was observed when heating a sample of form IV with traces of form III or re-crystallisation from water and acetonitrile solvent mixture.[199-200] Apart from having the most structurally determined forms, GCV also has the greatest amount of published literature, relative to FCV and VCV·HCl.[162, 197-201] Its composition with a variety of strong HB donors and acceptors, coupled with a considerable conformational space enabled by its torsional angles (shown in chapter three), provide the opportunity to exhibit polymorphism. However, an in-depth analysis is required to understand better the extent of thermodynamic stability of form I and define its ability to form other polymorphs.

By means of solubility screening, it was possible to transform form I into form II using dimethylformamide as solvent, whereas form III (trihydrate) was re-crystallised from water at room temperature (details in appendix). These forms were identified by comparing their PXRD patterns and FTIR spectra (figure 4.2) with those present in literature.[197-198, 201] None of the samples contained any trace of form IV (monohydrate). This was expected because the formation of the monohydrate is only preferred when the drying temperature falls between 50 - 60°C.[197]

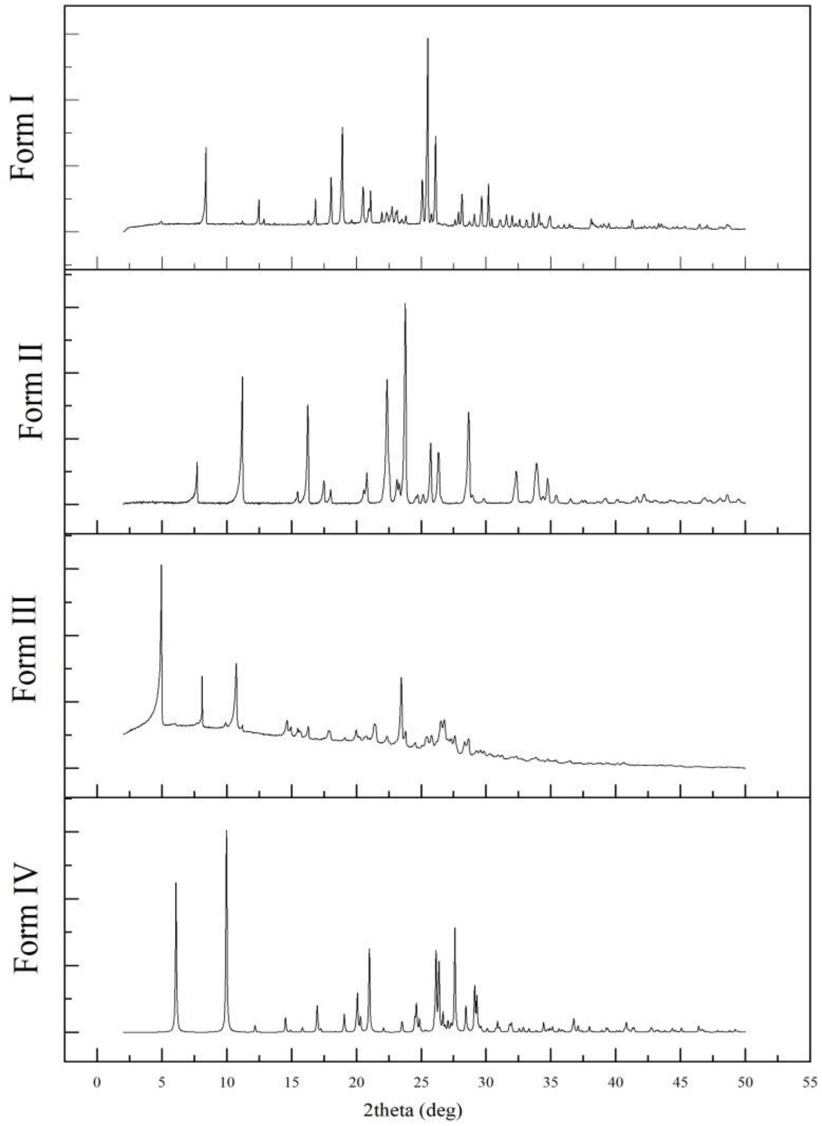


Figure 4.1: PXRD patterns of the structurally determined forms of GCV, all of which were collected during the characterisation process at room temperature, except for the one associated with form IV (simulated by *Mercury*).

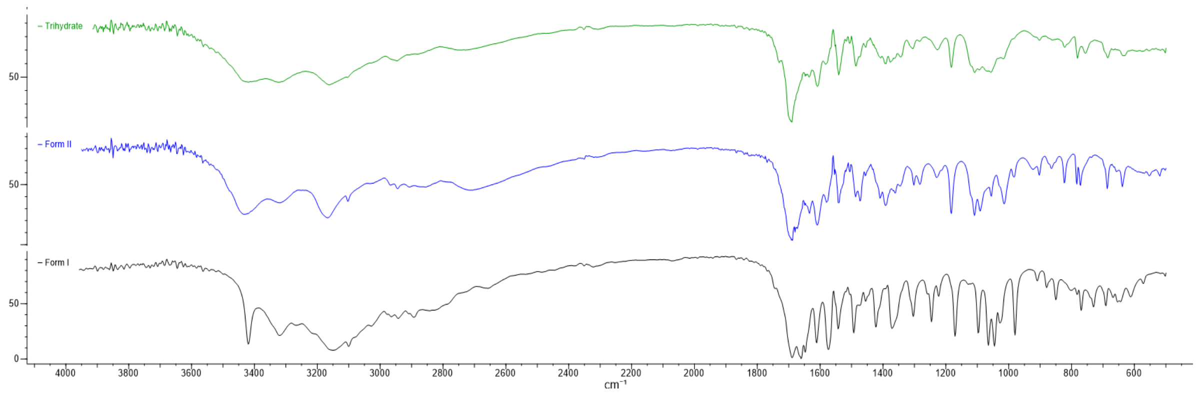


Figure 4.2: FTIR spectrum of form I (black), form II (blue) and form III (green).

The absorption band position, intensity and shape in IR spectroscopy can be influenced by various factors, including HB bonding because it alters the vibration mode of the participants. In general, a HB is expected to induce a shift in the band of X-H stretching to a lower frequency and increase its intensity.[279] Despite of not having a spectrum of isolated GCV molecule for comparison, the spectra in figure 4.2 still portray valuable information through their differences.. The most distinctive bands in the FTIR spectra are associated with the NH/OH and C=O stretching mode in the 3435 cm^{-1}–3420 cm^{-1} and 1750 cm^{-1}–1600 cm^{-1} regions, respectively. In particular, the broadness of the NH stretching band for form III is caused by the participation of this functional group in three distinct contacts, as well as due to an overlap with bands representing water. The unique sharp absorption band at 3419 cm^{-1} in the spectrum of form I might indicate a difference in the HB concerning its NH donor relative to those present in forms II and III.

4.3.1.2 Thermal stability – Experimental characterisation

Given the objective of the analysis, it was important to investigate the thermodynamic relationship between form I and II. Published studies demonstrate how the structure of form I remains consistent until the temperature range of 222°C to 228°C, where it transforms into form II which melts at around 252°C.[197, 201] For repeatability and completeness purposes, the thermal stability of form I and II was studied again in this work, using a slower heating of 5°C/min. This analysis employed the use of hot-stage microscopy (figure 4.3), STA (figure 4.4) and vt-PXRD (figure 4.5).

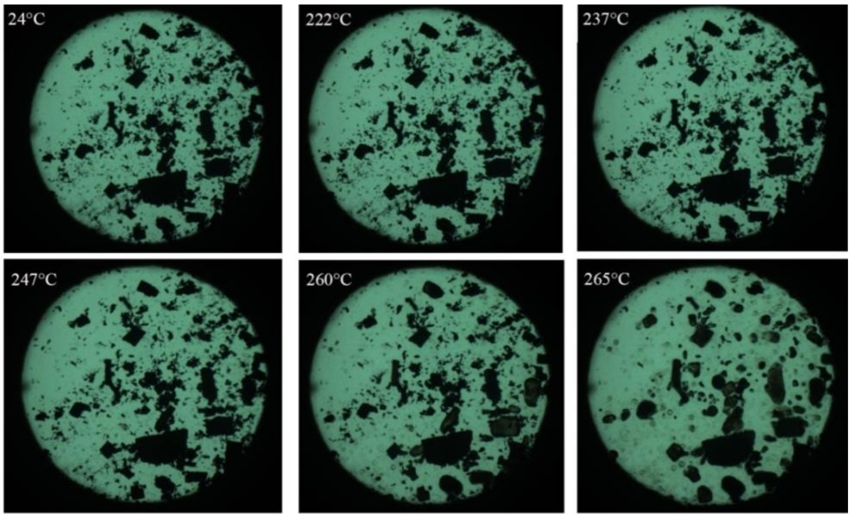


Figure 4.3: Hot-stage microscopy images (x2.5 magnification) starting with form I, set up with a temperature profile: heating from room temperature till 220°C at 10°C/min and decreasing the rate to 5°C/min till 280°C.

Due to the lack of single crystals, it was difficult to detect any morphological features from powder samples under the microscope, apart from its preference to aggregate in rectangular crystals. A sample of form I was heated and no significant changes occurred till the temperature rose above 234°C, apart from minor movements of small particles over the range of 217°C to 230°C. The first disappearance of minuscule crystals due to melting was first observed at 234°C, and melting gradually proceeded as temperature increased. During the melting process, it was not possible to determine the temperature at which decomposition started, however it was noted that these two processes overlapped.

The DSC graph of form I shows an endotherm starting at 224°C, without any weight loss, indicating the phase transition to form II. The presence of this endotherm confirmed the enantiotropic relationship between the forms, as claimed in literature, in agreement with the temperature range. As observed through the images of microscopy, this solid-solid transition was not preceded by a complete melting process. PXRD data demonstrated that there was no point prior to the formation of form II, where the sample did not retain its crystalline nature. In particular, the diffraction patterns highlighted the presence of a mixture of both forms at 220°C and pure form II at 223°C, thus suggesting that the phase transition occurred at a slightly lower temperature range (prior 220°C) than that defined in literature.

The second peak in the DSC plot of form I relates to the melting of form II because, despite a minor displacement, this endotherm (peak: 246°C) parallels the only major peak (peak: 241°C) present in the DSC graph of pure form II. A similar difference was observed in literature between the temperatures at which the melting peaks of form II occurred, with the one occurring after phase transition (252°C) being around 3°C higher.[197] TG curves display an initial decrease in mass along the melting range of form II, indicating an overlap with decomposition as suggested through the microscopy images. The temperature range disparity for the melting/ decomposition process of form II obtained in this work and in literature can be attributed to different heating rates, with the latter being at 12°C/min. In fact, this endotherm peak occurred at 253°C when the heating rate was increased to 12°C/min (see appendix). The kinetic features of decomposition might be the cause of such dependence on the heating rate. The degradation of form II prior to a full melting process was also supported by Roque-Flores *et al.*[201]

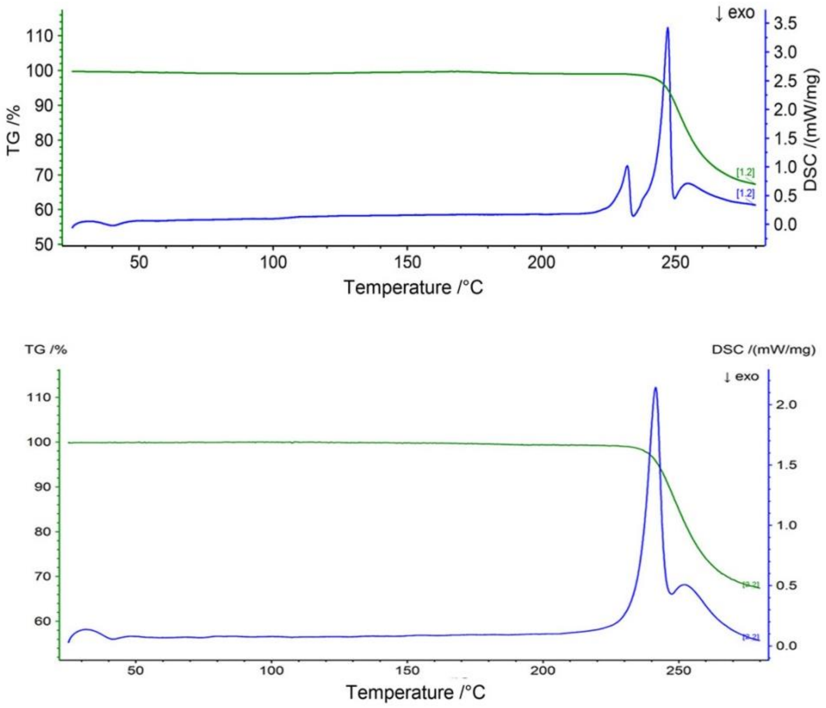


Figure 4.4: DSC-TGA overlay of form I (top) and form II (bottom), at a heating rate of 5°C/min.

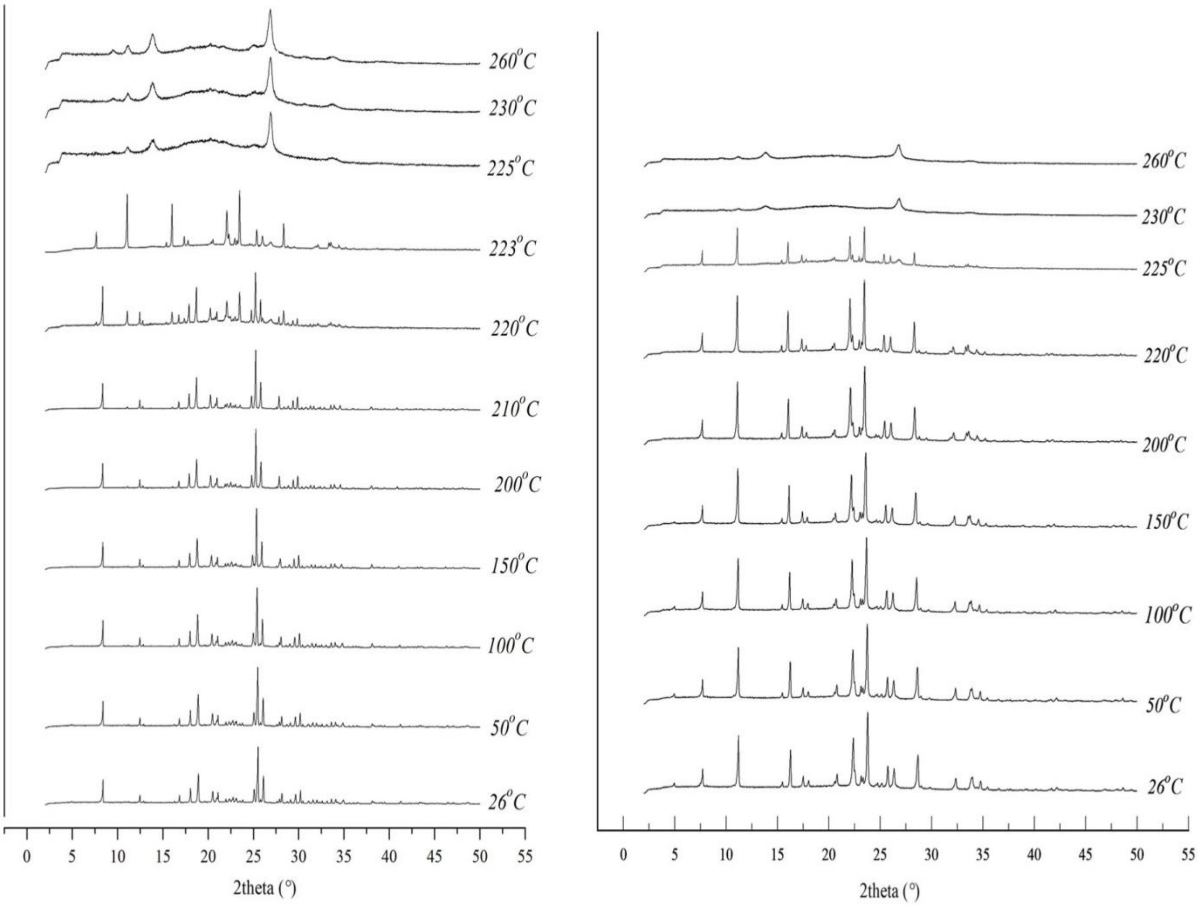


Figure 4.5: PXRD patterns for form I (left) and form II (right) obtained during vt-PXRD experiment carried out over a temperature range of 26°C to 280°C, at heating rate of 5°C/min, with more frequent collections over the range of 220°C to 230°C.

The major disagreement among the experimental evidence revolves around the stability of form II, because, as opposed to the DSC results and literature, the PXRD patterns display the disruption of its crystal lattice at a much earlier temperature.[197, 201] All repeated trials led to similar observations. This discrepancy might be due to the lack of controlled environment conditions during vt-PXRD experiments (matching vt-PXRD results in literature were obtained from air tight and thermally insulated set-ups) or the presence of an impurity which was not detected in the diffraction patterns.

4.3.1.3 Crystal structure analysis

The following study entails a detailed computational analysis aimed to characterise both anhydrates at different molecular levels, so as to identify the features which drive the exhibition of the above experimental results. The investigation of the thermodynamic stability of this pair of polymorphs using the CCDC toolkit was conducted in earlier work (see appendix), and hence these results will only be included in cases to complement observations obtained by new work produced here (parts of which was published[162]). The molecular flexibility defined in the previous chapter is very well expressed through the comparison of the molecular conformations of these forms, presented in figure 4.6. One can immediately notice how the orientation of the branch plays a crucial role towards the formation of each polymorph and its respective properties. The distinct extent of folding does not only have implications on the intramolecular stability of each molecule, but also dictates the accessibility with respect to intermolecular bonding.

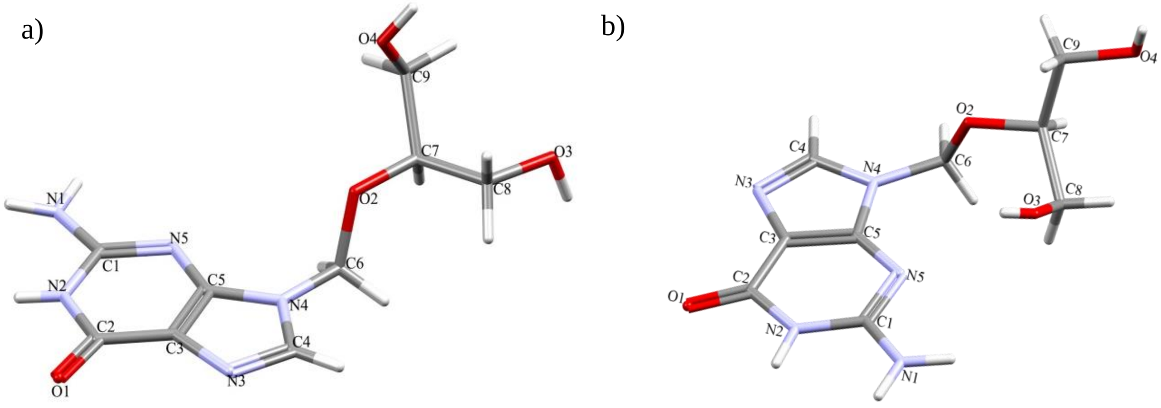


Figure 4.6: The molecular structure of GCV in the asymmetric unit of: a) anhydrous form I (CCDC refcode: UGIVAI01) and b) anhydrous form II (CCDC refcode: UGIVAI), with the numbering relevant for this chapter.

4.3.1.3.1 Hirshfeld surface analysis

Hirshfeld surface analysis was conducted to have a deeper insight of the variances in the network of HB and the involvement of weaker interactions within these conformational polymorphs.[84, 280-281] This tool was employed due to the nature of the Hirshfeld surface, which depends on the spherical atomic electron densities of a particular molecule within its crystal structure, hence being able to differentiate between polymorphs.[246] Surfaces in figure 4.7 exhibit the distinct traits of each form which primarily concern the position and intensity of red spots highlighting the presence of close contacts. The Hirshfeld surface of form I is populated with more red areas in comparison to that of form II, due to its extended conformation that amplifies its accessibility to hydrogen bonds. This observation was already noted through the full interaction maps, in figure 2.17, where the superior accessibility for HB in form I was exposed through larger, more intense areas.

The data mapped on the Hirshfeld surfaces was translated into a 2D format through the construction of fingerplots in figure 4.7, by plotting d_i against d_e, so as to illustrate the nature of all non-bonding interactions. The plots are overall significantly similar, probably due to the comparable properties between these conformational polymorphs. The H···H interactions are responsible for the most substantial contribution in both forms, with a higher percent (figure 4.8) and smaller minimum distance $d_i \approx d_e \approx 1.1$ Å. Even though such contacts are usually considered as repulsive in nature, Matta *et al.* portrayed how the net result of their presence is a stabilizing effect, with a maximum decrease of 42 kJ/mol in the total energy.[244]

The dominant appearance of N···H and O···H contacts represented by the spikes at the bottom left of both plots, is a product of the highly polar functional groups in GCV. The participation of these polar groups was already suggested by the electrostatic map of GCV, and *IsoStar* contour maps (see appendix). The spikes present in the fingerplot of form I are characterised by greener shades, thereby indicating a higher frequency of contacts with relatively shorter distances. The percentage contribution of O···H interactions is equivalent for both forms, but the global minimum distance $d_i + d_e$ of around 1.7 Å is located in form I.

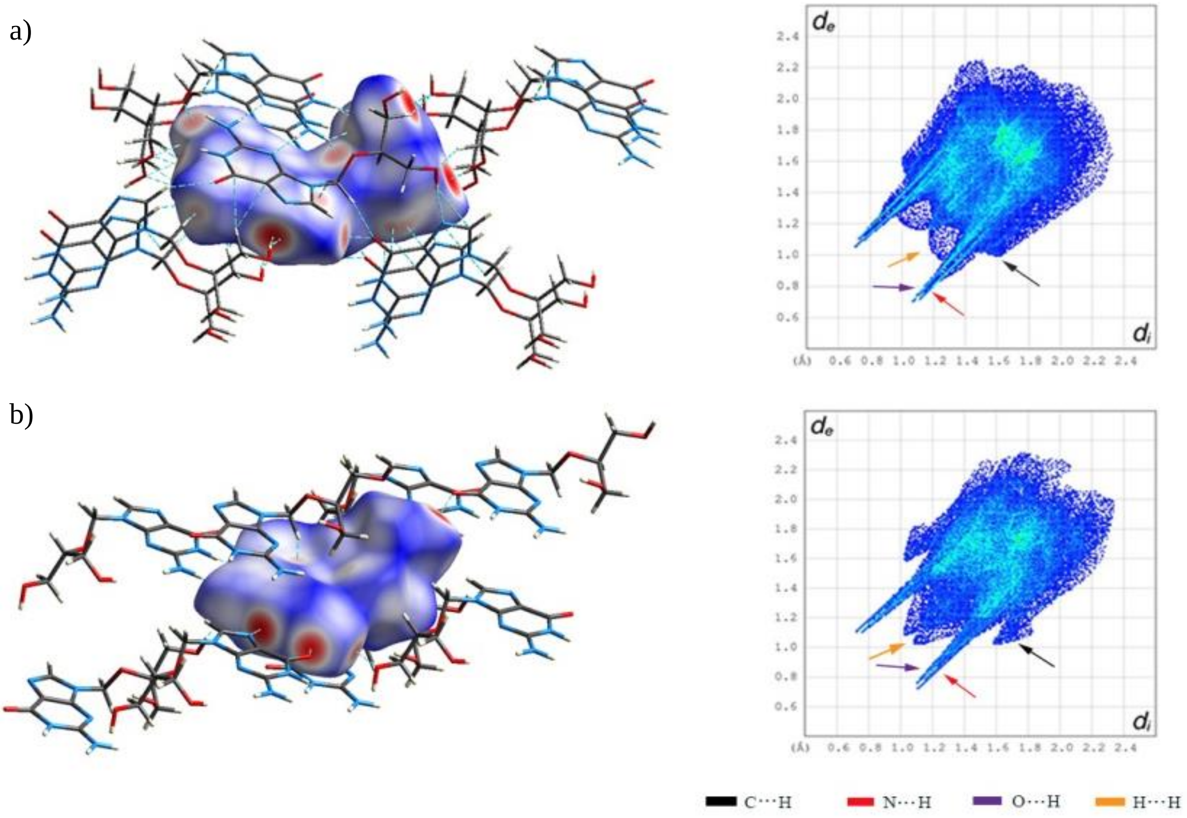


Figure 4.7: Hirshfeld surfaces with d_{norm} as mapped function, together their corresponding fingerplot, constructed for: a) form I and b) form II, with neighbouring molecules to illustrate the network of non-bonding contacts, represented by dotted lines (not only hydrogen bonds). The arrows represent interactions between different participants as indicated in the legend below. Given that the plots are approximately symmetrical, the arrows can be mirrored through the x-y diagonal.

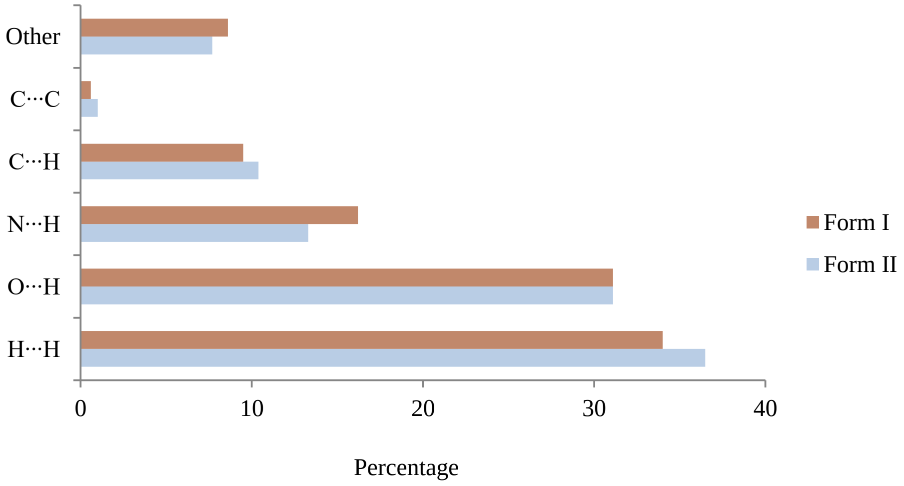


Figure 4.8: Bar-chart representing the percentage contribution of the main contacts in the Hirshfeld surfaces of forms I and II. The category of "other" includes C···N and C···O interactions.

Hirshfeld surface maps exhibit visually similar C-H···π contributions in relation to both polymorphs, however it is more pronounced in form II (see figure 4.8). Such difference in extent of C-H···π contribution was also indicated by full interactions maps illustrating methyl carbon and aromatic carbon as probes, which produced relevant maps only for form II (see appendix). Even though these interactions are less directional and weaker than the traditional HB, they still attribute in self-assembly and molecular recognition processes, as they strengthen the stability within the supramolecular structure.[282] Conversely, the absence of large peaks around the area at $d_i + d_e \approx 3.8$ Å and of the typical bold red and blue pairs of triangles on the shape index surfaces, indicate the less relevant contribution of π···π contacts.[283] Analysis of the network of interactions of the two forms led to the observation of how the side chain does not allow compact stacking of the aromatic systems, thereby reducing the influence of π···π contacts. C···C interactions are characterised by relevantly long distances and located mainly on the perimeter of the guanine ring system, hence suggesting that the aromatic systems are either distant and/or shifted relative to one another.

4.3.1.3.2 Thermodynamic stability – Energy frameworks

The network of non-covalent bonding described above constructs the packing arrangement of each polymorph as depicted in figure 4.9. The extensive network of HB in form I is highly visible along its packing pattern, which seems to follow zigzag lines when viewed along the *c*-axis. This type of pattern allows the layers of molecules to be very close to each other, hence enabling the optimal use of the space available, with a packing coefficient of 0.770978.[284] The complexity of this network is enhanced by the participation of carbon atoms as HB donors. The packing pattern of form II is composed of parallel guanine back-bones, with the “out of plane” branches. The folding in form II enables the presence of an intramolecular bond, but leads to a less populated network of intermolecular HB. The visible layering of the guanine rings creates an off-centre parallel stacking, but the geometric parameters of the guanine ring centroids stand at the upper bounds for effective π···π interactions, as suggested by the literature and by the fingerplots.[285-286]

Apart from the packing arrangement, the nature and strength of non-covalent interactions also dictate the thermodynamic profile of polymorphs. The calculation of energy frameworks was applied in order to comprehend the topological dissimilarities of the energy components of the forms and potentially derive a correlation to their respective packing and thermal behaviour. These frameworks are visualised in figure 4.10, depicting both Coulomb forces / electrostatic potential force and dispersion energy contributors.

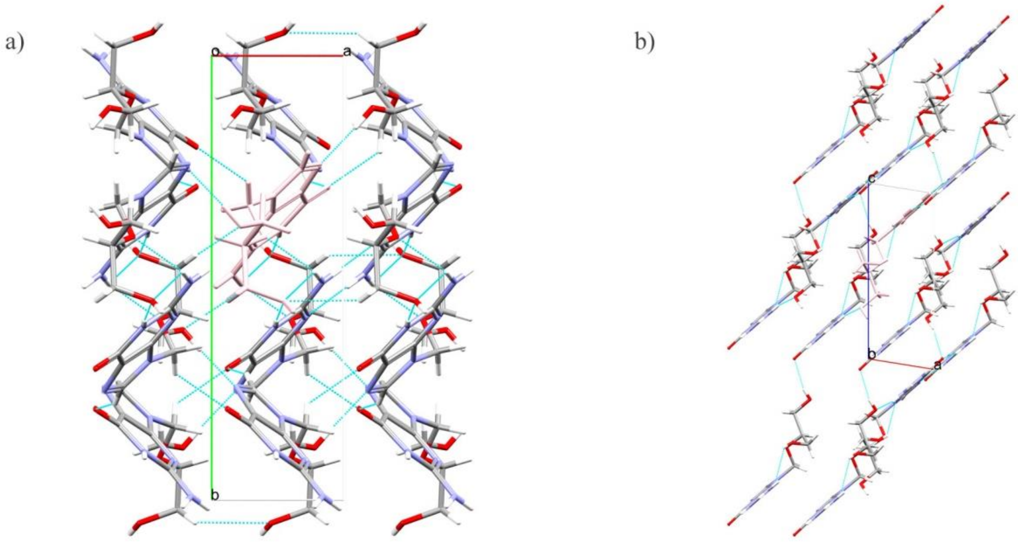


Figure 4.9: The packing diagrams of a) form I and b) form II, with a central molecule having the same back-bone orientation. Hanging contacts were removed for better visualisation.

Form I is centrosymmetric, crystallising in space group *P*2_1/*c*. Its unit cell contains two pairs of molecules that are identical within the group, and inversely-related between groups. The supramolecular structure of form I is stabilised with energy components directed along both parallel and non-parallel molecules, in a multidirectional fashion. The dispersion forces mainly can be traced along the off-centred parallel stacking (along the a-axis), as well as between adjacent molecules whose aromatic systems lie along the same plane. N1-H···O2 and N2-H···O4 interactions compose the strongest energy contribution (-164 kJ/mol), which is situated between parallel molecules. The distinctive involvement of the oxygen ether, O2 as a HB acceptor is also portrayed through a shift in its IR peak towards a lower frequency (1171 cm^{-1}) relative to its respective peaks in other forms (1183 cm^{-1}). The strength of such stabilizing contacts is a result of both electrostatic and dispersion forces, with the former being the dominant one. The presence of O4-H···O3 contact coupled with O3-H···N3 and C6-H···O1 interactions induce a further cumulative stabilizing effect of -120.1 kJ/mol. These observations represent the strong correlation between packing arrangements and energy frameworks (figure 4.9 and 4.10), both of which are a direct implication of intermolecular bonding.

Form II crystallises in the non-centrosymmetric space group *P*2_1, thus lacking an inversion centre. The most dominant electrostatic energy component in form II (-73.6 kJ/mol) is attributed to N1-H···O1 and N2-H···N3 interactions, located along the planes of aromatic rings. A secondary energy contribution (-45.2 kJ/mol) towards the same direction is present due to N1-H···O4 contact. The O4-H···O1 electrostatic interaction combined with dispersion contacts constitute the out-of-plane interactions, whose contribution is relatively smaller (-18.3 kJ/mol) compared to those along the aromatic systems plane.

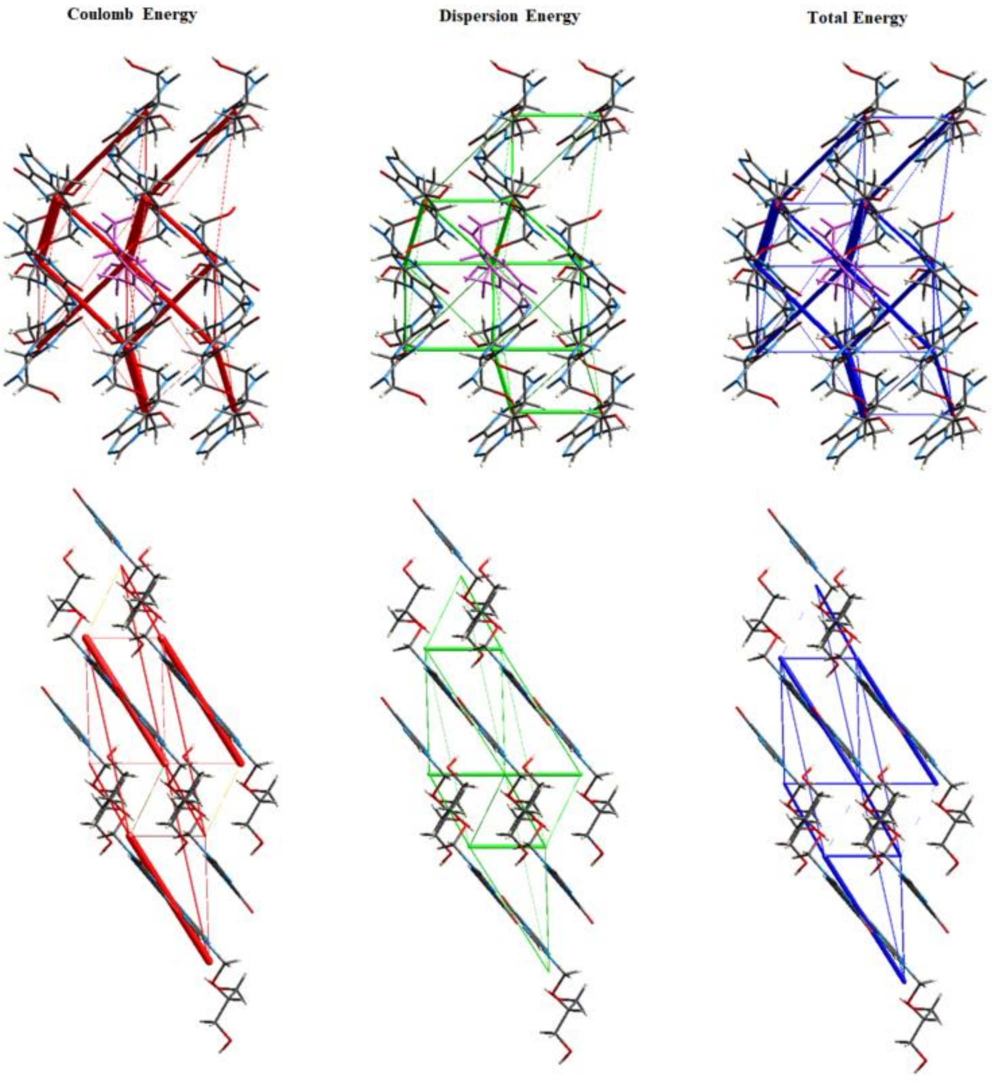


Figure 4.10: Energy framework diagrams, using CE-B3LYP model, using a scale tube size of 20 and cut-off value of 0 kJ/mol. Top row belongs to form I (viewed down c-axis) and bottom row is associated with form II (viewed down b-axis). Cluster of molecules with 3.8 Å radius around the central molecule (pink).

The less exhaustive network of intermolecular bonding of form II, together with the presence of an intramolecular bond, seems to prioritise the stability within the molecule, as opposed to form I which entails a very complex extensive framework of intermolecular contacts, with 16 HBs per molecule (with eight different neighbouring molecules). The comparatively thicker cylindrical radius of the total energy tubes along several planes in form I prove the greater degree of stabilising effect caused by its multidirectional framework of non-bonding intermolecular contacts. Such influence in conjunction to a more compact packing arrangement, might contribute towards a lower enthalpy in form I relative to form II, which according to L. Yu, has to be accompanied by a smaller level of entropy in order to conserve the enantiotropic relationship between the two polymorphs.[46, 201] With an increase in temperature, in the range of 222°C to 228°C, enough energy is absorbed to break the intermolecular framework of form I, which then transforms to form II.[197, 200-201] The thermodynamic stability of form I is also evidenced through the fact that form III (hydrate form) converts to form I at the temperature above 180°C; an exothermal transition which proves the monotropic relationship between the two forms.[201]

The intermolecular framework of form II seems to be equipped to cater for higher energy environments, with relatively loose packing (0.725452). The energy framework of this polymorph illustrates how the stabilizing effect is spread along planes, whereas much less robust contacts are observed in between layers (see figure 4.10). This lower level of intermolecular forces between parallel planes in form II might create an environment which can accommodate higher entropy within the crystal structure, hence allowing it to be thermodynamically stable at higher temperatures. In addition, this feature is also exhibited through the difficulty of obtaining form II through solubility screening, both in literature and in this work (only one trial).[200-201]

4.3.1.3.3 Thermodynamic stability – Lattice energy

The lattice energy of GCV polymorphs was calculated based on a static model of each crystal structure at 0 K, following the general equation *2.10.* As explained in the literature review, despite being of small magnitude, the results obtained do not cater for any thermal contributions, thereby introducing a degree of error. These forms exemplify cases where the calculated relative stabilities are sensitive to the method used for computation, due to the subtle interplay between the intramolecular conformational energy and intermolecular interactions present.[102] For the first method, U_{inter} was computed using the CE-B3LYP model, which decomposes the interaction energies between a central molecule and its neighbours into four components; electrostatic, dispersion, polarisation and repulsion, as displayed in eq. *3.5*.[274] The cluster of molecules was increased until the cumulative lattice energy reached convergence, with the electrostatic component being the last to converge due to its long range nature. The progression of this computation is illustrated in figure 4.11, with form I having a more negative cumulative lattice energy, probably as a result of its extensive HB network. Such outcome was already hinted through the components of the energy framework, and the density of the crystal structure given in table 4.1.

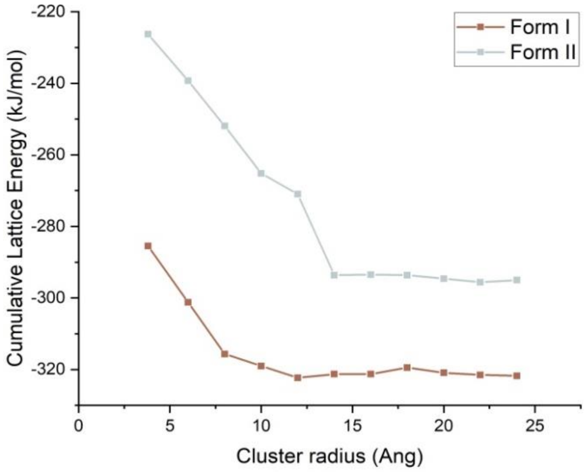


Figure 4.11: Graph showing the cumulative lattice energy (kJ/mol) at each cluster increment, until convergence was reached, using the CE-B3LYP model.

Despite being fairly accurate for several applications, the presence of conformational polymorphs required the addition of further refinements onto the cumulative lattice energy, as specified in *eq. 4.2*. The calculation of the ΔE_{intra} was performed as suggested by Thomas *et al.* whereby the molecular structure was geometrically optimised with fixed torsion angles at experimental values, followed by a single point calculation at MP2/cc-PVZ level of theory.[84] The ΔE_{intra} was evaluated as the difference in energy between such molecule and its associated low energy conformer in vacuum. The molecular structure within the solid-state was partially allowed to relax because due to possible errors in the experimental values, especially those corresponding to the hydrogen atom positions, that can lead to excessively highly single point energies.[117] Nonetheless, it was crucial that the general features of the conformation within the crystal structure were still preserved, so as to ensure any stabilising or strain effect present can be accounted for. Such comprise was more complex due to the considerable number of polar functional groups in a relatively small area of GCV. These features coupled with a relatively high flexibility, increased the propensity of forming intramolecular bonds during geometry optimisation either as O3H⋯ N5 or even between the hydroxyl groups, thereby amplifying the importance of the correct estimation of ΔE_{intra}.

The second method relied on the evaluation of the crystal lattice using a static, infinite crystalline solid, approximated through the use of periodic boundary conditions, and its related gas-phase conformer, as per *eq.4.4*. One should note that Stein *et al.* recommended the use of B97-D due to its lower mean absolute error when compared to PBE-D3, however the latter was utilised because B97-D was not accessible in the *Turbomole* version available.

Table 4.1: Lattice energy values for GCV form I and II. (Please see appendix for more details)

	Description of calculation	
Form I	Cumulative lattice energy, $\boldsymbol{E_{tot}^{CE-B3LYP}}$ (kJ/mol)	-321.750
	Cell dipole Energy correction, $E_{cell\ dipole}$ (kJ/mol)	Not required
	Intramolecular energy penalty, ΔE_{intra} (kJ/mol)	48.125
	Lattice Energy (kJ/mol), E_{lat} (*CrystalExplorer* model, *eq.4.2*)	**-273.625**
	Lattice Energy (kJ/mol), E_{lat} (PBC, *eq.4.4*)	**-278.567**
	Denisty (g/cm^3) (Hirshfeld surface analysis)	1.653
Form II	Cumulative lattice energy, $\boldsymbol{E_{tot}^{CE-B3LYP}}$ (kJ/mol)	-295
	Magnitude of cell dipole moment in unit cell (D)	7.705
	Cell dipole Energy correction, $E_{cell\ dipole}$ (kJ/mol)	-0.114
	Intramolecular energy penalty, ΔE_{intra} (kJ/mol)	19.823
	Lattice Energy (kJ/mol), E_{lat} (Crystal Explorer model, *eq.4.2*)	**-275.291**
	Lattice Energy (kJ/mol), E_{lat} (PBC, *eq.4.4*)	**-272.094**
	Denisty (g/cm^3) (Hirshfeld surface analysis)	1.549

Given the cumulative lattice energy values and the densities, it was expected for form II to have a higher lattice energy, in accordance with the values obtained for method two.

The lowest energy conformer of GCV in the gas-phase resulted to have a strong directional intramolecular HB, similar to the one present in form II. This observation was expected because in an environment with no intermolecular constraints, the molecule is likely to endorse a conformation which provides the greatest extent of intramolecular stability. A study on the possible molecular strain induced by intermolecular interactions within the crystal lattice with molecules having **no** intramolecular bond, concluded that in some cases the resultant distortion may still add up to 20 kJ/mol.[90] This implies that penalties associated with molecules involving possible intramolecular HB can be expected to be of a larger magnitude, due to a higher distortion (48.13 kJ/mol for form I).This distortion often leads to the formation of an extended molecular conformation in the crystal structure, with higher energy than the isolated molecular conformation, as observed in form I.[90] The effect of such intramolecular strain is compensated by the larger surface area of the extended molecular conformation of form I, which enhances its potential to engage in stabilizing intermolecular interactions within the crystal, as observed through its network of HB and energy frameworks.[78] Despite having such large penalty, the lattice energy values of both methods are in a considerable agreement (see table 4.1)

4.3.1.4 Risk assessment of polymorphism

Apart from determining solubility (see appendix), the screening was intended to explore routes for potential polymorphs. Due to the low solubility of GCV, only a few of manual solubility tests were successful, most of which required heat. Due to such tendency, it was beneficial to conduct the screening using crystallisers so as to be able to control the conditions better, particularly temperature. This technique was also used because manual solubility screening using a variety of mixtures was already performed in literature, with no novel polymorphs formed.[200-201] Furthermore, given that main objective revolves around the polymorphism of anhydrates, parallel to the following computational work, solvent mixtures very often required the presence of water, which would ultimately lead to hydrates.

However, same outcome is obtained here through the screening, with no new forms despite using heat. Even though new forms might have formed during cycles, none of them was stable enough to re-crystallise at room temperature by slow evaporation. The exploratory experimental run conducted using vt-PXRD with a cycle of heating (to 223°C), cooling (to -50°C) and heating (to 280°C), at a rate of 20°C/min did not lead to the formation of new crystal forms. This outcome implies that the likelihood of the formation of other anhydrous GCV polymorphs is rather small, re-confirming the thermodynamic stability of form I at room temperature.

4.3.1.4.1 HB propensity model

The objective of the construction of a logit regression model in *Mercury* concerns the stability assessment of the network of HB present within both forms, given the complexity of the chemical environment of GCV. The molecular structure was represented by substructures, ensuring that all available HB participants are included. The list of HB contributors also contained carbon as a HB donor, given its crucial role within the HB framework. The work regarding this procedure for GCV was conducted earlier, and all details and results can be found in literature.[162] The predictive ability of the model was reflected by the reduction from the null to residual deviance, together with a value of 0.878 for the area under the ROC curve. The procedure yielded the same model when starting the analysis with different forms, apart from minor differences due to the non-deterministic nature of the method.[133]

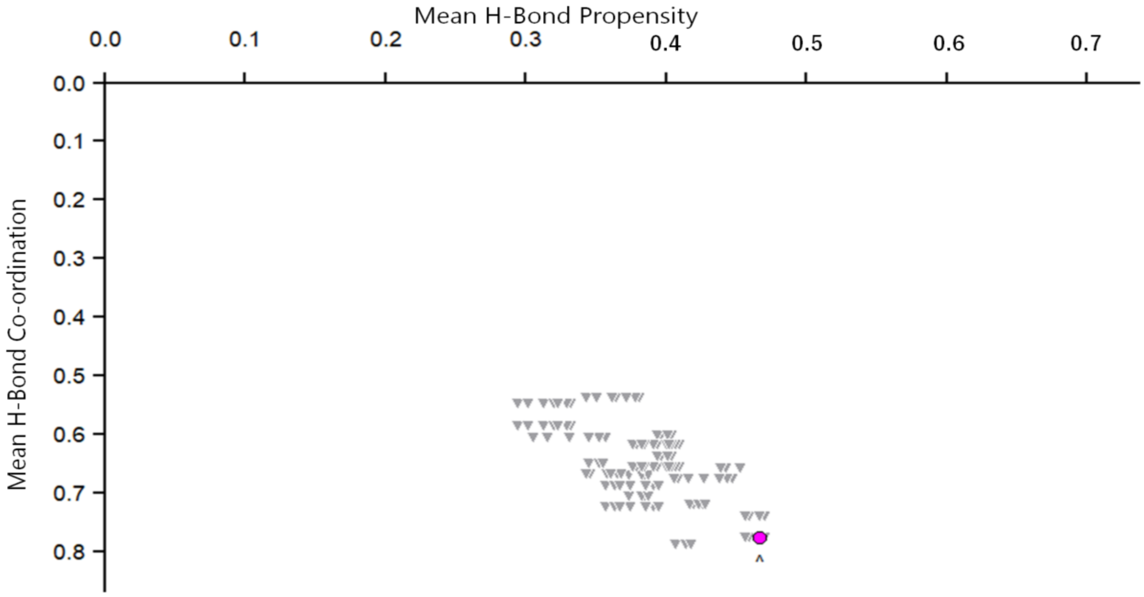


Figure 4.12: The propensity chart for GCV, showing form II (violet), together with other hypothetical forms.

The model was employed to calculate the propensity of all possible intermolecular HB, and each potential polymorph with a definite set of predicted HB was plotted in a putative polymorph landscape, as shown in figure 4.12. Given the high frequencies of the interactions found in form II relative to the structures in the CSD, this was classified amongst the forms with highest likelihood of formation. On the contrary, due to the nature of the contacts present, particularly those involving a CH donor, form I was plotted in the non-optimal area of the landscape, implying a lower extent of stability and hence less likelihood of formation. This outcome is in disagreement with the experimental evidence as well as with the computational work done above. The fact that this approach cannot take into account directional features of interactions, their collective influence towards stability and the effect of other non-bonding contacts which span over a longer range, poses a great limitation with respect to its polymorph stability classification.

4.3.1.4.2 Crystal structure prediction

The conformational search based on the geometric parameters defined in the previous chapter was able to locate six unique conformers, all of which were used as the starting molecule. Multiple runs of polymorph generation were performed using different parameters until it was ensured that they converged towards the same low energy forms, while closely monitoring the behaviour of energy graphs. During this process, generated forms were scanned to locate the experimental forms, using crystal similarity tool, superimposition of XRD patterns through *Reflex* module and HB network comparison.

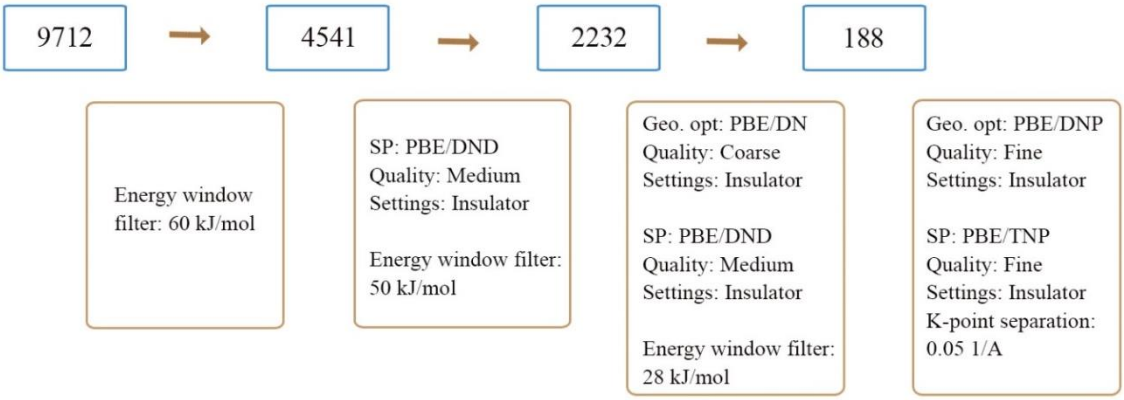


Figure 4.13: Progression of the DFT calculations and energy window filters employed to obtain the final 188 potential forms, with the number of structures (blue box) and settings (brown box).

The final clustering resulted in a list of 9712 potential structures, which was reduced even further using an energy window filter of 60 kJ/mol (on energy values obtained using COMPASS forcefield). The resultant 4541 structures were processed through a series of DFT calculations at 0K, illustrated in figure 4.13, each time refining criteria and reducing the energy window. The recommended ability of DFT calculations for the evaluation of lattice energy was witnessed through various studies discussed in chapter two. The final energy window filter was chosen based on an upper bound criteria suggested in literature.[215]

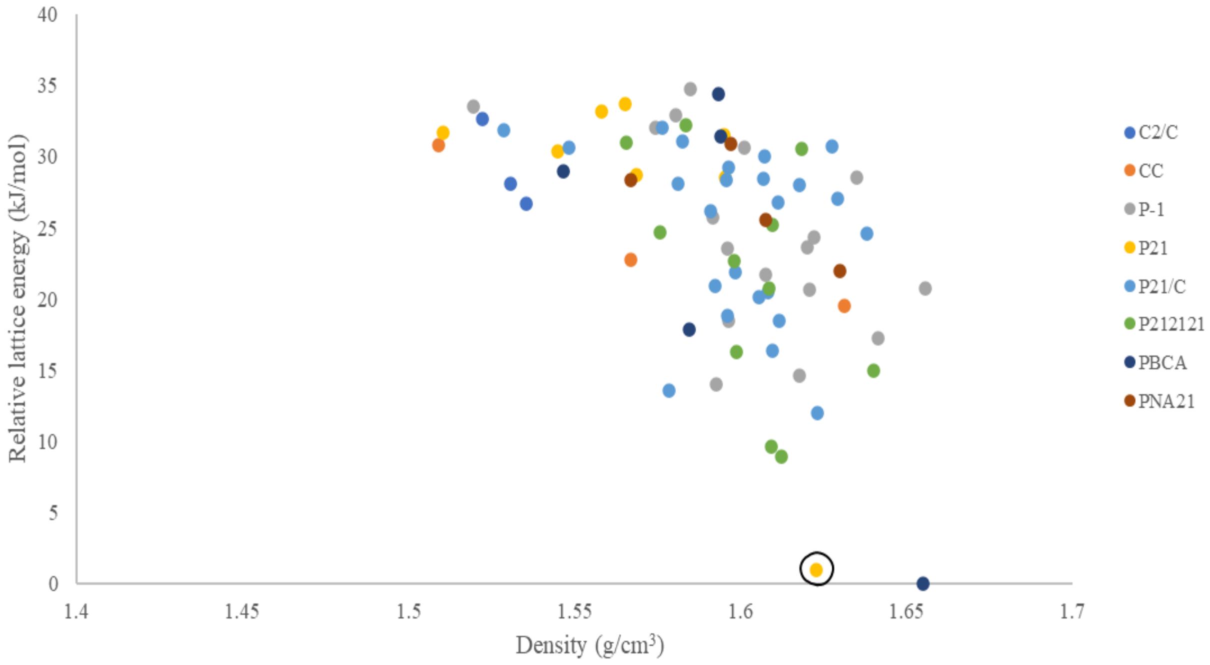


Figure 4.14: Crystal energy landscape of GCV, where each point represents a potential crystal structure and labelled according to its respective space group. The circled datapoint represents the generated crystal structure identical to form II.

The crystal energy landscape of GCV for anhydrates is displayed in figure 4.14, plotting relative lattice energy (kJ/mol) and density (g/cm^3). The most evident feature is the energy gap of around 9 kJ/mol separating the two lowest energy crystal structures from the rest, with one of them being form II (crystal superimposition = 0.9974). The agreement with the experimental unit cell details of form II is demonstrated in table 4.2. The global minimum involved a crystal structure with *Pbca* space group, composed of a molecular conformation similar to that present in form II, including an intramolecular bond at the same location (see appendix). The main difference resided in the orientation of other hydroxyl group which had a HB with the ether group of the neighbouring molecule. The discrepancy in energy of less than 1 kJ/mol might require finer energy calculations to be verified, particularly those which take into account thermal contributions. Nonetheless, the presence of this new hypothetical form in such low energy region might suggest inexhaustive solubility screening.

Table 4.2: Unit cell parameters of experimental and generated crystal structures.

	Form I		**Form II**	
	Experimental	Generated	Experimental	Generated
Space group	$P2_1/c$	$P2_1/c$	$P2_1$	$P2_1$
a	4.6448(10)	4.4867	4.3803(15)	4.2289
b	15.632(3)	16.3058	10.909(4)	10.8091
c	14.130(3)	14.2404	11.601(4)	11.6683
α	90	90	90	90
β	91.632(10)	90.8603	99.11(3)	101.6719
γ	90	90	90	90
Density	1.653	1.627	1.549	1.623
Crystal superimposition (target)	0.9661		0.9974	

The closest generated crystal structure to form I (crystal superimposition = 0.9661) was characterised by a similar HB network (figure 4.15a), except for the region concerning the hydroxyl groups. This potential form featured an intramolecular bond with a high degree of directionality, thus markedly affecting the lattice energy. This factor made it impossible to represent and rank fairly form I. In attempt to generate a closer crystal structure to form I, multiple settings concerning the polymorph generation procedure were altered. These changes included an increase in the maximum temperature possible during the Monte Carlo simulated annealing and an expansion in the cut-off distance of the VdW terms during geometry optimisation. However, such alterations did not result in the generation of form I.

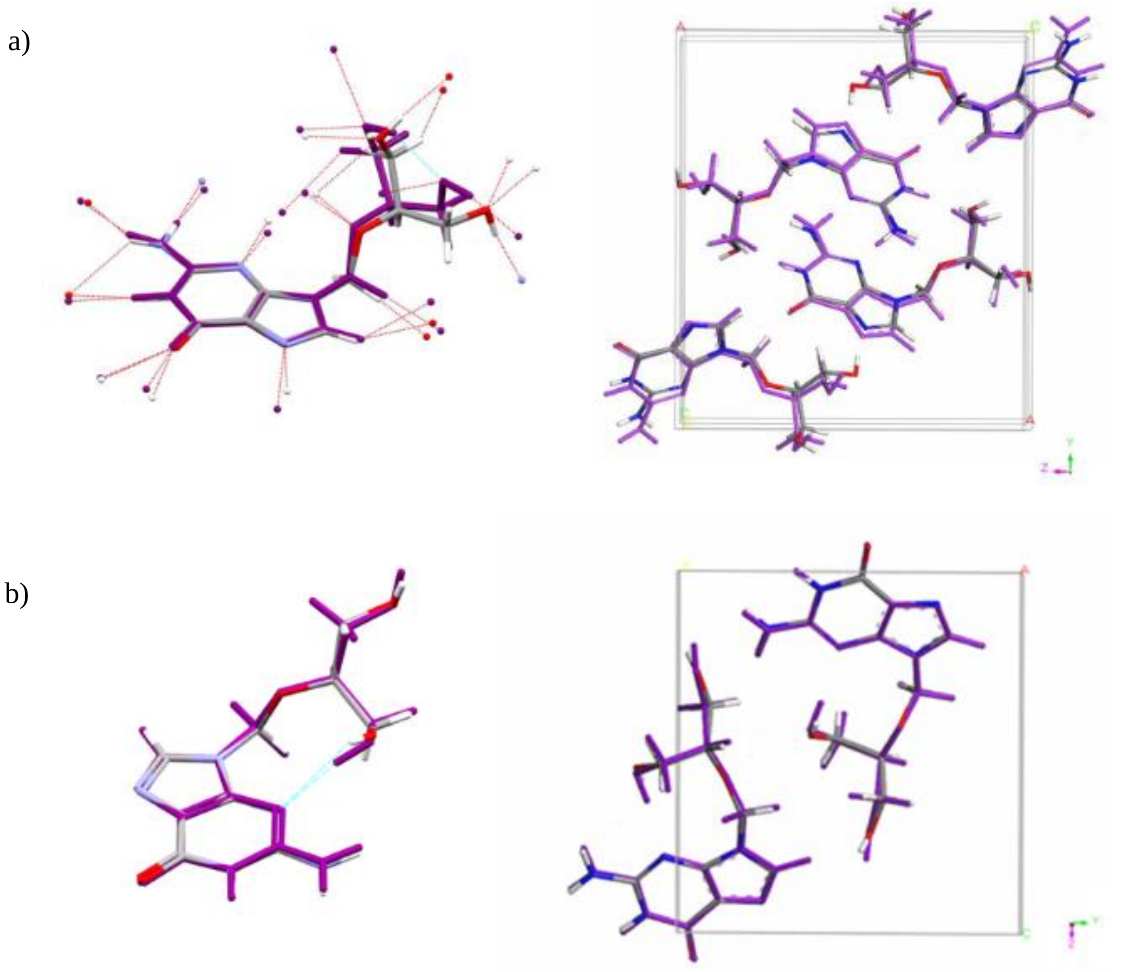


Figure 4.15: The molecular (left) and unit cell (right) superimposition of form I (a) and II (b) and their respective closely related generated crystal structure (violet). The intramolecular bond in the generated structure is coloured in turquoise.

The energy gap present in the crystal energy landscape promotes the suggestion that potential forms other than the global minimum, have a relatively small likelihood of being viable experimentally. This implies that having a considerable degree of flexibility, does not necessarily translates into a large amount of polymorphs. Such factor was also implied through the solubility screening results. Furthermore, given that form II was classified as the closest to the global minimum, also provides further reassurance for the thermodynamic stability of form I at low temperatures, which was proven both experimentally and computationally. However, this claim would have had more validity if the exact crystal structure of form I was generated and ranked as the global minimum.

4.3.2 Famciclovir

4.3.2.1 Form characterisation

FCV has five forms; three anhydrates (form I, II, IV), methanol solvate (form III) and a monohydrate (form V), with form I being present in the commercially available sample and patented as administered form (Famvir®).[203, 205-206, 287] Nonetheless, only form I and V have determined crystal structures. Despite not having a balanced number of HB donors and acceptors, FCV still exhibits three anhydrous polymorphs, characterised with different degrees of stability. This fact coupled with the increase in flexibility provided mostly by the carbon chain (flexibility scores: 0.65 to 0.71), pose the need of investigating the extent of the risk of polymorphism.

The polymorphs obtained from solubility screening, pure or otherwise, were characterised using PXRD data and FTIR, where possible. Only samples of forms I and V were obtained with the minimal presence of other forms (figure 4.16). One should note that the structure of form I was determined based on single-crystal data collected at 150 K, and such diffraction pattern differs slightly, particularly with respect to reflection positions at higher angles, from the PXRD data collected at room temperature (see app.). However, it was concluded in literature, that due to the similar trends in the relative reflection positions, the structurally determined structure is still a fair representation of the dominant form at room temperature, obtained from a commercial source.[206]

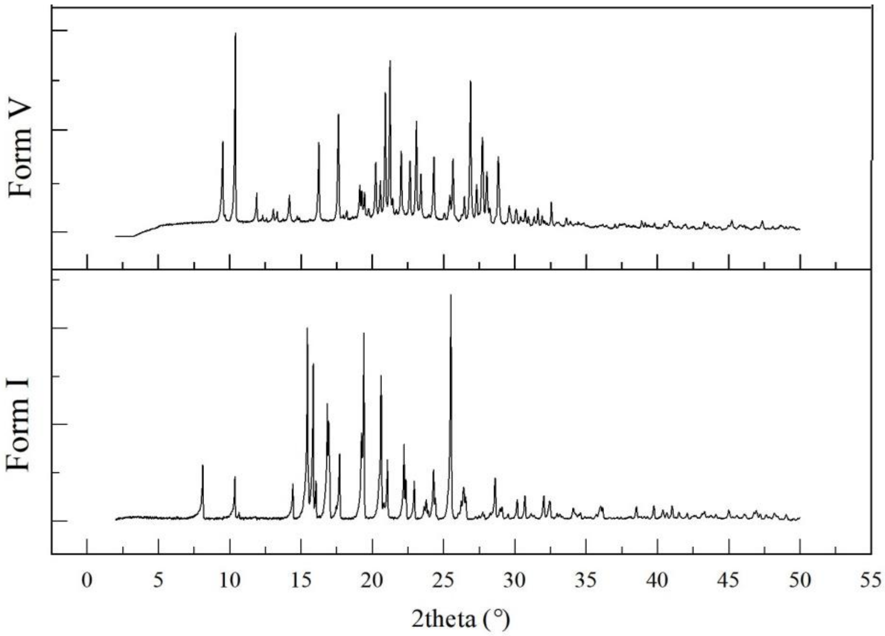


Figure 4.16: PXRD patterns for forms I and V, collected at room temperature.

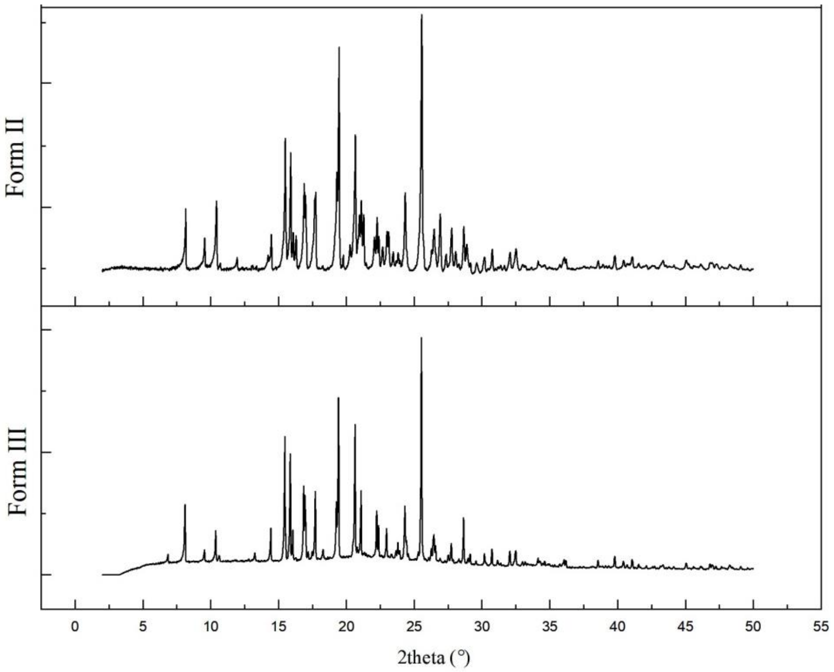


Figure 4.17: PXRD data of mixtures including form I and V, with traces of form II (top) and III (bottom). The presence of the monohydrate form is less dominant in the methanol solvate mixture.

The PXRD patterns in figure 4.17 illustrate two samples of mixtures, with different proportions of form I and V, but including also traces of form II (present characteristic peaks: 16.2°,16.4°,19.7±0.2°) and form III (present characteristic peaks: 6.6°, 13.0°, 18.4±0.2°). Re-crystallisation from alcohols, particularly ethanol and n-butanol should produce a sample with a dominant presence of form II.[205] However, maybe due to exposure of relatively high ambient humidity during the slow evaporation of the solvent, a mixture of form I and hydrate was obtained, with possible traces of form II.[205] The formation of the monohydrate might have also been instilled by the presence of some water content in the alcohol. On the other hand, the formation of polymorph I was not unexpected because there are various methods in literature whose end-product of re-crystallisation was composed of a mixture of both forms I and II.[205] Despite several trials, form II was not produced as a pure form and hence it was not possible characterise it further. Its presence in a mixture was not even possible to be detected by STA because its profile cannot be differentiated from that of form I.[205]

The FTIR spectra of form I (blue) and a mixture of hydrate and form I (black), re-crystallised from n-butanol, are displayed in figure 4.18. Given the presence of an identical component in samples, the spectra have similar trends, but with broader bands in the mixture sample, thus suggesting a variation in the HB network between both forms.

The stretching mode of OH group involved in HB in the hydrate form is displayed by the broad absorption band around the 3500 – 3200cm^{-1} region. The absorption band for NH stretching at 3334 cm^{-1} is present in both spectra, but it is accompanied by another overlapping band at 3358 cm^{-1} in the mixture spectrum. Such difference might either be a result of OH stretching or a difference in the NH stretching mode. The C=O stretching bands are situated at higher wavenumbers (1745 –1730 cm^{-1}) compared to those associated with carbonyl groups in a conjugated system. Aliphatic CH vibration modes are fairly visible through bands occurring at 3000 – 2840 cm^{-1} and 1460 – 1375 cm^{-1} regions, particularly due to the methyl groups.

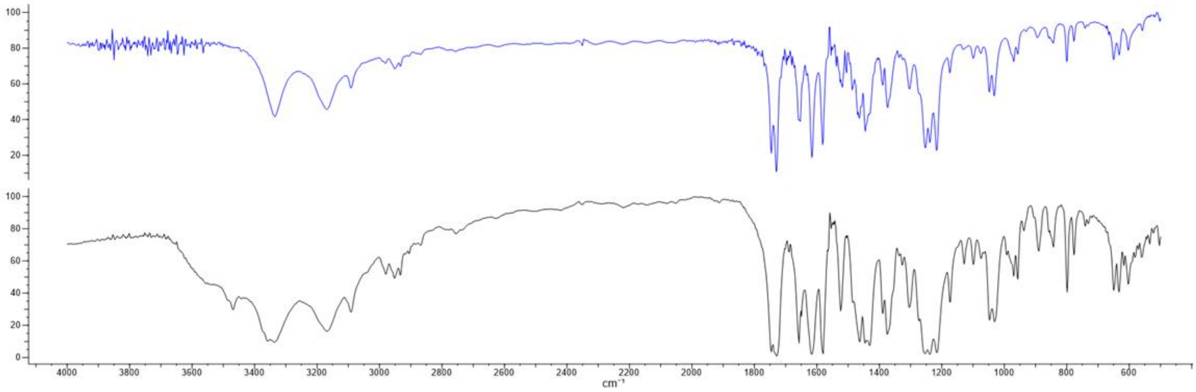


Figure 4.18: FTIR spectrum of pure form I (blue) and mixture of hydrate and form I (black).

4.3.2.2 Thermal Stability – Experimental characterisation

The wide range of different methods available to transform one polymorph to another portrays the difficulty of understanding the thermodynamic relationships between FCV forms.[205] Solubility screening results show the dominance of the presence of form I, with only traces of other forms in mixtures. This result concords with the calculations done by Wu *et al.*, which proved that FCV form I has a satisfactory level of thermal stability at normal climatic conditions, even if stored for a long time.[288] Furthermore, according to literature, FCV (form I) is non-hygroscopic at room temperature, with relative humidity below 85%.[287] However, better controlled experiments are required to determine the exact parameters of such property. It is noteworthy to add that form I can be produced by heating form II at 40°C for two hours under vacuum, thus suggesting a polymorph transition to a more stable form.

The observation of the crystal habit of form I through the microscope exposed its monoclinic nature. Furthermore, the heating of a finely grinded sample was monitored, as shown through figure 4.19. No change was observed prior to the initial signs of melting at 104°C, which proceeded further till 108°C. The image at 107°C captures an intermediate stage of melting, where clusters were composed of a liquid phase on the periphery and a group of crystallites at the centre, which still required more heating to melt. A detailed thermal analysis for FCV performed in both air and nitrogen atmosphere, displayed how decomposition starts at 205°C, and it occurs through two two-step processes.[288]

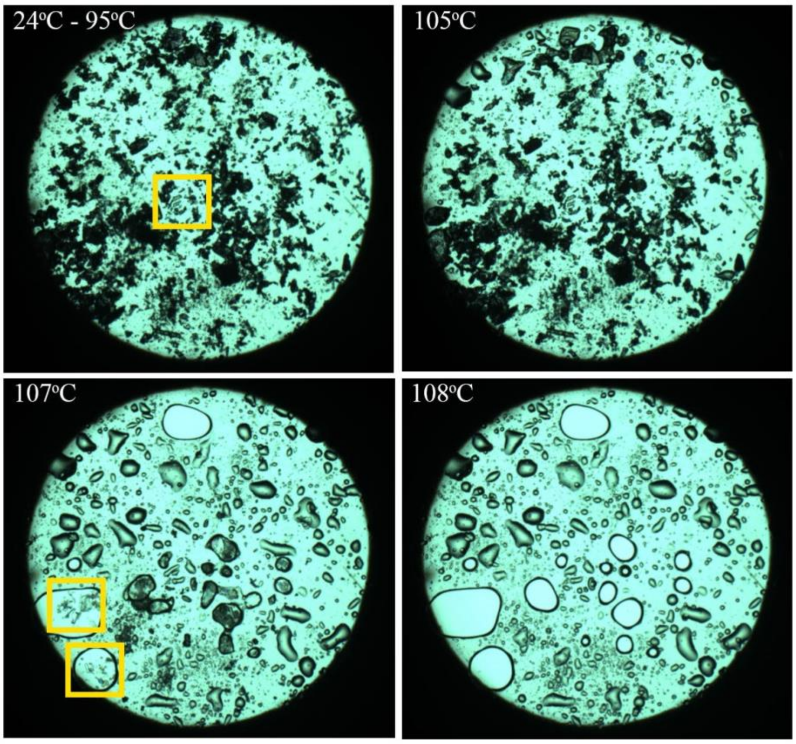


Figure 4.19: Hot-stage microscopy images (x2.5 magnification) starting with form I, set up with a temperature profile: heating from room temperature till 90°C at 10°C/min and decreasing the rate to 5°C/min till 110°C.

The melting without decomposition (no mass loss) over the temperature range of 102-107°C is shown by the second endotherm of the DSC graph of form I, displayed in figure 4.20a. This process was also monitored using vt-PXRD, which also verified such melting range, with the last crystals being detected at 104°C (figure 4.20b). The first small endotherm relates to the evaporation of water occurring at 60-65°C, and confirmed by MS spectrum. The removal of water indicates the presence of small content of a hydrate within the form I sample. During heating, it was noticed that small diffraction peaks, characteristic of the monohydrate, were no longer visible above the temperature of

65°C. Both STA and vt-PXRD results of form V show the removal of hydration at 60-65°C, accompanied by a mass loss of 3.65% and a phase transition to form I (see appendix). Same monohydrate characteristic peaks contributed towards the difference between the simulated PXRD pattern of the determined crystal structure of form I (refcode:IMADEI) and the collected PXRD pattern of the commercially available sample.

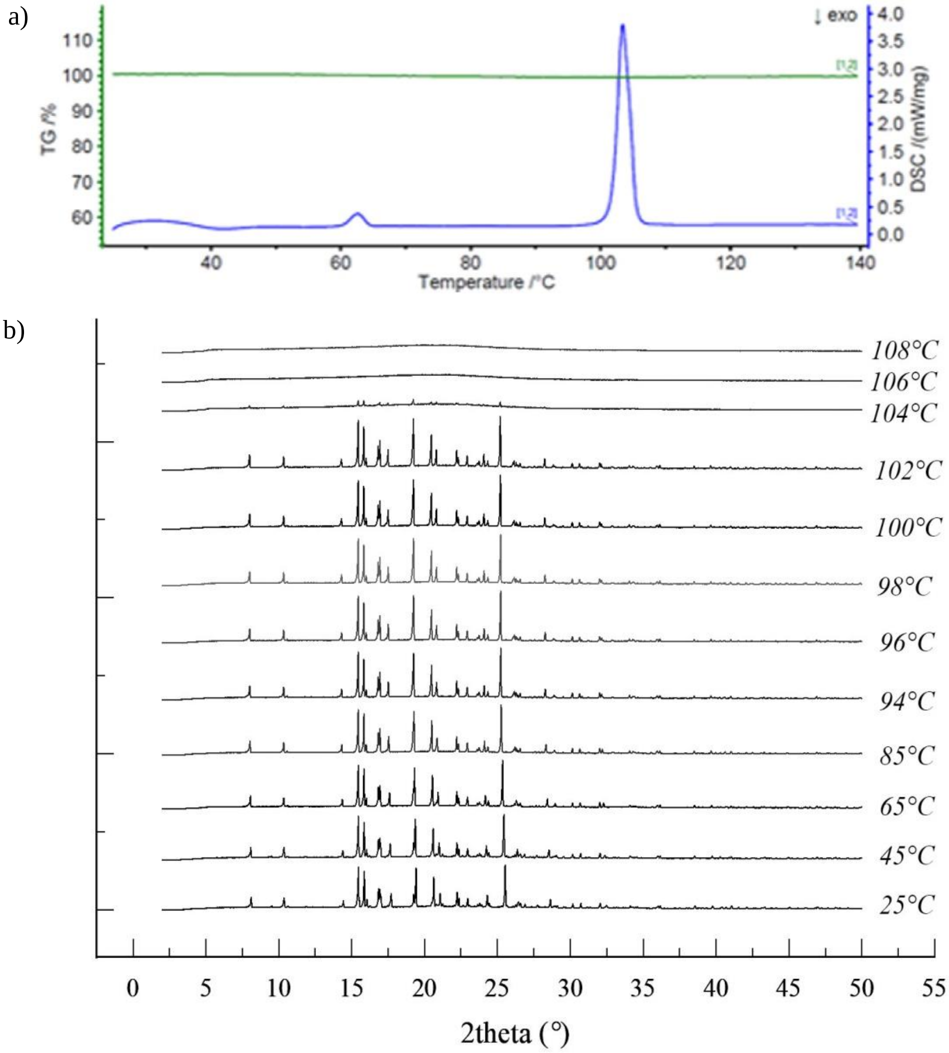


Figure 4.20: a) DSC/TG graph overlay for form I and b) PXRD patterns for form I obtained during vt-PXRD experiment carried out over a temperature range of 25°C to 140°C, at heating rate of 5°C/min, with more frequent collections over 94°C.

The exploratory vt-PXRD run of successive heating and cooling cycles was accounted for in literature, to produce form IV.[206] The same experiment was conducted to investigate the thermodynamic stability of form I within such parameters, while also using STA. The first heating cycle proceeded as described above, with a complete liquid phase at the maximum temperature of 110°C (figure 4.21), following the melting range defined earlier (as figure 4.20). The lack of crystalline material was sustained during the cooling till -50°C and heating again till 35°C, at which point the appearance of the first diffraction peaks was noted. Both diffraction patterns at 35°C and 45°C show the characteristic peaks of form I, together with additional minor peaks at around 5.8-5.9° 2θ, indicating the formation of traces of form IV.[206] These intensities were absent in the diffracted pattern collected at 55°C, suggesting a polymorph transition to form I, which later melted over its distinctive temperature range. This phase transition was also detected by the DSC/TG analysis (figure 4.22), through an exothermic peak at 68°C, with an onset temperature of 58°C and no mass loss. Further runs are required to understand better the cause behind the difference between the phase transition temperatures observed by the DSC and vt-PXRD, apart from potential instrumental differences.

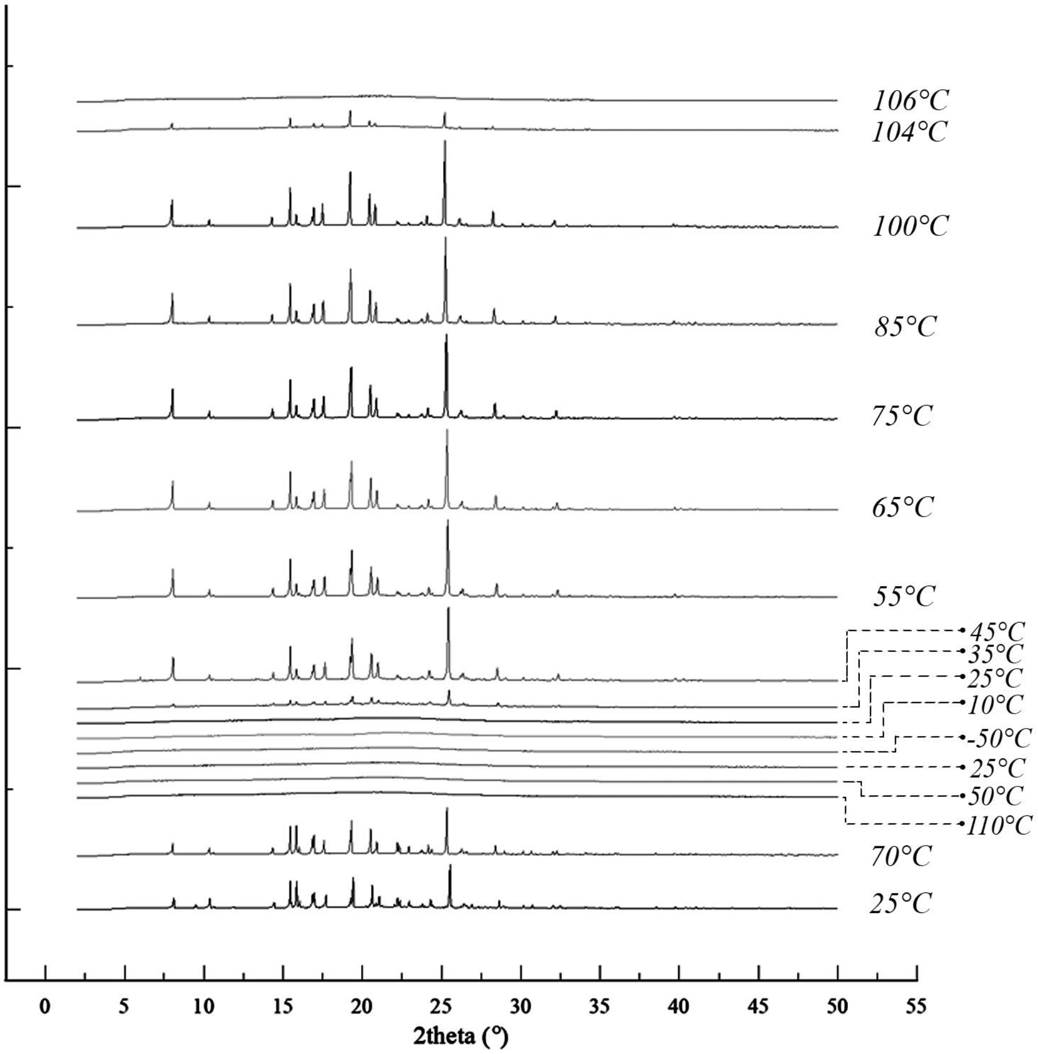


Figure 4.21: PXRD diffraction patterns collected during heating till 110°C, cooling till -50°C and heating again till 140°C, at a heating rate of 6°C/min, starting from form I.

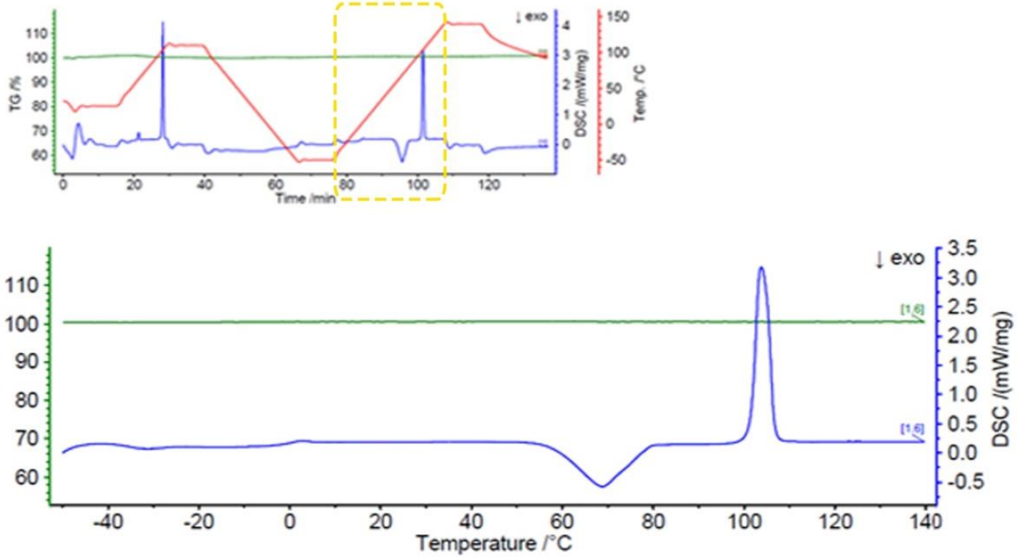


Figure 4.22: DSC/TG overlay of the exploratory run for form I, following the temperature profile defined for the vt-PXRD run in figure 4.21, also illustrated in the top graph. The enlarged image concerns only the second heating, indicated by the yellow dotted square above.

Same characterisation techniques were employed to observe any changes during the heating of sample containing a small content of form III (see appendix). The characteristic diffraction peaks of form III were present till 85°C, after which the diffraction pattern of pure form I was retained till its melting point. Despite not showing a separate endothermic peak, as in literature, the melting peak of form I had an onset at a much lower temperature (~87°C) than usual, probably due to the minimal content of form III in the sample.

The collection of measurements processed for the thermal analysis of form I provide evidence to support the premise that this polymorph is thermodynamically stable below its melting point. The evidence is mainly concentrated on the fact that all other forms transform to form I with an increase in temperature.

4.3.2.3 Crystal structure analysis

The following section is dedicated towards the analysis of form I at different molecular levels, to derive information about which factors influence its polymorph stability. The molecular structure in figure 4.23 also includes three unbonded hydrogen atoms around the methyl group, indicating the degree of disorder present within the unit cell. It is interesting how despite being equivalent with respect to composition, only one methyl group experiences such disorder. This is probably due to the difference in their chemical environment, caused by their respective orientation relative to the backbone. The effect of disorder will be investigated in terms of intermolecular interactions and also from a thermodynamic point of view.

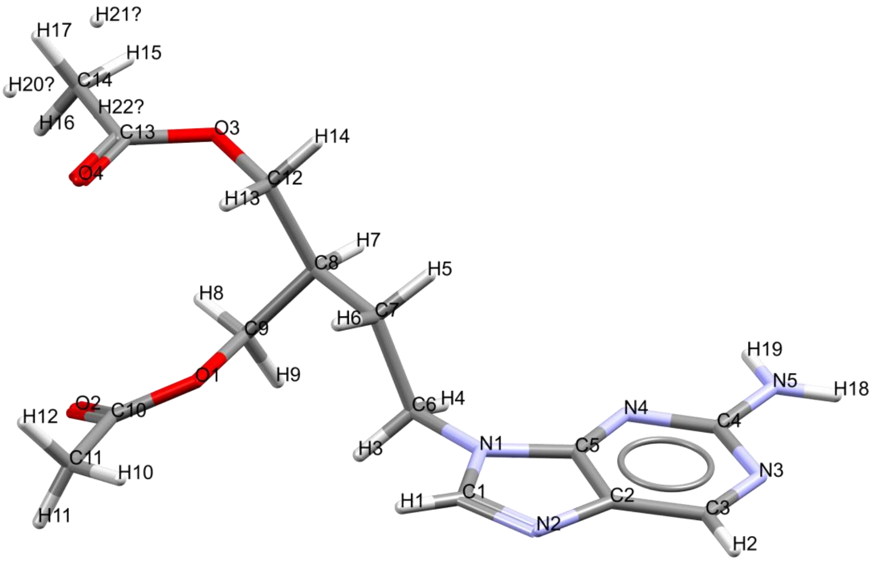


Figure 4.23: The molecular structure of FCV present in the asymmetric unit of IMADEI, including the degree of disorder represented by three unbonded hydrogen atoms.

4.3.2.3.1 Intramolecular level

Given the relatively large branches of FCV, it was essential to ensure that the molecular geometry of the structure was defined appropriately. This procedure was done through the utilisation of *Mogul,* which classifies each geometric parameter based of its compatibility with the corresponding predicted preference, as suggested by relevant entries available in the depository of CSD-libraries.[166] The search concluded that form I has no unusual parameters, suggesting the absence of specific issues with the structure. Furthermore, this result might also indicate the lack of extreme strain present along the structure, which might have been exposed through unusual geometric features.

The torsion angle C1N1C6C7 (82.434°) positions the branches almost perpendicularly with respect to the aromatic system, thereby reducing steric hindrance influences. The histogram in figure 4.24 demonstrates how despite being within the usual range, the majority of values for this torsion angle lies at larger angles, with a local density of 0.22. Another crucial location relates to C8 which is located along the straight chain and positions the three main components of FCV (see appendix). The angles around C8 highlight how the distribution prioritises the maximisation of the distance between the backbone and the ester group (starting with C9), due to the small torsion angle of C6C7C8C9 (45.25°), which places such groups almost at parallel positions. Conversely, C7C8C12 is slightly less than the standard tetrahedral angle (108.51°), as the exertion of repulsion is of much lesser degree. This illustrates how the considerable length of carbon chain provides enough space to optimise the intramolecular geometry, hence reducing strain. The possibility of having multiple ways of optimising the three groups around C8 might be a potential factor which enables FCV to exist in three anhydrous forms. This might be the case for form II, given the similarity of its experimental data relative to those of form I.

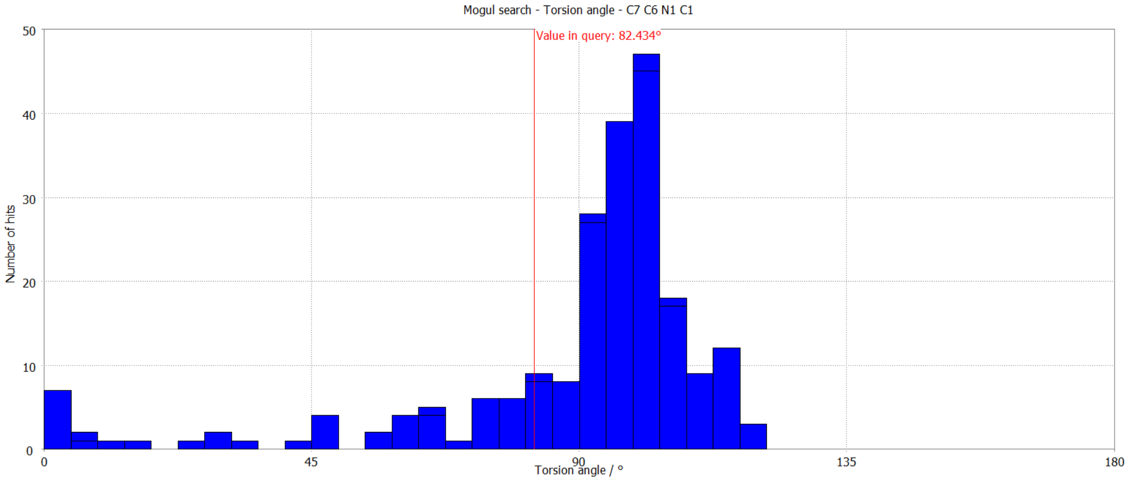


Figure 4.24: Histogram showing torsion angle C1N1C6C7, with a high amount of hits located within the 90° to 110° region.

The electrostatic potential map of form I was constructed by mapping the total potential energy experienced by each fragment onto the molecular shape, while retaining the spatial coordinates (see appendix). Apart from the expected extreme areas around the carbonyl oxygen atoms, the aromatic nitrogen atoms and the hydrogen atoms of the amine group, the majority of the map was rather within the neutral range. Regions associated to the oxygen atoms of the ether group had higher potential energy compared to the oxygen atoms in the carbonyl groups. A slight decrease in the potential energy was observed above the aromatic ring system, due to the presence of delocalised π-electrons. These observations suggest that the intermolecular network of form I is likely to be composed of few strong HB assisted with either C-H···π interactions or π···π stacking.

4.3.2.3.2 Intermolecular level

4.3.2.3.2.1 Exploring Expectations

The initial screening for potential intermolecular interactions was performed through *IsoStar*, which constructs contour plots to indicate the amount of entries involving particular interactions. In order to obtain a more representative plot, only contacts whose distance was less than the sum of VdW radii with a tolerance of 0.5 Å, were included.[289] Given the absence of purine ring as a central group in *IsoStar*, information regarding the backbone of FCV was derived from the plots constructed for GCV using aromatic nitrogen atom as central group. Data points representing nitrogen, sulfur and oxygen probes (in a X-H format, where X=N,O,S) exhibited the strong HB accepting ability of the nitrogen in the central group, resulting in contacts with a considerable degree of directionality (see appendix).[133] In addition, data associated with carbon as contact group was clustered above and below the aromatic nitrogen, with the majority of hits being incorporated in aromatic rings.[133] This implies high propensity associated with involvement of delocalised π-electrons in the intermolecular network of structures as FCV.

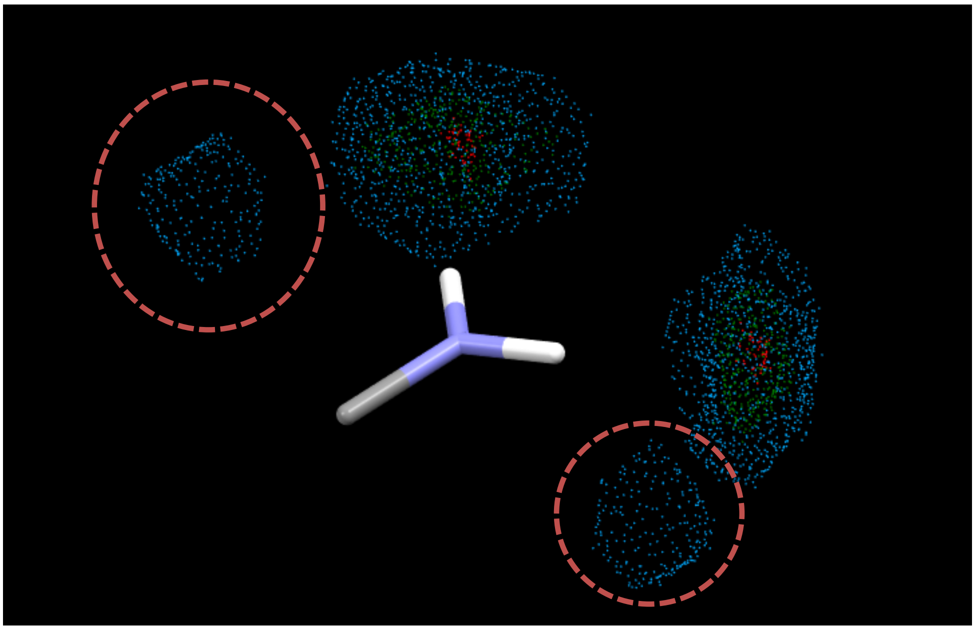


Figure 4.25: Scatterplot showing uncharged aromatic amine as the central group, with any nitrogen or oxygen as contact groups. The contour plot is scaled internally with level 25 (blue), level 50 (green) and level 75 (red).

The contour plot in figure 4.25 illustrates an uncharged aromatic amine as a central group with any nitrogen and oxygen as contact groups. The shortest interactions were located directly in front of the hydrogen atoms, highlighting the significant contribution of this group as a HB donor. The circled data points represented interactions with the neighbouring atoms of the central group, as in case of TIMCUO and ETAYUU, thereby not relevant to be taken into account.

Another scatterplot was constructed to explore the possible interactions around an ester group, as displayed in figure 4.26. Even though the oxygen atom in the –COC- is relatively polar, there were not any detectable contacts at the plotted levels. The absence of the participation of such oxygen atom in HB was also remarked in GCV polymorphs. As expected, the majority of data points were concentrated around the carbonyl group, due to the accessibility and electronegativity of the oxygen atom. These interactions were characterised with strong directionality and short distances (red points).

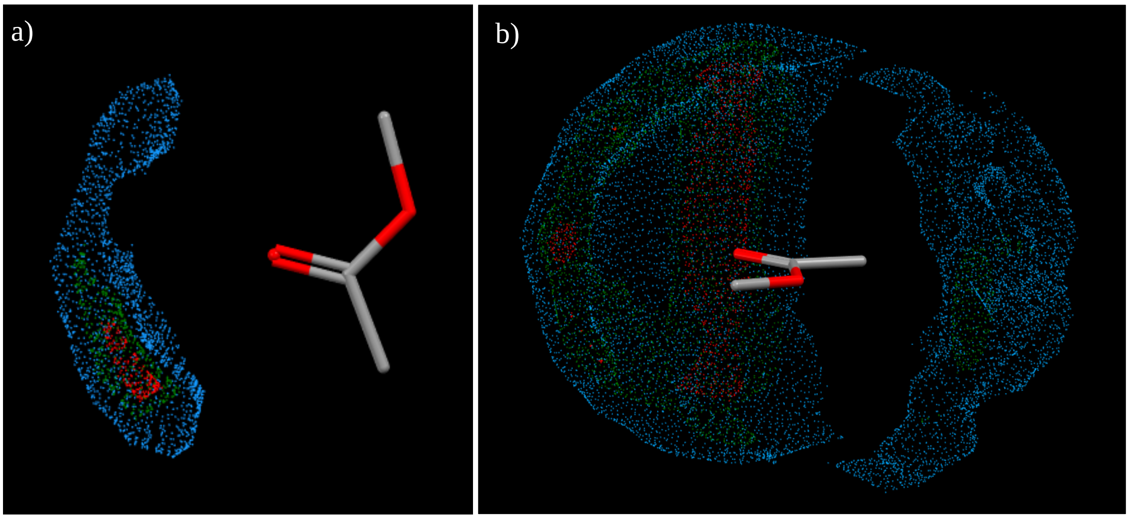


Figure 4.26: Scatterplot showing ester as the central group, with a) any polar group X-H, where X= N,O or S, and b) any alkyl -CH as contact groups. The contour plot is scaled internally with level 25 (blue), level 50 (green) and level 75 (red).

The likelihood associated with the participation of carbon as a HB donor in the presence of this ester group is exposed in figure 4.26b. However, given the common occurrence of alkyl groups in the CSD entries, there were several hits (as IHOHOC and GOCFOT) which despite containing a –CH in the vicinity of an ester group, their relative orientation still made it highly improbable to form an intermolecular HB. These cases were mainly clustered around the carbon ends of the ester group, and presented a clear overestimation of the propensity associated with ester - alkyl interactions. Nonetheless, the HB donating ability of carbon was demonstrated through the red points near the oxygen in the carbonyl group.

Full interactions maps were utilised to provide a better overview of the expected interactions within FCV, with contour plots adjusted according to the local chemical environment. The interaction preferences derived from such maps can be used to explore possible HB synthons which might be present in different polymorphs, as well as to identify regions where co-former or solvent might interact.[267] Figure 4.27 displays a contour plot at level 6, implying that the program mapped only regions where the density of the contacts located in the CSD data was six times relative to what was encountered at random.[273] Hotspots were also included to locate the areas with superior local interaction densities at that particular level, labelled according to the propensity associated with each probe.

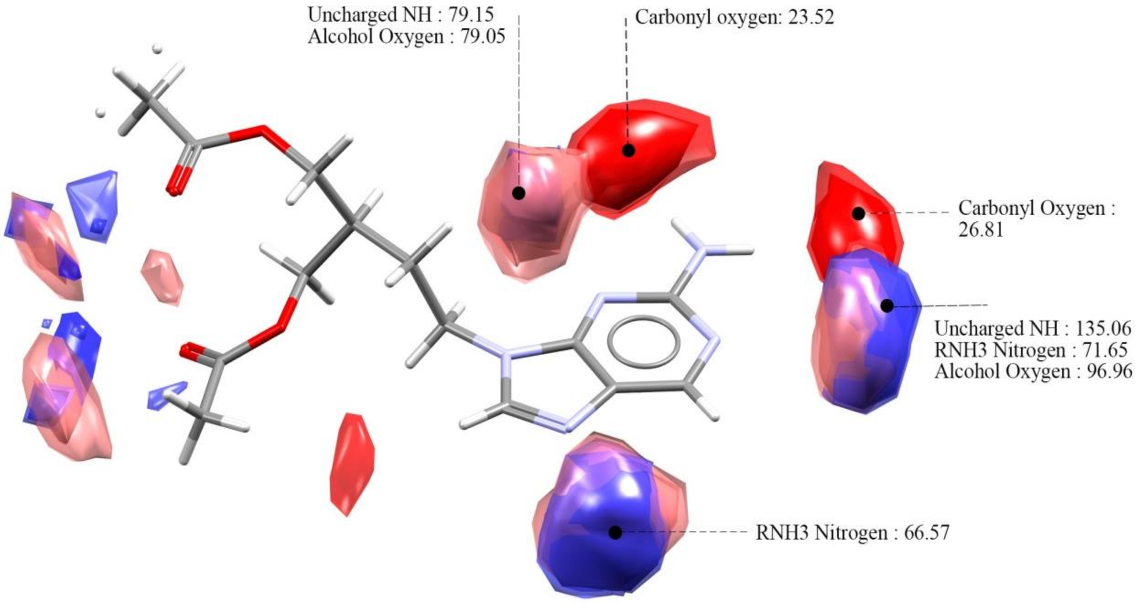


Figure 4.27: Full interaction map of FCV form I, displaying only contour plots at level 6, together with the corresponding hotspots, and the highest two propensities for each probe; uncharged NH nitrogen (dark blue), RNH3 nitrogen (light blue), alcohol oxygen (light red) and carbonyl oxygen (dark red).

This interaction map illustrates a conventional pattern of HB with preferences concentrated around polar groups, with the highest propensities corresponding to plots directed to aromatic nitrogen as a HB acceptor. This might have been a result of the particular molecular conformation in form I, which provides a greater degree of availability for aromatic nitrogen atoms compared to the carbonyl groups. Furthermore, the nitrogen atoms in the aromatic ring retain the lone pair of electrons in their respective

sp^2 orbitals, outside the delocalised π-electron system, thereby increasing their likelihood to contribute as HB acceptors. The lack of participation of O1 and O3 atoms (-COC) in HB was also highlighted, even at lower levels, in agreement when the preferences suggested by *IsoStar*. The unconventional participation of carbon as a HB donor can be observed through some red regions around C1 and the methyl group. Such contribution was highlighted even further when plotting the maps at more distant levels (see appendix). A similar plot was constructed using aromatic carbon as probe, but the mapped areas were very minimal, even when plotted at level 1.5 (see appendix). This was not compatible with the results derived from the *IsoStar* plot for the aromatic nitrogen – carbon pair, whose data points were distributed above and below the aromatic system.

The predicted HB network of form I by *Mercury* was compatible with the geometric parameters in the literature, consisting of six distinct contacts, with seven neighbouring molecules.[206] The superimposition of the full interaction map at various levels onto the molecule with its contacts showed how almost all the potential HB contributors are satisfied (figure 4.28). This is the initial evidence supporting the considerable extent of thermodynamic stability of form I shown experimentally.

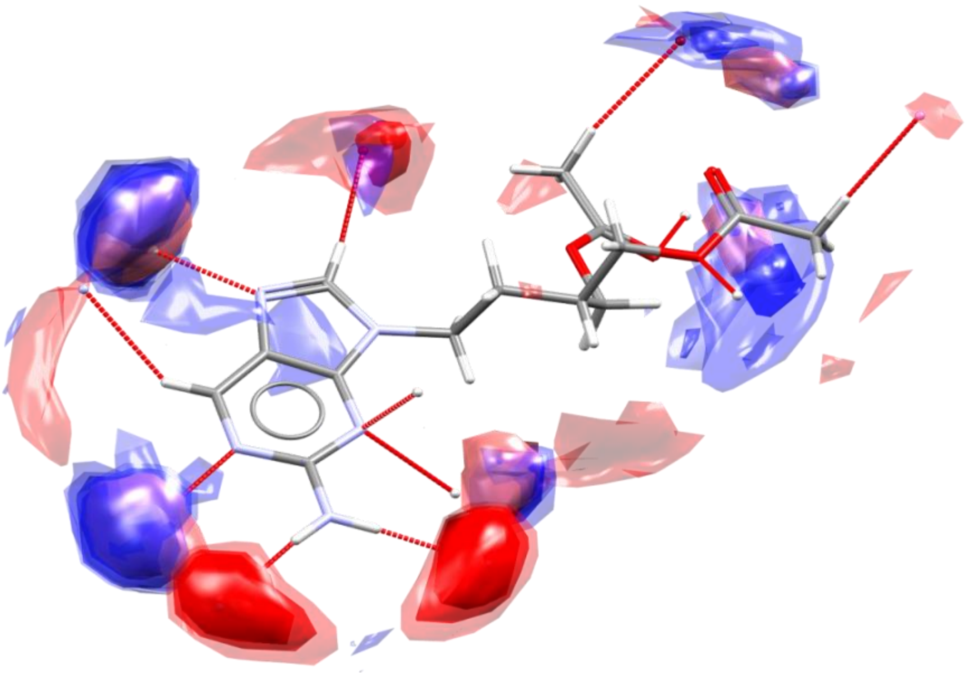


Figure 4.28: Full interaction map of FCV form I, displaying contour plots at level 2,4 and 6, together with the corresponding hotspots for each probe; uncharged NH nitrogen (dark blue), RNH3 nitrogen (light blue), alcohol oxygen (light red) and carbonyl oxygen (dark red). Hanging intermolecular contacts are shown in red.

The participants of HB depict the important role of carbon as a HB donor, probably due to the low proportion of conventional HB donors in FCV relative to HB acceptors. These interactions are characterised by relatively longer distances, and despite being classified as weak contacts according to criteria published by Desiraju and Steiner, they still can contribute up to 16 kJ/mol.[290] The absence of multiple strong HB might be the reason which allows FCV to be fairly soluble in polar protic solvents, as shown in the solubility screening. Relatively weaker intermolecular contacts might be comparatively easily broken and replaced by stronger HB with the polar groups present in such solvents. In these scenarios, the energy given out by the solvation, through the formation of such new stronger interactions would be enough to compensate for the disruption of the HB network within the pure crystal.

4.3.2.3.2.2 Hirshfeld surface analysis

IsoStar and full interaction maps provided considerable information concerning the role of conventional HB contributors, but less definite indications regarding the contribution of weaker contacts. The importance of investigating weaker interactions was emphasized through the calculation of the intermolecular potential in *Mercury*, shown in appendix. The strongest potentials (-41.4 kJ/mol and -41.7 kJ/mol) are directed towards neighbouring molecules which have no HB with the central molecule, hence exposing the crucial role of other non-bonding forces in the process of form I crystallisation (see appendix).[106, 270] Hirshfeld surface analysis was employed to assist the characterisation of such bonds, together with other non-conventional HB.

The construction of Hirshfeld surface illustrated in figure 4.29, with d_{norm} data mapped on it, highlighted similar traits concerning the framework of HB, through numerous red spots on the periphery of the molecule. Nonetheless, the decrease in the accessibility of some oxygen atoms is illustrated by light red spots, indicating relatively long distant contacts. Furthermore, the d_{norm} values for FCV form I had a wider range, with a more positive mean, when compared to the data corresponding to both GCV anhydrous polymorphs (see appendix).

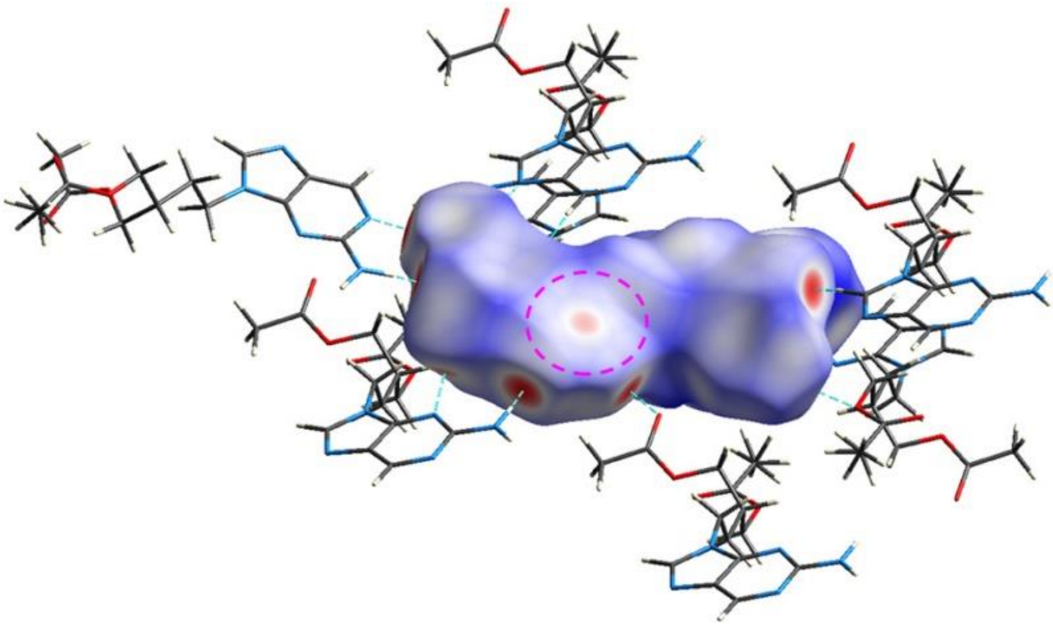
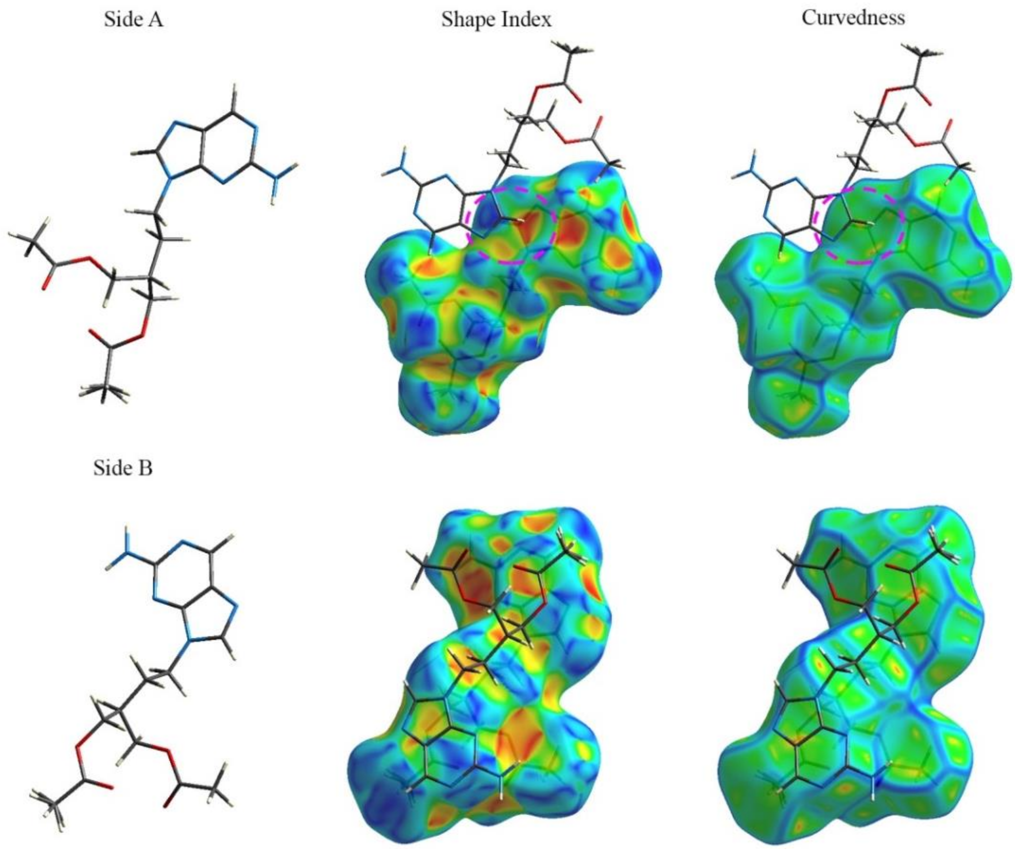


Figure 4.29: Hirshfeld surface for form I, mapped with d_{norm} over the range -0.447 to 1.5847. Neighbouring molecules which participate in the HB network (turquoise dotted lines) of the central molecule are shown, while excluding any hanging contacts.



Side A

Shape Index

Curvedness

Side B

Figure 4.30: Opposite interfaces of the molecular structure of form I, with their corresponding shape index and curvedness maps. Neighbouring molecules are also included to facilitate plot interpretation.

Despite both having distant interactions, the opposite sides of the molecule are characterised by very distinct type of non-bonding interactions due to the orientation of the branches. This variation can be observed through the distinguishable shape index and curvedness maps in figure 4.30. Both plots for side B captured its irregular surface due to the lack of shape compatibility between its interface and the molecule above it. The shape index plot for side A has complementary patch, outlined with a magenta circle, due to the presence of another molecule above it, related by an improper rotation, as defined in the space group. The region associated with this complementary pattern is characterised by a smooth, flat area, indicating close $C\cdots C$ interactions ($d_i \approx d_e \approx$ 1.6-1.8Å). This information coincides with the only extreme point present in the d_{norm} values above/ below the aromatic ring system, circled in figure 4.29. In fact, the presence of $C\cdots C$ contacts was mainly limited to such complementary patch only (see appendix). This result is in agreement with the predictions illustrated by the full interaction map using aromatic carbon as a probe, as opposed to the observations extracted from *IsoStar* contour plots. This happened because the packing arrangement of form I did not allow the formation of $\pi\cdots\pi$ interactions along a larger area. Such effect could only have been detected by the full interaction maps, given that they take into account the local orientation of the molecule, instead of isolating each functional group.

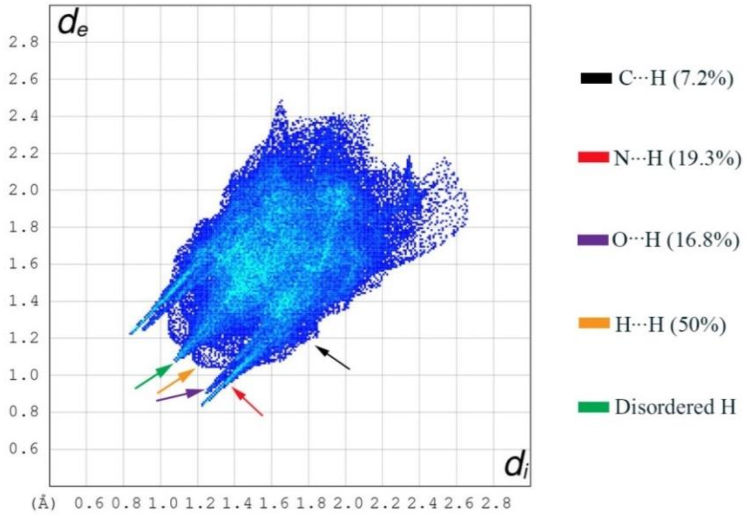


Figure 4.31: Fingerplot for form I, plotting d_e (Å) against d_i (Å). The arrows represent interactions between different participants as indicated in the adjacent legend. Given that the plots are approximately symmetrical, the arrows can be mirrored through the x-y diagonal.

The composition of the interatomic pairs detected by the Hirshfeld surface can be explored further through fingerplots, shown in figure 4.31. The significant population of

H···H contacts reflects how conventional HB is not the most abundant type of interaction around the overall surface. The percentage contribution of HB in the GCV polymorphs is more dominant relative to FCV form I, possibly due to the lower ratio of polar groups present in the latter antiviral. Despite the abundant presence of H···H contacts, their majority is associated with distances larger than sum of their VdW radii, hence possibly making them less influential towards the stabilisation of form I. Furthermore, the nature of the interactions corresponding to the minor group of dihydrogen contacts with $d_e+d_i <$ 2.4 Å, can be either repulsive or attractive.[280] In addition, the H···H contacts with the smallest interatomic distances corresponded to the disordered hydrogen atoms, whose presence increased the percentage contribution by 0.7 %.

N···H/H···N and O···H/H···O interactions (36.1%) are represented by pronounced spikes, with a minimum distance of $d_e+d_i \approx 2$ Å and 2.1 Å, respectively. They are responsible for the contacts identified by the red spots on the Hirshfeld surface, and can be considered to be the major contributors towards the electrostatic forces within the crystal structure of form I. The wide variation in the distances of such interatomic pairs can be partially attributed to the nature of the HB donors, namely either nitrogen or carbon. Despite having fairly long distances (most distances between the closest C-H groups and the purine ring system exceed 3.5 Å), the percentage contribution of C-H···π interactions was rather significant.

4.3.2.3.3 Supramolecular level

The crystal structure of form I is centrosymmetric ($P2_1/c$), starting from one molecule in the asymmetric unit, which then can be used to provide the positions of the other three molecules through the symmetry operations and translations defined by the space group (2_1 screw axis, *c*-glide plane and inversion), as illustrated in figure 4.32a.[206] The packing arrangement is composed of two consecutive layers consisting of alternating parallel pairs of purine systems and opposite branched chains. The strongest HB interactions involving nitrogen as a HB donor, contribute significantly towards the network within the aromatic layer, whereas contacts characterised by carbon as a donor, dominate the rest of the structure. This led to a more spacious packing (coeff.= 0.703882), relative to GCV forms, possibly to allow the bulky branches to endorse an optimal orientation, thus reducing strain.

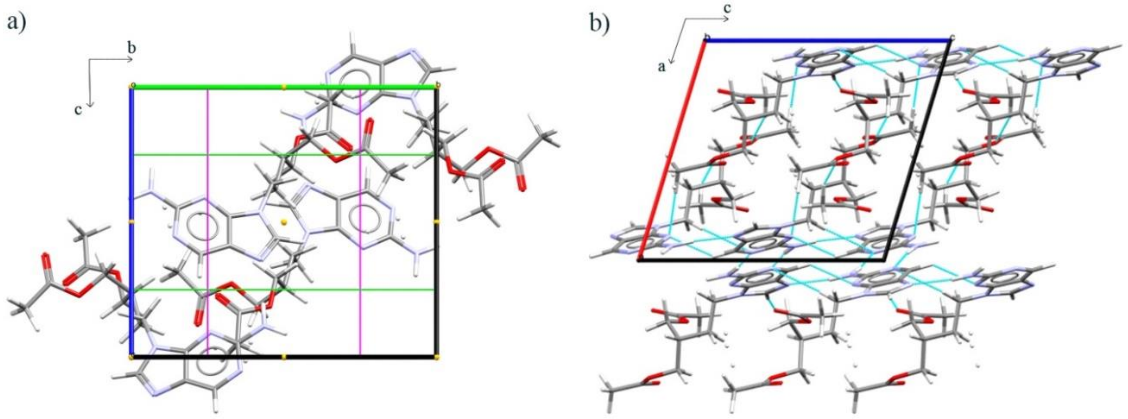


Figure 4.32: a) The unit cell of FCV form I, viewed down the *a*-axis, together with its symmetry operations; two 2-fold axes (dark green), two glides (pink) and a central point of inversion (yellow). b) Packing in form I, excluding hanging contacts for better visualisation.

The lower extent of compactness within the layer of the branched chains led to the estimation of voids, using a probe radius of 0.7 Å, with an approximate grid spacing of 0.7 Å. The voids were searched through the solvent accessible surface, so as to increase the likelihood of identifying voids which can accommodate solvent molecules. The present voids within the structure constitute of 0.3% of the unit cell volume, with an area of 4.3 Å^3. The voids are mainly positioned between the branched chains of the molecule, bounded within layers of consecutive aromatic rings (see appendix). The volume percentage occupied by the voids is almost negligible, therefore implying that the structure is highly unlikely to host guest molecules without having a change in its crystal structure skeleton. This factor can be proved by the distinct HB framework and packing arrangement present within the crystal structure of the monohydrate, relative to that of form I.[206] This feature is also expected to be observed for the methanol / ethanol solvate of FCV.

Crystal morphology plays a crucial role in the macromolecular properties of a crystal structure, particularly on the efficiency of industrial processing. This was highlighted through research studies as that of Tedesco *et al.* where it was shown how the modification of the morphology of a particular drug highly improved its industrial process.[291] Morphology is also a critical factor with regards to the distinction between polymorphs, as shown by Calvo *et al.* during their investigation involving carboxylic acid polymorphs.[292] Particularly, the calculation of growth rates and morphologies is used as a tool to decrease the number of potential polymorphs, and therefore help towards the identification of the actual risk of polymorphism.[35]

The morphology of FCV form I was predicted using the BFDH law, which states that relative growth rate of a plane is inversely proportional to its inter-planar distance. This implies that the faces with the highest surface areas are expected to be characterised by the largest inter-planar distances and lowest surface energies. Given the space group of this crystal structure, it was expected that the corresponding habit manifests itself as a monoclinic form. This prediction is in agreement with the monoclinic shape featured through the microscope images in figure 4.19. The most important morphological faces of this form are (100) with d_{100} =10.606 Å, (0-11) and (011) with d =8.459 Å, as well as their respective mirrored planes. Their dominant existence in the crystal habit is also shown visually in figure 4.33, as well as through their BFDH relative areas. Conversely, planes with the smallest d-spacing, namely (10-2) and (020), are also defined by the smallest BFDH relative areas, because they are positioned perpendicularly to a direction with a fast growth rate. In fact, the plane (020) exposes the nitrogen in the amine group which participate in the two strongest hydrogen bonds.

The crystal habit is determined by the growth rate of each plane, which in return depends on the rate at which molecules attach themselves onto such surfaces. When viewed down the *a*-axis, the layering of molecules becomes very evident, with strong hydrogen bonds from the aromatic ring systems acting as forces to attract more molecules (see the contacts at the periphery of the monoclinic figure). This leads to (100) and (-100) planes to have the highest BFDH relative surface areas and slowest growth rate. These contacts are directed both towards the adjacent aromatic system and also towards molecules in the consecutive layer, thus ensuring a continuous strong build-up of layers.

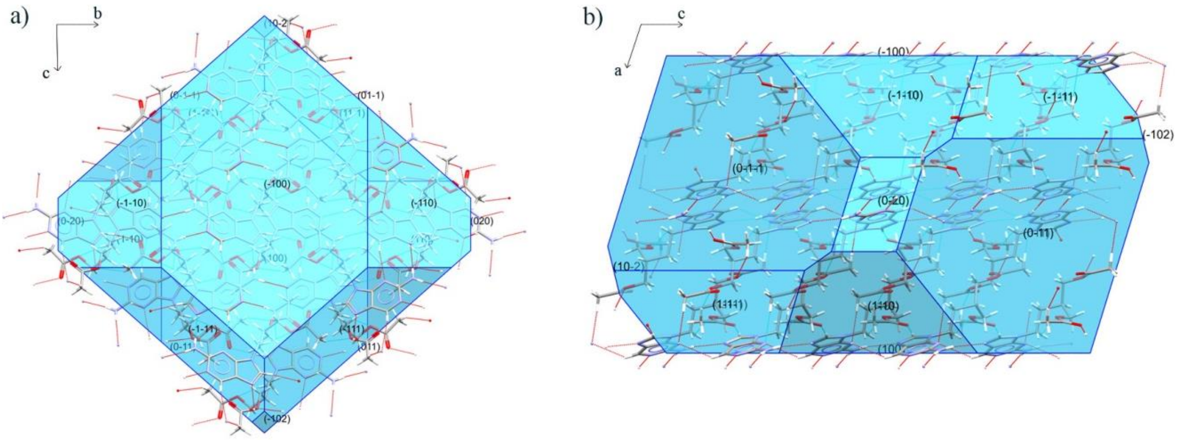


Figure 4.33: Predicted monoclinic morphology of FCV with molecules, together with a network of HB (turquoise) (scale: 100) as viewed down a) *a*- axis and b) *b*-axis. Hanging contacts are coloured in red.

4.3.2.3.4 Thermodynamic stability – Energy frameworks

The examination of different energy components can facilitate the process of understanding the mechanism behind the thermodynamic stability of form I. Thermal analysis in section 4.3.2.2 showed its considerable extent of stability through the fact that all other forms can transform directly into form I with a specific increase in temperature.[205] Conversely, the production of form II requires the recrystallization of form I or of a mixture of both forms, and very often the final product will still contain form I component.[205]

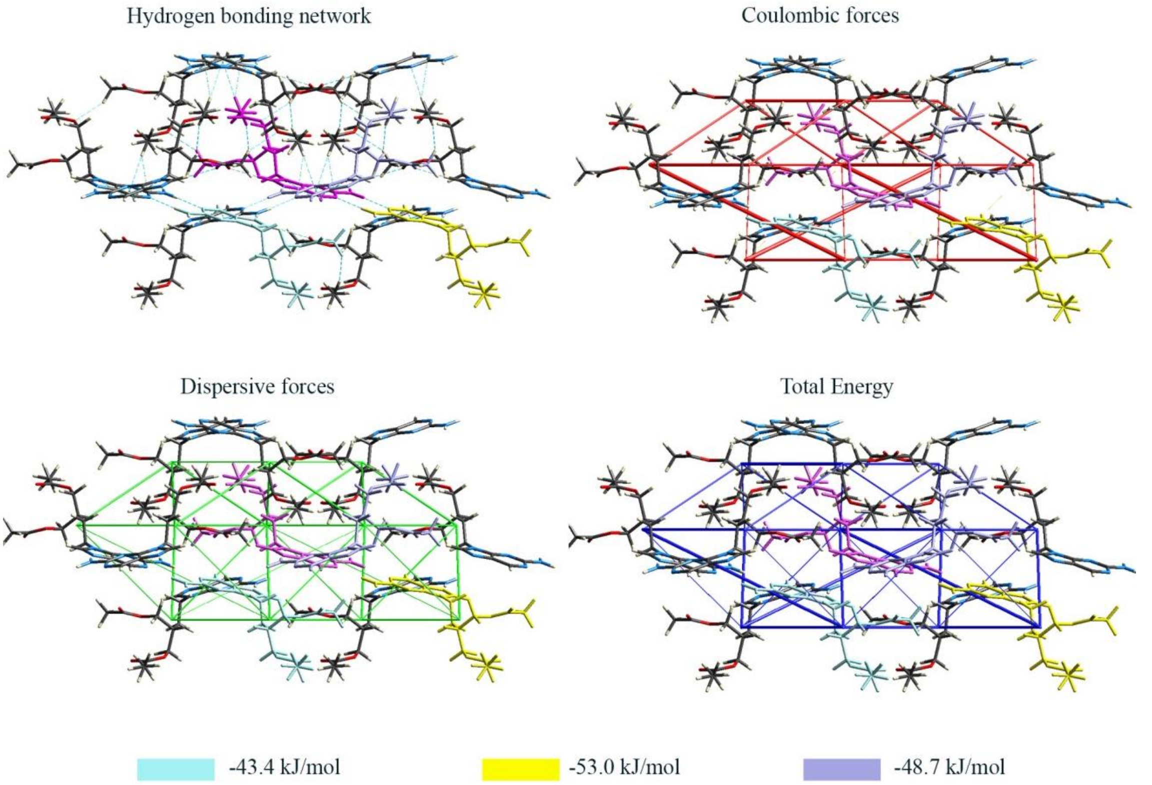


Figure 4.34: Energy framework diagrams, using CE-B3LYP model, with a scale tube size of 20 and cut-off value of 0 kJ/mol. Cluster of molecules with 3.8 Å radius around the central molecule (magenta). The solid coloured molecules (referred to in text), have their corresponding interaction energies (kJ/mol) listed in the bottom legend.

The strongest energy contributor (-53.0 kJ/mol) is directed towards an adjacent molecule (yellow) whose purine ring system is parallel to the central molecule (magenta), as shown in figure 4.34. This significant contribution of electrostatic nature is a result of a short, directional HB involving N5-H···N3, which was not detected by the UNI intermolecular potential calculation in *Mercury*. Strong Coulombic forces induced by the presence of C3-H···N4 and N5-H···N2 interactions are located between alternating pairs (each consisting of two parallel purine rings) in a non-parallel fashion (magenta and lilac molecules). These contacts coupled with a relevant dispersive component (-48.7 kJ/mol) provide a substantial degree of stability across the layers of purine rings.

While electrostatic forces seem to be concentrated along the layer of aromatic ring systems, as illustrated by thick diagonal tubes, dispersive forces can be found in all directions. This might be a result of the presence of several C-H···X (X=O,N) interactions, because their electrostatic nature is of a smaller magnitude, and it becomes even lower than dispersion term in contacts of a weaker nature as those involving a methyl carbon (C11-H···O3 and C14-H···N4).[290] The electrostatic term decreases with a reduction of the polarity of the C-H···X contact, whereas the dispersion term remains unchanged. The cumulative stabilising contribution of these relevantly weak interactions might be significant during the process of crystal formation, especially given the size of FCV molecule.[293]

On comparison with the Hirshfeld surface analysis, the UNI intermolecular potential and the scatterplots of *IsoStar* were indicative of a more dominant contribution of C-H···π and π···π interactions. Nonetheless, the interaction energy between the pair of molecules (magenta and light blue) identified through the complementarity patch in the shape index plot (figure 4.30, side A) is -43.4 kJ/mol. Such pair of molecules were highlighted earlier as they were associated with the most negative UNI intermolecular potential, despite not having any HB contacts between them. This demonstrates how despite having minimal percentage contribution and moderate centroid to centroid distance (3.812 Å), π···π interactions still participate towards the thermodynamic stability of form I.

4.3.2.3.5 Thermodynamic stability- Lattice energy

Before evaluating the lattice energy of form I, it was important to investigate the effect of disordered hydrogen atoms within the unit cell. Firstly, both alternative molecular conformations were geometrically optimised at B3LYP/def2-TZVPD level of theory, with dispersion correction. Both distinct conformers converged to the same minimum, suggesting that the positional disorder has low energy barriers.[294-295] This was followed by the energy minimisation (using PBC) of the two alternative experimental structures of FCV, which resulted in almost geometrically identical unit cells, with fairly equivalent energy (ΔE = 0.0173 kJ/mol). The d-DFT energy minimisation reproduced the experimental structure very well, with a minimal r.m.s Cartesian displacement (< 0.20 Å), in-line with the validation study for d-DFT calculations suggested in literature.[295] These results indicate that such disorder is not influential with respect to the lattice energy. This observation was also supported by negligible difference obtained between the cumulative lattice energy values of the unit cell with and without disorder, using the CE-B3LYP model with dispersion correction (see appendix). Such result might be due to the location of disorder, which only relates to one methyl group with one relatively weak intermolecular bond. This is also in agreement with the small change in the percentage contribution attributed to the presence of disorder. Further analysis on such disorder can be assisted using molecular dynamics calculations and variable temperature single crystal XRD.

For the final evaluation of the cumulative lattice energy in *Crystal Explorer*, the disordered hydrogen atoms were retained, as shown in the fingerplots. The computation of the ΔE_{intra} penalty was less complex due to the low likelihood of forming a strong intramolecular bond within the low energy conformer of FCV. This resulted in a small ΔE_{intra} penalty of 5.360 kJ/mol, possibly due a weak interaction between the carbonyl oxygen and a hydrogen atom from a methyl group (see appendix). The lattice energy values obtained from both methods are fairly compatible, highlighting their good extent of accuracy in cases with no strong intramolecular HB.

Table 4.3: Lattice energy values for FCV form I. (Please see appendix for more details)

Description	**Energy value (kJ/mol)**
Cumulative lattice energy, $\boldsymbol{E_{tot}^{CE-B3LYP}}$	-203.700
Cell dipole Energy correction, $E_{cell\ dipole}$	Not required
Intramolecular energy penalty, ΔE_{intra}	5.360
Lattice Energy, E_{lat} (Crystal Explorer model, *eq.3.6*)	**-198.340**
Lattice Energy, E_{lat} (PBC, *eq.3.8*)	**-203.455**

4.3.2.4 Risk assessment of polymorphism

The high solubility of FCV in polar protic solvents, particularly water, can be explained as a consequence of the lack of strong HB donors within its molecular structure, thus leading to more energetically favourable solvent-solute interactions. However, FCV resulted also to be soluble (>40mg/ml at room temperature) in aprotic solvents as acetonitrile, implying that there are other factors promoting such solvent-solute contacts. Such relatively high solubility might be due to presence of an overall fairly weaker HB network relative to other antivirals, namely GCV, shown by smaller energy contributors when analysing its energy framework and lower melting point.

Nonetheless, this characteristic seems to be only relevant in comparison with other antivirals and not relative amongst FCV forms. This was illustrated through the dominant presence of form I after re-crystallisation from various solvents, implying that its HB framework is still the most energetically favourable compared to other known FCV forms. Furthermore, this feature was the resultant outcome of the experimental thermal analysis, where by all analysed forms transformed to form I when heated at particular temperatures. Crystal structure analysis explained how given the available molecular composition, all intermolecular contributors including C-H···X contacts and π···π interactions, work collectively and fairly optimally to maintain the thermodynamically stable character of form I. These outcomes imply that the ability for form I to transform itself into another unknown anhydrous stable form is rather minimal. This characteristic will be lastly tested through the following HB propensity model tool.

4.3.2.4.1 HB propensity model

The molecular structure of FCV was represented by five substructures, ensuring that all available HB participants are included (see appendix). The list of HB contributors also contained carbon as a HB donor, given its crucial role within the HB framework of form I. One should note that the disordered hydrogen atoms were required to be removed before initiating the search in CSD, and this was accepted assuming that their impact on the overall lattice energy of form I is minimal (as discussed in section 4.3.2.3.5).

Due to the nature of weaker interactions, the bounding parameter for a HB angle was lowered to 110°, while retaining the default settings for the contact distance. The training set consisted of 4608 hits, each of which was composed of at least one of the targeted functional groups. All the available donor and acceptor categories were integrated within the model due to substantial amount of hits respectively. After minimal refinements for optimisation, the final model contained several statistically significant predictors, including some catering for competition, steric density and aromaticity. The goodness of fit was confirmed by the value for the area under ROC curve (0.888), ensuring that the model is able to differentiate between the two categories.[296] Furthermore, the reduction from the null to the residual deviance exhibited the significance of all the predictors present in the parsimonious model.[297]

The final model produced the propensities of all potential HB, including those corresponding to form I (see appendix).The coefficients of the regression model provided an indication of the favourability of each particular HB participant, with N5 having the most positive influence as a HB donor. Conversely, the oxygen atoms within the ester group (-COC) were accompanied by the lowest coefficient (-1.814), due to their uncommon appearance as HB acceptors in the presence of carbonyl oxygen atoms. All the factors associated with carbon donors had relatively small coefficients, indicating their mild influence on hydrogen bond propensity, as opposed to that related to nitrogen in the amine group.

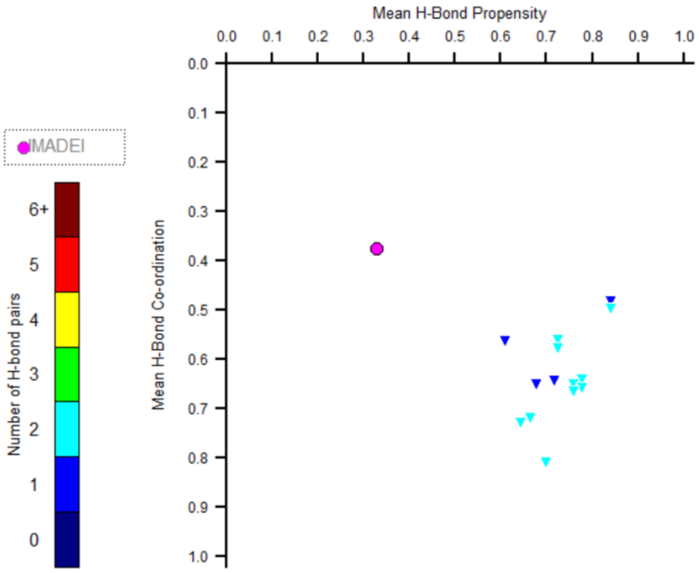


Figure 4.35: The propensity participation chart for form I (violet), together with other hypothetical forms (blue). The colour associated with the potential forms corresponds to their number of HB pairs.

The model also produced a propensity participation chart (figure 4.35), featuring potential forms with different combinations of non-bonding contacts. The propensity chart and the interactions present within more favourable potential forms exposed how the underlying mechanism of the statistical model was mainly based on conventional HB, with N5H···O2/O4 being the most likely interactions. These hypothetical forms only contained one or two interactions, involving only N5, given that it is the only conventional donor. One can also notice how even though the carbonyl oxygen atoms were involved in the most probable contacts, their highest propensity values were still associated with zero coordination. This constrained the list of contacts in the most stable hypothetical form to N5H···N2 and N5H···N3 interactions only.

The model placed form I in a non-optimal area, with relatively low values for the mean HB propensity and coordination. This classification was a result of the presence of some statistically unfavourable interactions within form I, particularly concerning the participation of O3 and N4 (twice) as HB acceptors; a participation which was possible due to the inclusion of CH as a HB donor. In fact, the favourability of form I in the propensity chart was categorised in an optimal area when HBP model was reconstructed with the exclusion of CH HB contributors, despite not capturing the entirety of its HB network. This outcome was probably possible due to the few conventional HB donors in FCV (the exclusion of CH for GCV HBP model still failed to predict accurately the stability of form I).

Given the size of FCV, it is unlikely that the most stable forms would include such low amount of HB, without the presence of weaker interactions. However, the relevant degree of stability provided by such weaker contacts involving carbon was not taken into consideration. This factor was also observed during the analysis of the HBP model of GCV form I, thereby making it a presumably persistent limitation of this approach. Given these observations in conjunction to the experimental and computational outcomes described earlier, it would be inaccurate to rely on the HBP model results to rank the stability of form I.

4.3.3 Valacyclovir Hydrochloride

4.3.3.1 Form characterisation

The existence of multiple VCV·HCl polymorphs can be witnessed in various patents, which mostly include PXRD patterns for identification, accompanied by the respective acceptable water content.[209-214] The comparison of such diffraction patterns suggests that these forms are characterised by low symmetry, due to the considerable population of peaks, as well as low crystallinity. The detection and classification of possible disorder can be difficult to be extracted from such PXRD data, because atoms are spread over a larger amount of unique positions, thus resulting in reduced peak intensities, strong correlation with background at large 2θ values and increased peak overlap. These factors led to the absence of structurally determined VCV·HCl crystal structures. In addition, several forms share multiple characteristic peaks along the whole range of 2θ values, hence making it difficult to identify the forms solely on PXRD data, especially in the presence of mixtures.

The anhydrous polymorph described in one of the earliest patents as the active ingredient, written by Carter *et al.* is referenced as form I, while the other confirmed anhydrate phase is noted as form II.[210, 213] The rich list of hydrates includes a hemihydrate (form III), monohydrate (form IV), sesquihydrate (form V), dihydrate (form VI) and others.[209] The numbering of forms used in this document is novel due to the lack of a standard system used amongst published material. It is noteworthy to add that the ratios between water to API molecules were established using Karl Fischer method, and not by solving the crystal structure. There are other patented forms with less than 2% water content, hence possibly also considered as anhydrates.[211]

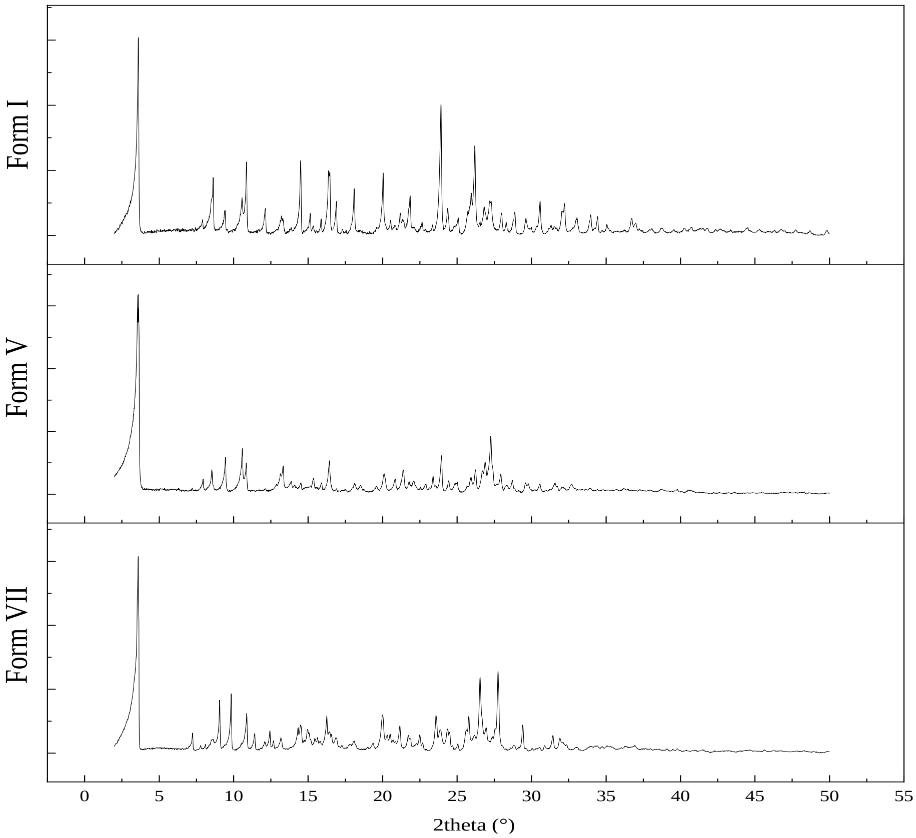


Figure 4.36: PXRD patterns of different forms obtained from solubility screening, collected at room temperature.

Due to the low crystallinity of VCV·HCl, PXRD patterns obtained from flat plate setup were of lower quality. In addition, due to the high population of weak peaks, it was difficult to determine the presence of an isolated form or a mixture. When possible, measurements were collected again using capillary setup, to increase the quality of the data. In general, solubility screening results contained mixtures, with a varying degree of percentage compositions. The three samples with relatively smaller amount of impurities were selected for analysis and characterised by the diffraction patterns in figure 4.36. The identification of the contents in other mixtures will be investigated in further work. The first PXRD pattern collected from the commercially purchased sample, represents dominant content of anhydrous form I, with some traces of other forms. The peaks located at 9.40° and 10.58° 2θ are characteristic of hydrates, however not distinguishable between the monohydrate and sesquihydrate. A maximum water content of 3% w/w is a fairly common occurrence in samples of form I.[213] The second PXRD pattern corresponds to the sesquihydrate form (V), which re-crystallised from a binary solvent mixture that included water. This sample also contained minimal content of the hemihydrate form (III), indicated by diffraction peaks at 7.84°, 15.31°, 20.78° and 23.41° 2θ.

The polymorph re-crystallised from a solution of water and 2-methoxyethanol, with a ratio of 1:9, is associated with a novel diffraction pattern, with characteristic peaks defined in table 4.4. The PXRD pattern of this form (VII) was compared with all diffraction data present in literature to the author’s knowledge. Furthermore, the possibility of having a mixture of only known forms was eliminated, since none of the possible combinations of existing patterns reproduced exhaustively the PXRD pattern of form VII. In particular, the existence of consecutive diffraction peaks at 9.08° and 9.85° 2θ is unique amongst other data. This feature does not exclude the presence other forms in the sample, namely form I.

Table 4.4: Characteristic diffraction peaks of VCV·HCl forms, collected during solubility screening.

Polymorph	**2θ of characteristic peaks (±0.2°)**
Form I	3.6, 8.6, 10.8, 12.1,14.5, 16.4, 16.9, 18.1, 20.0, 21.8, 23.9, 26.2,30.5, 32.2
Form V	3.6, 8.5, 9.4, 10.5, 10.8, 13.3, 16.4, 20.1, 20.8, 21.4, 23.9, 26.8, 27.2
Form VII	3.6, 7.3, 9.1, 9.8, 10.9, 16.2, 20.0, 21.2, 23.6, 25.8, 26.6, 27.8, 29.4

The FTIR spectra of the re-crystallised forms are illustrated in figure 4.37, and the similarity in trends makes it difficult to extract reliable information regarding the involvement of specific functional groups in HB. The presence of various fragments in different local chemical environments, such as the NH on the guanine ring and the NH on the chain, contribute towards the broadness of absorption bands over the range of 3600 – 2800 cm^{-1}. The degree of such broadness varies across all spectra, with the most defined bands corresponding to form I, which is known to contain only a minimal content of water. The spectra of form I and VII are characterised by the dual bands at 3444 cm^{-1} and 3325 cm^{-1}, which are typical for OH/NH stretching vibrational modes, as seen for GCV and FCV. A notable shift concerned the CO stretching band, which occurred at 1093 cm^{-1} for form I and at 1101 cm^{-1} for the other forms.

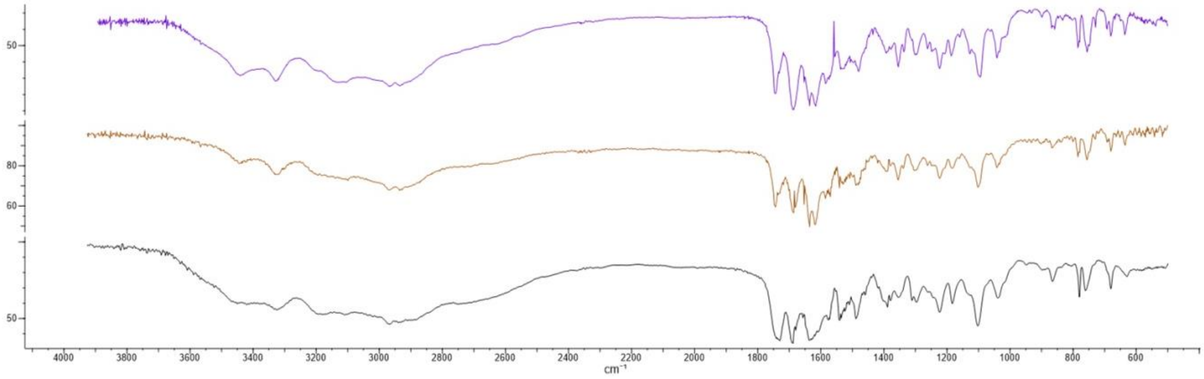


Figure 4.37: FTIR spectra of anhydrate form I (purple), form VII (brown) and form V (black).

4.3.3.2 Thermal Stability – Experimental characterisation

Literature provides a wide range of methods to produce pure forms or a mixture thereof, most of which vary in terms of the solvent used and conditions of drying. The formation of particular VCV·HCl polymorph is influenced by various factors, including nucleation and seeding, but mostly water content in solvent mixture and in the initial form. Variations in these criteria, even minimal, can lead to unexpected outcomes.[210] This feature together with the high occurrence of mixtures imply that the difference in the energetic feasibility between forms is not very substantial, thereby increasing the risk of polymorphism. This is also noticeable through the similarity in the melting points of the anhydrates form I and II, with an endothermic peak at 227°C and 223°C, respectively, followed by decomposition.[209] Given the absence of any other peaks in their DSC graphs prior to melting/decomposition, one can infer that there are no phase transitions between pure form I and II with heating.[209, 298]

Firstly, the thermal analysis of form I was conducted using hot-stage microscopy, with no considerable changes till around 230°C, except for some granule movements over the ranges of 75-100°C and 150-190°C. Decomposition was clearly noted at temperatures above 230°C (figure 4.38a), and it occurred in various stages without being preceded by a proper melting process. A melting point apparatus was utilised to obtain a rough estimation of the initial temperature of decomposition, through visual monitoring. Only minor signs of melting were visible prior to the first appearance of decomposition at around 195°C.

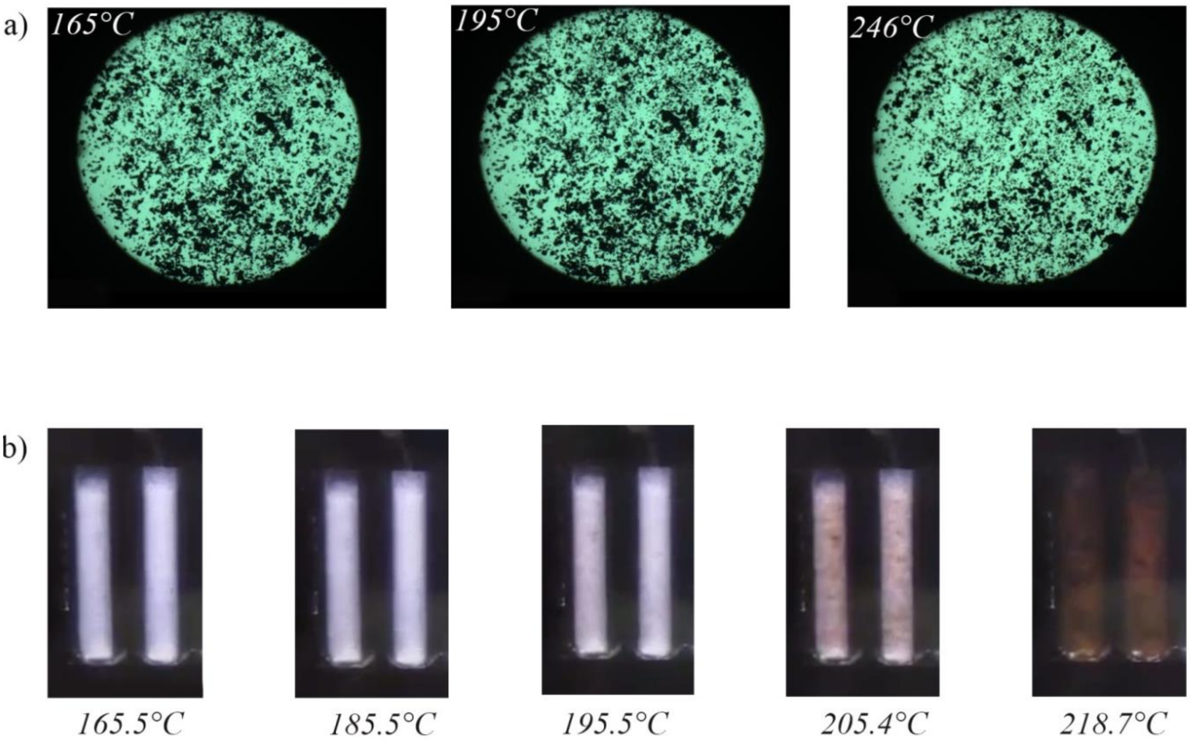


Figure 4.38: Heating of form I (commercially available form) at a rate of 5°C/min as observed through a) hot-stage microscopy and b) melting point apparatus.

Given the absence of vt-PXRD data in literature, the following discussion will include all three samples under study. This data is particularly crucial given the tendency of VCV·HCl to form mixtures, hence making it more difficult to understand stand-alone experimental results. Only DSC peaks with an associated change in signal (DDSC) were taken into account. The DSC graph of form I in figure 4.39a, features the typical endotherms for the monohydrate and sesquihydrate forms, thereby confirming the presence of either hydrate.[209] As expected, the PXRD peaks associated with hydrates disappeared when heating over 100°C, with a mass loss of 1.63%. The crystal structure of form I retained its crystalline lattice over all the range of temperatures till 190°C, thereby in agreement with the melting point apparatus observations and the onset temperature for the melting peak in the DSC graph. The overlap between melting and decomposition is indicated by the negative gradient of the TG graph starting at around 200-205°C, parallel to the melting endotherm. This was also accompanied by an increase in the evolved species detected by the FTIR spectrometer. Furthermore, unlike the water of hydration, MS spectra showed that the chloride ion was detected only during decomposition, thus highlighting its integral presence within the lattice structure of VCV·HCl. DSC runs with higher heating rates displayed similar melting/decomposition trends but at higher temperatures due to the kinetic nature of decomposition (see appendix).

Multiple decomposition peaks imply either that the VCV·HCl molecule decomposed into several fragments which required different energy intakes or the presence of impurities (including other phases) which were not detectable through PXRD data. Conversely, published DSC graphs of pure form I only, display a typical single endotherm at 227°C and a single decomposition exotherm.[209, 298] This disagreement (particularly in melting point) suggested that the traces of hydrate present in the initial sample of form I or other impurities, might have impacted the stability of form I with temperature. Other influential factors concerned the experimental setup, where in contrast to those in literature, due to the analysis of the evolved species, the Al pan was pierced, and thereby introducing differences in the physical conditions. This factor surely affected the results, especially those related to the decomposition trends as material was pushed out of the pan during the process. Nonetheless, discrepancies in melting points are also present in published material, such as those associated with anhydrous form II whose established melting points vary from 211 - 223°C.[209-210, 214, 298]

The DSC graph of form V (figure 4.39c) illustrates two adjacent endotherms over the temperature ranges of 60 - 104°C and 120 - 137°C, respectively. These endotherms were accompanied by a decrease in mass of 6.28%, which falls within the recommended range of water content in sesquihydrate polymorph. The presence of two endotherms indicates that the dehydration of form V occurs in two stages, each representing a particular type of crystal water in the structure.[209]

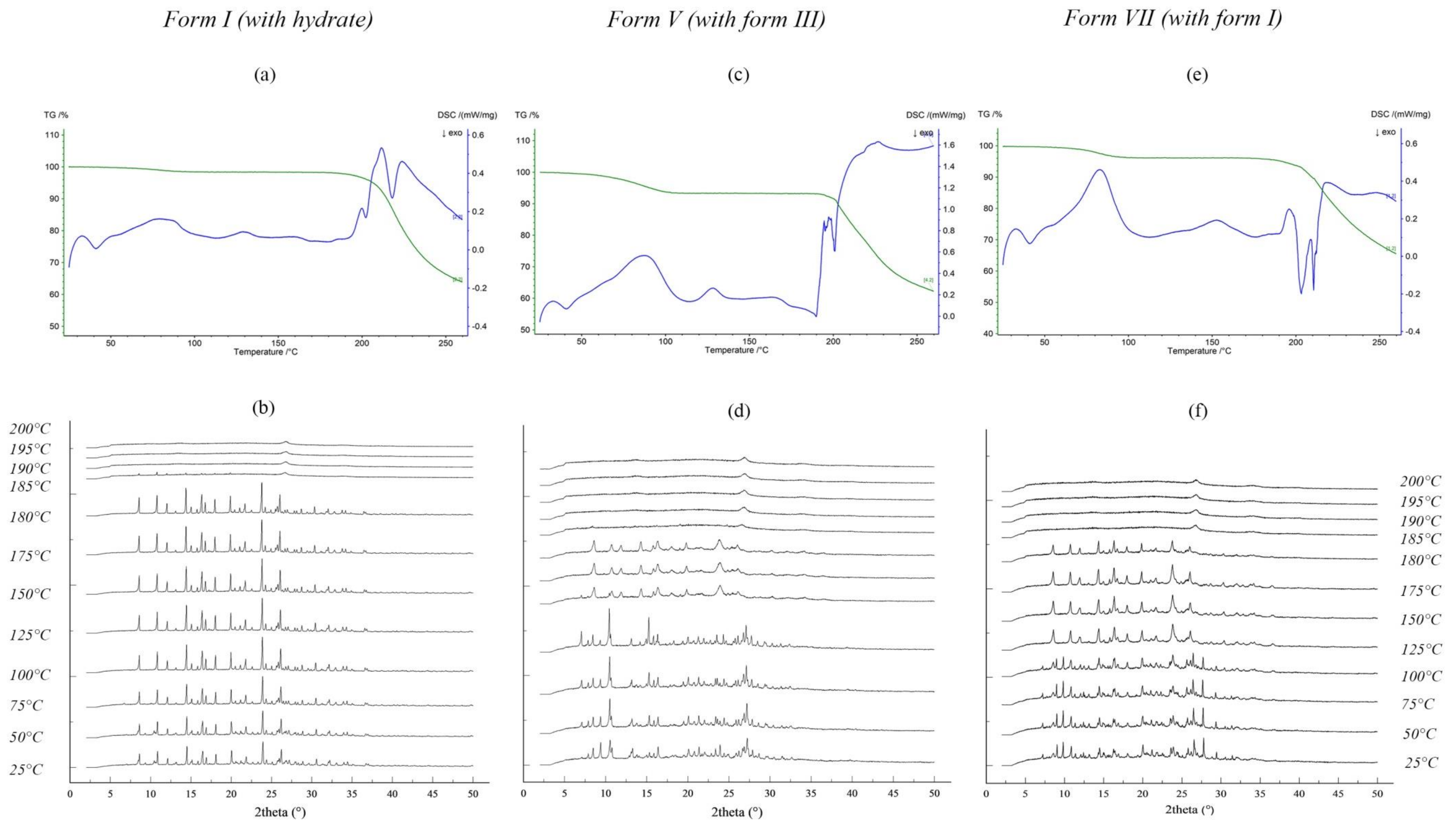


Figure 4.39: DSC/TGA overlay graphs and their respective vt-PXRD diagrams for samples with predominant content of form I (a,b), form V (c,d) and form VII (e,f). The temperature profile was set up to heat from 25°C till 260°C at a rate of 5°C /min, for both techniques (PXRD data collected at the same temperatures for all three samples). The first endotherm in each DSC graph is ignored because it is an effect due to unstable heating rate. The diffraction peak at 3.6° 2θ is missing from all vt-PXRD data due to instrumental purposes.

The corresponding PXRD data collected during the heating of form V in figure 4.39d, illustrate an interesting shift in the absolute intensities of peaks, with an increase in temperature. In particular, unique diffraction peaks of form V, such as 9.4°, 13.3°, 16.4° and 23.9° 2θ decreased in intensity, opposed to characteristic peaks of form III. These observations suggested that upon partial dehydration, sesquihydrate form transformed into the hemihydrate polymorph. However, due to the lack of a pure form V sample, it was difficult to determine whether the trace of form III in the initial sample influenced the route of phase transition.

The PXRD data (figure 4.39d) collected at 100°C show a mixture of hydrates, with hemihydrate (form III) as the predominant form, whereas the pattern at 125°C presents a sample of amorphous anhydrate with hemihydrate content. Subsequently, a further increase in temperature (150°C) led to complete dehydration, indicated through the absence of peaks such as those at 7.1°, 7.8°, 10.4°and 15.3° 2θ. This implies that the second endotherm represents the dehydration of the hemihydrate form into an anhydrate. The peak positions in the diffraction pattern associated to the amorphous anhydrate, are in line with those of form I. The low crystallinity might have caused a premature melting, with the disappearance of peaks at 180°C, relative to the melting endotherm in the DSC graph with an onset temperature of 189°C (peak at 197°C).

The DSC graph of form VII (figure 4.39e) illustrates an endotherm spanning over the temperature range of 60 - 105°C, coupled with a mass loss of 3.37%. This event proved that form VII is a hydrate, but it is distinct from form III, IV and V due to the absence of an endotherm at around 120 - 140°C.[209] The vt-PXRD run of form VII, displayed in figure 4.38f, was instrumental to confirm the presence of any other phases in the sample. In fact, on comparison, there were some peaks, associated with form I content, which were constant throughout the heating process, except for an increase in their intensity at temperatures over 100°C. Conversely, characteristic peaks of form VII, namely at 7.3°, 9.1°, 9.8°, 23.6°, 26.6°, 27.8°, 29.4°, were absent in the pattern collected at 125°C, in agreement with the endotherm in the DSC graph. The FTIR spectrum of this sample in figure 4.36 also hinted the presence of form I. These observations imply that form VII is a distinct hydrate which upon dehydration transforms into anhydrate form I. The DSC graph also features an endotherm over the temperature range of 135-166°C, however, no changes were observed in the PXRD patterns. Such endotherm might represent the melting of an impurity which was not detected by PXRD. The investigation of pure form

VII sample is required to explain better the effect represented by such peak in the DSC graph.

It is imperative to note that complete dehydration of hydrates with consistent increase in temperature, particularly form V and VII, led to the formation of form I with lower crystallinity and an onset melting temperature at around 189°C. The crystallinity of form I was much improved when a sample of form VII was heated for an hour at 100°C, suggesting that it requires a longer period of time to crystallise properly. These results demonstrate that the thermodynamic stability of form I is influenced by the purity of sample and the conditions of crystallisation. Similarly, form I retains its crystalline structure without comprise when exposed to 75% relative humidity at 30°C for a year, given the absence of impurities.[213] However, when exposed to relative humidity of 100% for a week, water intake induces polymorph transformation to a dihydrate.[214] Despite being sensitive to certain parameters, the outcomes support the thermodynamic stability of pure samples of form I, particularly relative to form II, due to having a thermal relationship with the discussed hydrates and a higher melting point (heat of fusion rule). Nonetheless, this experimental work is not exhaustive enough to exclude the possibility of forming new anhydrates, especially in untested conditions.

4.3.3.3 Crystal structure analysis

Given the absence of a structurally determined crystal structure, several computational techniques, such as Hirshfeld surface analysis and construction of lattice energy frameworks, could not be performed. However, it was still possible to infer certain information regarding the potential HB network and stability, which can be helpful during crystal structure solution.

4.3.3.3.1 Intramolecular level

The introduction of HCl led to the formation of a multicomponent system, with the possibility of forming a salt, depending on the nature of the functional groups on VCV molecule. The strongest pK_a value for basic functional groups in VCV is related to the primary amino group on the branched chain, with a value of 7.48. This evaluation was performed using the pK_a calculator plugin in *ChemAxon Marvin* online suite.[299] Consequently, the isoelectric point of VCV occurs at a pH of 8.82, suggesting that even in an aqueous environment (pH=7), the population of VCV molecules contain both neutral and ionic analogues, with a microspecies distribution ratio of 1:3, respectively (see appendix). A search in the CSD (last updated March 2022) was performed using *Conquest* to analyse trends concerning organic crystal structures (composed of C, H, N, O, Cl) which include at least one HCl/Cl$^-$ moiety. Results displayed the unchallenged dominant presence of a chloride ion relative to the associated HCl, as in the case of GCV·HCl. Given this information, it was reasonable to consider VCV·HCl as a salt, with NH_3^+ and Cl$^-$ groups.

The structure of VCV·HCl was geometrically optimised using the B3LYP functional and 6-311++G** Gaussian basis set, with dispersion correction, in a COSMO environment.[277] The resultant structure (figure 4.40) was intended for qualitative analysis, and not to represent a specific configuration in the solid-state. In particular, more intensive calculations are required to predict the exact location of the proton position between the amine group and chloride ion. Nonetheless, this conformation seems fairly reasonable because it has an extended backbone similar to those featured in the most thermodynamically stable structurally determined forms of GCV and FCV. Furthermore, an extended conformation would increase the accessibility of HB, which is known to provide stability in solid-state forms.

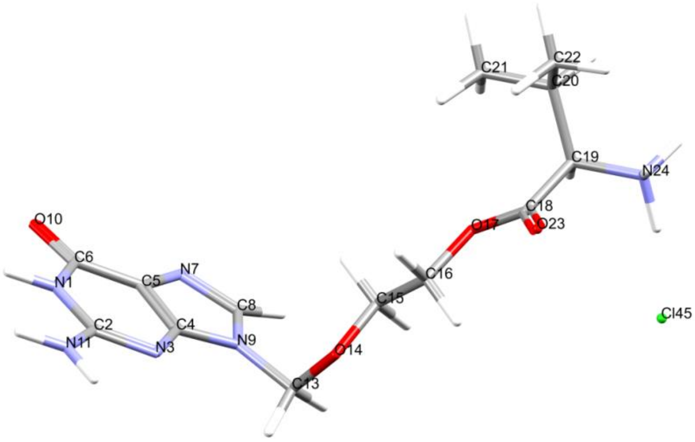


Figure 4.40: Molecular structure of VCV·HCl as a salt. Numbering system is valid for further reference.

Despite the effects caused by the presence of HCl, the flexibility of VCV·HCl can be still attributed to the long chain branch of VCV molecule with eight rotational bonds, whose individual extent of flexibility was described earlier. Such considerable allowance for rotation enables the molecule to exhibit a variation of folding modes, depending on its chemical environment. However, as noted in chapter three, the molecule also tends to produce different conformations with minimal adjustments along the end part of its branch, which is less restricted due to the lack of ether oxygen atoms. In fact, O17C18C19C20 and C18C19C20C21 torsion angles have the highest flexibility scores of 0.731 and 0.678, respectively. This end fragment opens the access to a significant conformational space, whereby conformations with minor differences can transform into each other without high energy barriers.

However, from the intermolecular analysis of GCV and FCV, it was concluded that even minor conformational discrepancies can cause disruptions within the HB network of a crystal structure. This is particularly valid for VCV·HCl, because the location of charged amino and chloride fragments is at the flexible end-part of the branch, thereby highly influential on the stability of the HB framework. The combination of these characteristics already hints the premise of having a substantial risk of polymorphism.

4.3.3.3.2 Intermolecular level

4.3.3.3.2.1 Exploring expectations

The expectations related to the intermolecular interactions of several fragments present in VCV·HCl were already discussed, since they were also present in FCV or GCV. However, *IsoStar* was employed to visualise the potential contacts using a charged amino and i-propyl as central groups. Same tolerance criteria and internal scaling settings were retained for comparison purposes.

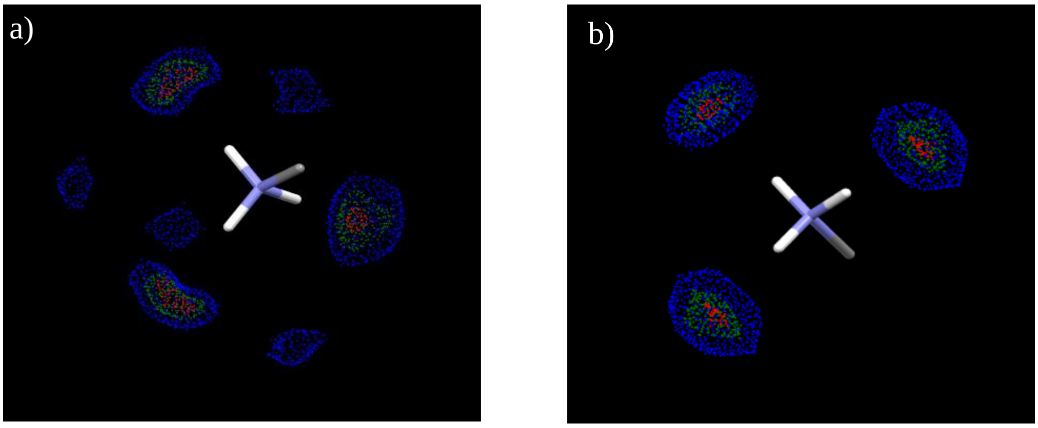


Figure 4.41: Scatterplot showing a charged amino as the central group, with N, O or S (a) and chloride ion (b) as contact groups. The contour plot is scaled internally with level 25 (blue), level 50 (green) and level 75 (red).

Given the strong HB donating ability of primary amine groups, it was also expected to observe similar traits for the charged amino group. Data points representing nitrogen, oxygen and sulfur (figure 4.41a) probes were distributed in multiple clusters around the charged amino group, not limited to areas directly in line with the hydrogen atoms. Conversely, the contour plot (figure 4.41b) featuring the chloride ion probe, involved only interactions with strong directionality, thereby exhibiting the superior HB accepting ability of such ion.

This characteristic was also highlighted through the scatterplots in figure 4.42, illustrating i-propyl as a central group. On comparison, contour plots associated with chloride ion were less dispersed than those representing nitrogen / oxygen atoms as contact groups, despite plotted at the same level (green). Data points at level 25 were removed from figure 4.42a because the majority of these hits represented cases whereby the relative orientation of the central and contact groups made it highly unlikely for them to form a HB. These scatterplots also exhibited the potential of C-H to contribute in the HB framework of VCV·HCl as a donor.

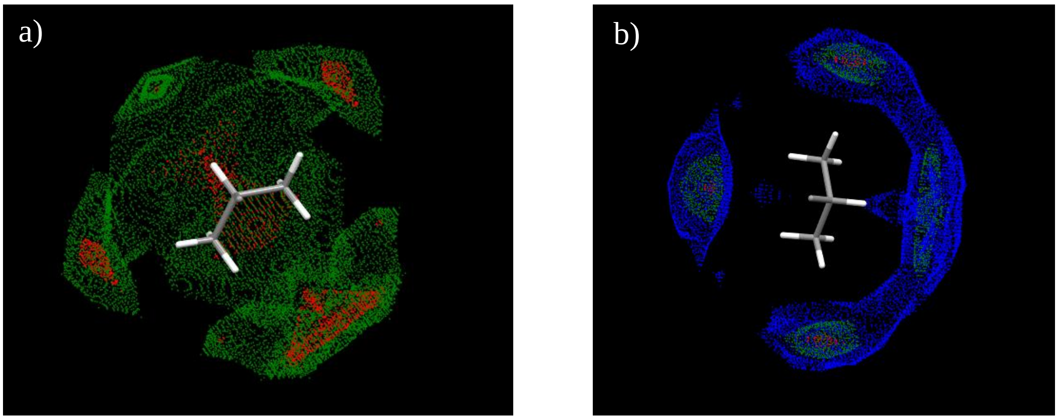


Figure 4.42: Scatterplot showing i-propyl as the central group, with N, O or S (a) and chloride ion (b) as contact groups. Scaling settings are identical to those in figure 4.41.

Despite of not being the exact conformation present in a solid form, the predicted conformer can still be regarded as a reasonable candidate. This implies that the main features of the full interaction map, shown in figure 4.43, can still be considered valid (to an extent) for the investigation of non-covalent bonding of VCV·HCl. This map illustrates how the most probable areas of interaction are located at opposite parts of the molecule, possibly allowing the branch to have some degree of freedom for rotation. This likelihood is supported by the presence of multiple polar groups around the ring system and the charged amino group, thus promoting the dominance of conventional contacts. On the other hand, relatively low propensity values (6.14, 6.77) characterised the hot spots located near the ester group, probably due to some hindrance in spatial access imposed by a comparatively large terminal fragment. Furthermore, the relatively small HB accepting ability of ether oxygen atoms was already referenced in the HB network analysis of FCV.

Similarly, one can notice the absence of HB acceptors involved in potential contacts with the two methyl groups, despite having exposed hydrogen atoms. This suggestion was unexpected, especially due to the inferences derived from the contour plots of the i-propyl group (figure 4.42). In addition, the HB donating ability of CH in a methyl group was also demonstrated in the full interaction maps for FCV. This difference highlights the sensitivity of the algorithm behind the full interaction maps towards each local chemical environment. Maps associated with CH probes in a methyl group or in an aromatic system, were only visible when plotted at a higher level. However, if spatial conditions permit, potential C-H···π interactions might be possible, as hinted by the corresponding orange hotspot.

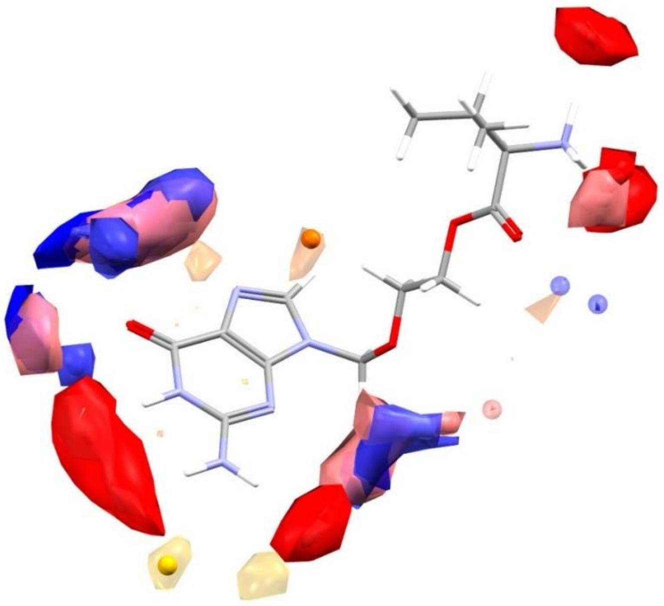


Figure 4.43: Full interaction map of VCV cation with the following probes: uncharged NH Nitrogen and RNH3 Nitrogen (dark blue: level 6), charged NH nitrogen (light blue: level 6), alcohol oxygen (light red: level 6), carbonyl oxygen (red: level 6), methyl carbon (yellow: level 2) and aromatic CH carbon (orange: level 2), together with the respective hotspots.

4.3.3.3.2.2 Chloride ion as HB acceptor

The list of organic crystal structures in the CSD was screened to locate hits containing a chloride ion, so as to comprehend the nature and extent of its potential contribution towards the HB network. Furthermore, this information can provide an indication of the relative position of VCV molecules with respect to each chloride ion, as well as be useful for co-crystallisation prediction. The HB definition was retained as default, given that no additional data was available to refine its geometric parameters.

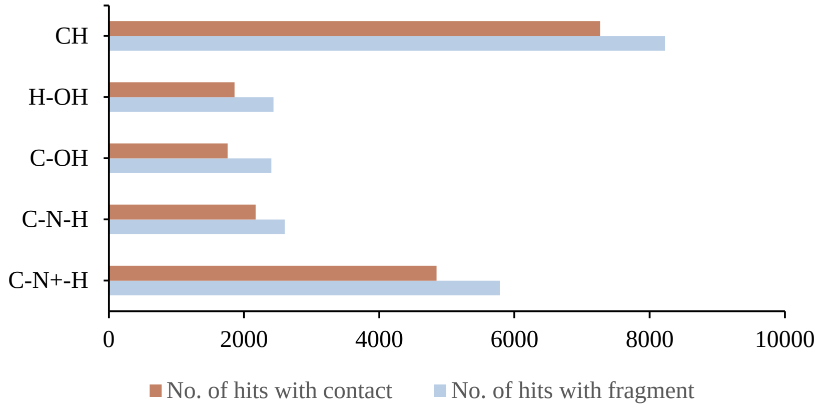


Figure 4.44: Bar chart showing the number of hits that included a chloride ion and a defined HB donor, as well as the dimension of their respective subset of hits which involved an interaction between such HB pair.

Table 4.5: Mean values for the geometric parameters of specific HB contacts involving a chloride ion.

HB donor			**Distance (Å)**		**Angle (°)**	
	No. of hits	No. of contacts	Mean	Standard dev.	Mean	Standard dev.
C-N$^+$-H	5783	10378	2.186	0.146	160.630	12.390
C-NH	2600	4864	2.314	0.167	157.691	12.596
C-OH	2404	2608	2.119	0.116	164.770	10.619
H-OH	2436	4363	2.219	0.114	165.549	10.907
CH	8229	31037	2.747	0.138	148.883	13.823

Firstly, it was noticed that every chloride ion was a strong contributor with respect to HB network, very often having a HB coordination number of three or four. The bar chart in figure 4.44 illustrates the number of organic hits in the CSD with chloride ion and another specific donor. In addition, the geometric parameters of the interactions per specific HB pair were collected for comparison purposes. The search for C-NH fragments included any type of covalent bonding between the carbon and nitrogen atoms. Almost 70% of the structures involving a chloride ion, also included a positively charged amino group, of which 83.8% had at least one HB between the two moieties. Results demonstrated that most of the crystal structures which contained both neutral and positively charged NH groups, also included a HB between the chloride ion and each group respectively. Nonetheless, on average, the interactions with the positively charged amino group tended to have a stronger degree of directionality than with their neutral counterparts (table 4.5). These figures suggest that each chloride ion in VCV·HCl is likely to form interactions between molecules in a linear manner, given that neutral and positively charged NH fragments are situated at opposite ends of VCV molecule.

Similarly, a high percentage (>78%) of crystal structures with any other conventional HB donor, such as a hydroxyl group, also involved contacts with the chloride ion. On comparison, this type of HB donor (OH in water or bonded to a carbon) formed the strongest interactions. This was indicated through the mean values of their corresponding distance and angle measurements. The accepting ability of this halide ion also extends to interactions with carbon donor, with more than 88% of the hits involving both fragments also having a HB contact. As expected, these contacts were characterised by the lowest directionality, as also displayed in the stacked histograms below.

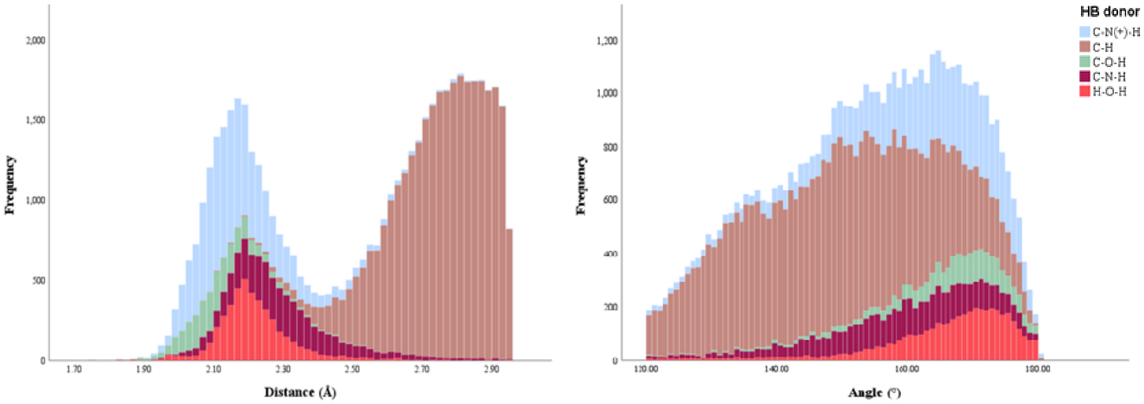


Figure 4.45: A stacked histogram showing the distribution of the distance (left) and angle (right) measurements of the interactions involving a chloride ion and specific HB donor.

4.3.3.4 Risk assessment of polymorphism

The solubility of VCV·HCl resulted to be limited to some polar solvents, particularly alcohols and water. Despite using heat, there were several trials where the solute did not dissolve completely, such as in case of 1-butanol, acetone and dichloromethane. Binary solvent mixtures contained a minimal amount of water in order to aid the dissolution of VCV·HCl in the presence of another solvent. As noted earlier, the majority of the products from solubility screening resulted in mixtures, some of which were characterised with very low crystallinity. The presence of sesquihydrate form and anhydrous form I was the most common amongst the resultant mixtures.

The thermal analysis of the produced hydrates supported the thermodynamic stability of form I, given its formation after the dehydration of the respective hydrates. Such characteristic was also indicated by its presence in various mixtures, unlike anhydrate form II which was not obtained in any trial. Moreover, the exploratory STA profile for form I, with successive heating and cooling cycles, did not result in new polymorphs. However, experimental work also highlighted how the stability of form I is dependent on other factors, such as the purity of the sample. Furthermore, patents show the existence of forms of VCV·HCl with very low water content (<2%), which might be considered as anhydrates, thus requiring further investigation.[211] In conclusion, the risk of forming other polymorphs is still relevant in case of anhydrous VCV·HCl.

4.3.3.4.1 HB propensity model

The advantage of this technique is that any system, even with multiple components, can be a suitable input, as long as each contributing fragment is well represented in the search. VCV·HCl was dissected into eight sub-structures, including the chloride ion. For similar reasons as earlier, the default HB definition was retained and carbon was included as a HB donor. The search was restricted to only organic structures with no errors or disorder. A training set of 5258 hits was utilised, while ensuring that every sub-structure was adequately represented by a sufficient amount of entries. The area under the ROC curve parameter (0.90) indicated that the final model had an excellent discrimination ability. Additional variables as steric density and competition were retained.

The logit model generated a propensity value for all possible interactions, ranging from 0.96 to 0.01. The most probable interactions involved the chloride ion as an acceptor with both neutral and positively charged amino group as donors, with a propensity value of 0.96 and 0.93, respectively. Such propensity was already hinted through the intermolecular analysis done earlier. The HBC table showed how the likelihood of the chloride ion to have three or four contacts simultaneously was almost equivalent (~0.4). The carbonyl oxygen atoms were also associated with interactions of significant probabilities, particularly the one located on the ring system. However, the oxygen atoms in the ester group had higher propensities for the formation of intramolecular bonds with the neighbouring CH donors, due to their location along the chain.

After obtaining a list of interactions, the algorithm generated 98 potential polymorphs, and ranked each according to their mean HB propensity and HBC, as displayed in figure 4.46. The absolute majority of the hypothetical structures were composed of six to seven intermolecular HB pairs, all of which involved multiples ones with the chloride ion. One should note that the propensity related to each possible HBC number is influenced by the molecular conformation of the inputted system. This implies that the ranking based on this parameter might be more prone to inaccuracies given that the inputted structure was not extracted from a structurally determined crystal structure.

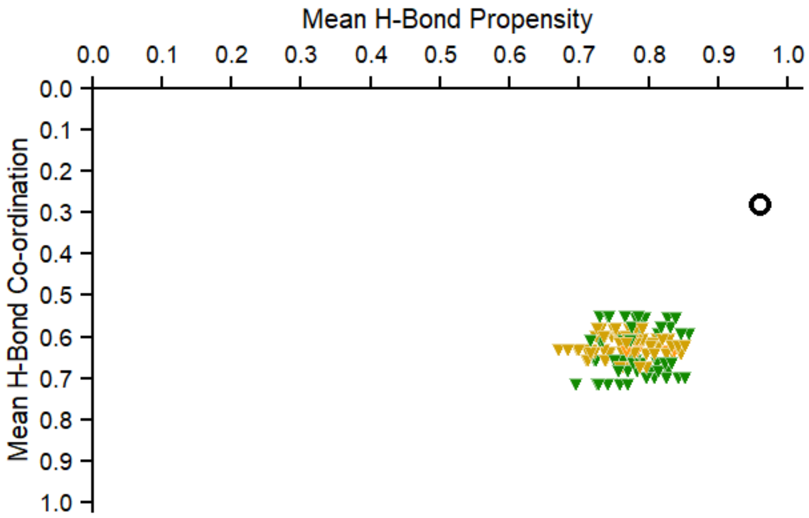


Figure 4.46: HB propensity landscape for VCV·HCl, with potential polymorphs having either six (green) or seven (mustard) intermolecular HB pairs. The circle represents the inputted system but it can be ignored as it was not part of a unit cell.

The presence of a fairly balanced ratio of HB acceptors and donors, coupled with HB participants with high HB coordination number and considerable molecular flexibility, increased the number of distinct potential compositions which were able to construct a stable HB network in VCV·HCl. This effect is reflected by the numerous potential forms ranked in the optimal section of the propensity chart. On comparison, some HB frameworks only differed to a minimal extent, but such variation is sufficient to exhibit polymorphism. Similar HB frameworks also imply that forms might transform into others without high energy barriers. Furthermore, this characteristic is even more highlighted in the presence of water due to the addition of strong HB participants, as proved through the various hydrates described in literature.

4.4 Performance of polymorph prediction methods

Polymorph prediction for GCV was conducted using two approaches of different nature; HBP models are based on the statistical analysis of geometric parameters of solved structures, whereas CSP is based on the energy evaluation of generated crystal structures. The appropriateness of these methods for polymorph prediction can be determined on two factors: (i) the estimation of the risk of forming other forms experimentally and (ii) the classification of the extent of stability of known forms. Both methods predicted correctly the feasibility of generating form II experimentally, relative to other hypothetical forms, in agreement with lab work results. However, they both failed to represent the crystal structure of form I appropriately and to capture its thermodynamic stability. The inability of HBP model method to account for enantiotropic relationship between polymorphs, as the one present between form I and II, was also noted in literature.[300]

The risk of polymorphism for GCV based on the HBP model was overestimated due to the presence of other hypothetical polymorphs in the optimal area of the chart, in close vicinity of the datapoint of form II. Conversely, such risk was more appropriately represented by the crystal energy landscape, whereby the significant energy gap between form II and most of the potential polymorphs illustrated a decrease in the probability of forming multiple other forms. Therefore, the CSP procedure satisfied both criteria for appropriateness, with respect to form II. Nonetheless, these results would have much greater validation if form I was generated and ranked during the procedure.

Given the less intensive procedure of HBP models, it was possible to employ it for all three antivirals and compare the outcomes. Even though FCV has a higher flexibility score, its propensity landscape was much less populated relative to the one for GCV. This occurred because, despite including CH donors, the HBP model is still highly influenced by the conventional HB participants. This implies that due to a more balanced number of HB donors and acceptors, the model was able to generate a larger number of potential polymorphs with unique HB frameworks for GCV. Similar effect can be noted for

VCV·HCl, due to its diverse functional groups and their ability to have higher HB coordination numbers. However, a large amount of generated polymorphs does not equate to a high risk of accessing such pool experimentally due to energetic reasons. The incorrect ranking of FCV and GCV form I expose the inefficiency of the HBP model system to account for factors which contribute towards the stability of a crystal structure, beyond traditional HB.

4.5 Outcomes and further work

Surely, the tendency for polymorphism does not solely depend on the molecular flexibility. A high flexibility score will probably enlarge the number of hypothetical polymorphs, but it does not reflect the actual amount of forms which are energetically feasible. These three case studies show how polymorphism is a direct consequence shaped by the extent of stability provided by all contributors within the intermolecular framework, particularly the HB donors to acceptors ratio, HB coordination number per group and the strength of other non-bonding contacts, beyond HB. The risk of polymorphism is imminent in cases where either an improved or equivalently stable intermolecular network is achievable, without considerable energy requirements. This case applies in particular for VCV·HCl.

Crystal structure analysis and experimental work conducted on FCV and GCV, coupled with literature, led to the conclusion that the administered form is the most thermodynamically stable form at room temperature, with minimal risk of polymorphism with respect to other anhydrates. Finer energy calculations are required to evaluate the actual feasibility of the hypothetical form ranked as the global minimum in the CSP of GCV. Even though it is sensitive towards certain conditions, such as sample purity and process of crystallisation, the thermodynamic stability of VCV·HCl form I proved to be considerable, particularly relative to form II. Nonetheless, the risk of polymorphism of VCV·HCl is still relevant, and hence requires further analysis to be fully understood and defined. Further work should also include the investigation of the kinetic stability of the administered forms, particularly in the presence of high relative humidity conditions, beyond the information present in literature. This is an important aspect which can lead to polymorphism, particularly during the production of the drug and storage periods.

Chapter 5

CO-CRYSTALLISATION

5.1 Introduction

The use of co-crystals in the pharmaceutical industry is gradually gaining momentum due to possibility of modifying the solid state properties of the active ingredient, in attempt to enhance its bioavailability, while still conserving its biological activity.[60] Given the extensiveness of the experimental process of co-crystal screening, computational techniques were devised to offer assistance by predicting the favourability of potential co-crystals. Due to the nature of the procedure, computational screening for co-crystallisation can cover a much greater variety of coformers, hence increasing the probability of success. The available computational methods are mainly constructed to either provide indications for coformer selection or to quantify / rank the potential successfulness of co-crystallisation. Such predictive techniques are based on theoretical methods such as HB energy calculations or on structure informatics obtained from known determined crystal forms, as noted in the literature review.[301] In several case studies, the results of distinct approaches were either integrated together or accompanied by outcomes of other prediction methods.[155, 175, 177, 179-180] It was also noticeable how their success rates were influenced by the size, chemical nature and extent of flexibility of the molecules under study. In fact, significantly reliable outcomes were very often only associated to relatively smaller, rigid, and neutral molecules.[177, 179, 302]

Given that the three antivirals deviated from such limited molecular domain to varying degrees, it was essential to investigate possible routes that can potentially improve the prediction abilities of standard approaches. These routes focused particularly on a selection of inputted molecular conformations, which are known to influence the quality of predictions. Despite their influence, so far, relatively minimal importance was given to molecular conformations in published computational studies related to co-crystallisation. Further modifications revolved around the inclusion of carbon as HB donor within specific procedures. The HBC approach was not employed because its performance resulted to be the least reliable in literature, even after attempted improvements.

5.2 Methodology

Similar to the previous chapter, the following methodology is composed of a combination of diverse approaches for co-crystal prediction, coupled up with experimental work for validation purposes. The selection of 17 coformers was based on the HB motif search results and molecular complementarity indications obtained in earlier work (see appendix).[133] It was demonstrated how the carboxylic group can outcompete the homo-interactions by a sufficient margin. The carboxylic group has an additional advantage because it is composed of two potential HB acceptors and a HB donor, thereby it can theoretically be a good candidate to form contacts with both the carbonyl oxygen and the N-H groups of the antivirals. The considerable benefits of having an aromatic nitrogen atom as well as a hydroxyl group were also highlighted. Furthermore it was ensured that all coformers were present in the GRAS list, because any hypothetical co-crystal was ultimately intended for human consumption.

5.2.1 Experimental method

Liquid assisted grinding (LAG) was the chosen route for co-crystallisation, using between one to three drops of solvent with 40 mg of API (commercially available sample). It was ensured that the manual grinding process amongst trials was performed coherently as much as possible, within a limited timeframe. All coformers were tested at different stoichiometries, but due to the complexity of computational methods, only outcomes of 1:1 molar ratio will be discussed. In case of GCV and FCV, trials were performed twice, using different solvents for comparison purposes. The synthesis and characterisation of GCV co-crystals were conducted by another research group member.

The resultant product of each trial was characterised using PXRD as well as FTIR spectroscopy. The details of these techniques can be found in section 4.2.1.2.

5.2.2 Computational methods

The aim of this section was to apply different approaches towards the attempt of predicting correctly the degree of favourability of hypothetical co-crystal formation. Given that there are no co-crystals of GCV, FCV and VCV·HCl published in the CSD, it was crucial to take into account multiple probable scenarios, with no assumptions. As opposed to the other APIs, the co-crystallisation prediction for VCV·HCl was restricted to only one approach (HBP models), due to the additional layer of complexity present in dealing with a multicomponent system prior to co-crystallisation.

5.2.2.1 COSMO-RS approach

Given that the method for the experimental screening for co-crystallisation employed LAG, with the involvement of solvents, it was of crucial importance to include the molecular conformation in solution as part of the factors being tested to potentially improve predictions. Since most probably, such dominant molecular conformation in solution would be different than the ones present at gas-phase or in the solid-state, calculations were performed in a COSMO environment. For all computations, the temperature was set to be 298.15 K, in attempt to replicate ambient conditions.

5.2.2.1.1 Selection of a set of conformers in solution – ReSCoSS

The process of selecting a set of relevant conformations in solution was based on the workflow called Relevant Solution Conformer Sampling and Selection (ReSCoSS), published by Udvarhelyi *et al.*.[303] The general steps of the workflow were followed, as presented in figure 5.1, with some adjustments based on the tools available. The first step involved conformer generation with diverse tools, and given that it was shown how certain traditional force-fields as UFF performed poorly for conformer searching and ranking, RDKit ETKDG v2 stand-alone algorithm (Conda environment) and CSD conformer generator were selected.[304-308] The input format was as a 2D molecule, and both generators produced 500 conformers each (see appendix for more details).

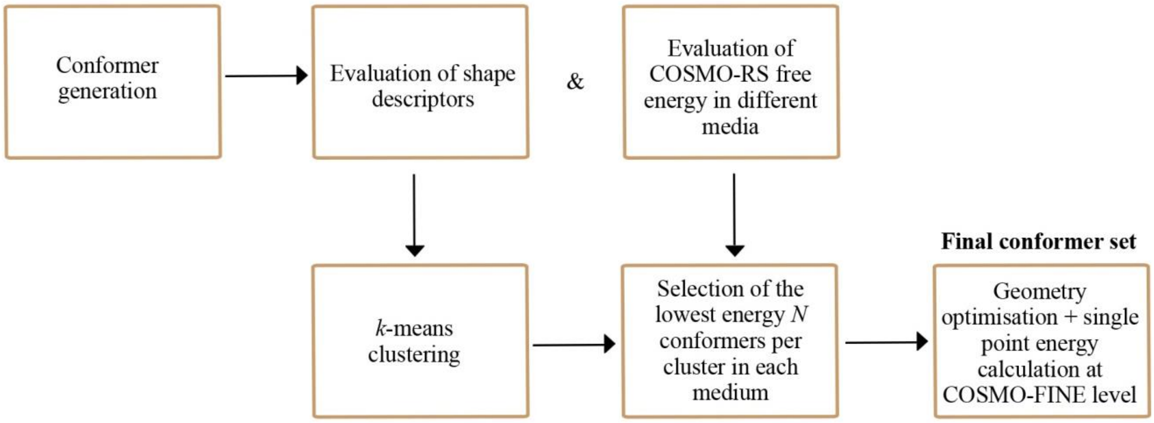


Figure 5.1: General steps in the ReSCoSS framework for the generation of conformer sets.[303]

The set of 1000 conformers was inputted in *COSMOconf* V4.2, so as to be assessed and reduced only to a minor extent (loose criteria). The process involved an initial geometry optimisation using the semi-empirical method AM1 in a COSMO environment, followed by single-point energy calculations at BP86/def2-TZVPD/COSMO (FINE) level of theory, the latter being the highest level of parameterisation available in *COSMOtherm* V.18.02 (see appendix).[309] The COSMO-FINE level uses a refined cavity construction algorithm which provides an improved system compared to the standard COSMO cavity, particularly when dealing with concave molecular shapes.[310]

Before processing the COSMO files in *COSMOtherm*, some of the settings were adjusted according to the chemical nature of the molecules under study and the tools applied.[143] These changes included the authorization to allow the addition of VdW and HB contributions to the chemical potential. Furthermore, there was also the acceptance of the integration of the COSMO-RS combinatorial term, which is recommended to be switched off only in case of polymers.[309] Default settings were retained with respect to the options for the σ-profile composition.

The processing in *COSMOtherm* was required to obtain four descriptors derived from the σ-surface for each individual conformer (not as a set), namely the dipole moment, the σ-total surface area, the σ-HB donor moment and σ-HB acceptor moment.[303] Moreover, it was also required to integrate the use of solution energies within the process of conformer selection. This was done by calculating COSMO-RS free energy of each conformer, at the same parameterisation level, in solvents covering a wide range of polarity index and HB capabilities: vacuum, cyclohexane, acetonitrile, acetone, chloroform, methanol, 1-octanol, dimethyl sulfoxide and water. The corresponding gas phase energy file of each conformer was allowed to be automatically estimated by *COSMOtherm,* rather than performing a stand-alone geometry optimisation in gas-phase as this led to no improvement in the final results.

After collecting all information, clustering based on the four descriptors using k-means was performed, so as to be able to partition the conformers in every solvent into *k* groups. The free energy variable was then utilised to sort the list of conformers in each cluster and the first *N* conformers from each cluster were selected. This procedure was performed for all solvents, and any duplicate conformers were omitted from the list. For example, in case of *k1N2* set, it would contain a maximum number of 18 conformers, given that nine different media were utilised. Different combinations of *k* and *N* parameters were tried, with balanced ones as *k3N3* being the most recommended in the literature.[303]

The selected conformers were refined through a geometry optimisation using the B97-3c composite scheme in conjunction to def2-mTZVP basis set, in a COSMO environment, using *Turbomole*. This DFT method is known to provide sufficiently accurate molecular geometries, at a relatively low computational cost when compared to other hybrid functionals.[311] The geometry optimisation was followed by a BP86-D3(BJ)/def2-TZVPD/COSMO-FINE single point energy calculation, enabling the resultant conformer COSMO files to be adequate to input them in *COSMOtherm* for prediction.[230, 235, 238, 276]

In order to examine the geometric variability within the selected sets, RMSD values were evaluated using *CSD Python API,* with the exclusion of hydrogen atoms. In case of having conformers with RMSD < 0.1, only the one with the most negative free energy was retained. The exclusion of such highly geometrically similar conformers only took place if they had negligible Boltzmann’s weight factors. The remaining conformers within the set, composed the final group of conformers, relevant to the conformation of each antiviral in solution. These final sets of conformers (one set for each antiviral) were analysed, compared to conformations in known solid-states and utilised to perform solubility predictions in attempt to compare the values with experimental results (see appendix for more details).

5.2.2.1.2 Co-crystal prediction

An automatic conformer search was performed for each coformer used in order to ensure that all relevant solution conformations were taken into account. This was conducted through *COSMOtherm* interface, at BP86-D3(BJ)/def2-TZVPD/COSMO-FINE level. The chosen *kN* set of conformers for the antiviral was set as target (API) and the calculation for co-crystal screening was initiated after adjusting the stoichiometry to 1:1.

In order to analyse the dependence of the results on the solvent used, all the conformers produced at a BP86-D3(BJ)/def2-TZVPD/COSMO-FINE level (irrelative of the set), were inputted in a mixture with the solvent used for LAG co-crystallisation. The ten conformers with the highest Boltzmanns's weight were also treated as a set and used for prediction. Lastly, the procedure was repeated using only one API extended molecular conformation (to represent the molecular conformation in the solid-state crystal structure), for comparison (summary in figure 5.2).

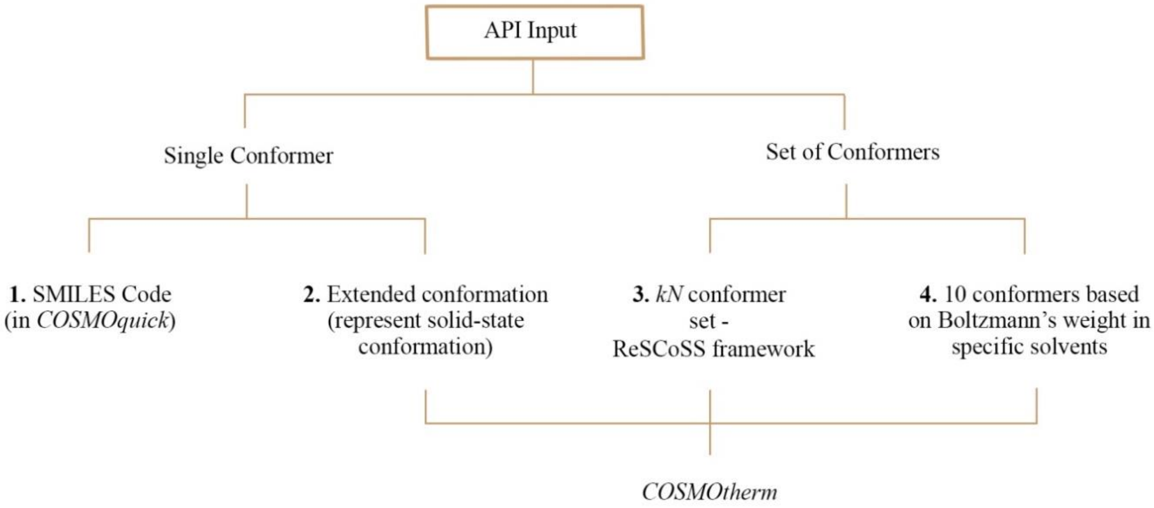


Figure 5.2: Different API inputs for co-crystallisation prediction in *COSMOtherm* and *COSMOquick*.

The results of the *COSMOtherm* co-crystal screening provide the enthalpy of mixing or excess enthalpy, H_{ex} (eq. 5.1), the free energy of mixing, G_{ex} (eq.5.2) and the contribution of HB in enthalpy of each subcooled co-crystal liquid, A_jB_k. As opposed to *COSMOquick*, only these thermodynamic properties are utilised for the evaluation of the theoretic co-crystallisations.

$$H_{ex} = x_j H_{AB}^{A} + x_k H_{AB}^{B} - x_j H_A - x_k H_B \tag{5.1}$$

$$G_{ex} = x_j \mu_{AB}^{A} + x_k \mu_{AB}^{B} - x_j \mu_A - x_k \mu_b + RT\left(x_j ln x_j + x_k ln x_k\right) \tag{5.2}$$

$$x_j = \frac{j}{j+k} \qquad\qquad x_k = \frac{k}{j+k} \tag{5.3}$$

$$f_{fit} \sim H_{ex} + a(\max(1, nrot_{API}) + \max\left(1, nrot_{co-former}\right)) \qquad \text{a=0.5102} \tag{5.4}$$

H_{AB}^{A} and H_A represent the enthalpy of component *A* in the mixture (j:k) and in the pure state, respectively, and similarly applies for component *B*. Furthermore, the partial empirical function, f_{fit} (eq. 5.4) which is the default ranking scheme in *COSMOquick*, was also used to rank the propensity of co-crystal formation. This function takes into account the number of rotatable bonds present in both API and coformer, in attempt to compensate for the lack of kinetic effects considerations. The compatibility between each antiviral - co-former pair was also analysed through the prediction of the Hansen solubility parameters based on the QSPR approach, and the evaluation of the Hansen distance.[133, 312]

5.2.2.2 Molecular Complementarity approach

Molecular Complementarity was conducted in *Mercury*, using the principles outlined by Fábián with respect to co-crystal screening, based on five molecular descriptors.[178] Given that such properties are dependent on the molecular conformation, different combinations of API and coformer conformers were tested (table 5.1). Dominant conformations in solution were chosen from those produced at a COSMO environment (section 5.2.2.1.1). This was done by selecting the conformer of both antiviral and coformers which had the highest Boltzmann’s weight factor in specific solvents. These solvents were chosen based on what was used during the experimental LAG co-crystal screening. Furthermore, the procedures were repeated using API conformations produced earlier (ReSCoSS) which were the most similar to the extended molecular conformations present in known determined anhydrous forms (GCV: UGIVAI01 and FCV: IMADEI). Other trials using solid-state molecular conformations for coformers (selected based on the PXRD data collected from the pure compound used in experimental screening) were also assessed in conjugation to different API conformations. A percentage pass rate for each specific active ingredient and coformer pair was generated.

Table 5.1: Combination of inputted conformers per route per each hypothetical co-crystal. The extended conformer for API was chosen from the list of conformers generated in the ReSCoSS procedure, and it resembled its solid-state molecular conformation. The coformer solid-state conformation was extracted from its corresponding crystal structure from the CSD.

	Different combinations of inputted molecular conformations			
Route	1	2	3	4
API	Dominant conformation in solvent	Extended conformation	Extended conformation	Dominant conformation in solvent
Coformer	Dominant conformation in solvent	Dominant conformation in solvent	Solid-state conformation	Solid-state conformation

5.2.2.3 Hydrogen bond energies approach

This approach was based on the methodology introduced by Musumeci and Hunter, whereby they proposed a hierarchal method of determining HB pairs, using extreme points on molecular electrostatic potential (MEP) maps on gas-phased conformations for identification.[171, 313] This procedure required a wavefunction file for each API and coformer, so as to be able to construct the MEP maps. Similar to the previous approach, different input molecular conformations were tested (table 5.1). These conformers included those dominant in particular solvents for both co-crystal counterparts, as explained in section 5.2.2.2, solid-state extended conformer representatives for the antivirals and solid-state molecular conformations for coformers. The conformers of coformers were extracted from the unit cell of their geometry optimised crystal structures, following the procedure suggested by Khalaji *et al* for similar chemical contexts and objectives.[180] The unit cell of each coformer was optimised through *CASTEP* in *Material Studio*, using PBE functional with Grimme dispersion correction, under periodic boundary conditions at ultra-fine level of criteria.[232]

Turbomole was used for the production of wavefunction files per conformer, but some of them failed normalisation checks, probably due to the presence of molecular orbitals which were not sufficiently orthonormal.[314] Therefore, another route was chosen, starting from the input of each molecule in *Multiwfn* so as to produce an appropriate input file for *ORCA* V.5.0.[247, 315-316] A B3LYP-D3(BJ)/6-311++G** single point calculation was carried out for each molecular conformation, while retaining default settings (see appendix for more details). These settings were employed because they were in line with the parameters used for the construction of eq. *5.5* and *5.6*, thus ensuring appropriate applicability.[171] The output molden file from *ORCA* was inputted in *Multiwfn* to produce the required wavefunction file, which was then processed for normalisation checks. Each valid wavefunction was used in *Multiwfn* to construct the MEP surface (0.002 e/au isosurface), yielding interaction energies (kcal/mol) between a positive point probe and the surface of the molecule at the site of interaction.[317]

The list of potential HB contributors was compared with the list of extreme points on the surface, choosing only local maxima and minima that were directly located upon such contributors. In case of having acceptors with no direct extreme point, the minimum energy value associated with that particular atom was taken into account. Extreme points related to either no specific atom or to a potential HB contributor which was unavailable to form intermolecular contacts, were omitted from the list.

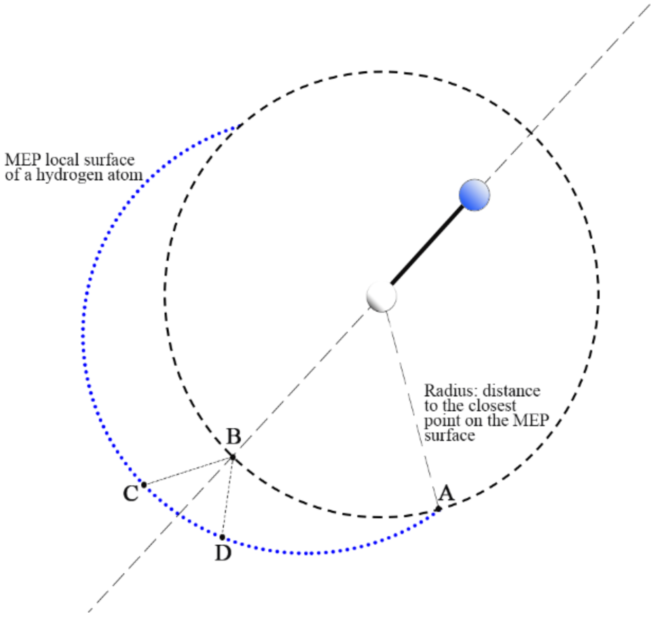


Figure 5.3: A graphic illustration of the method used to locate points on the MEP surface for HB donors.

In order to take into account O-H or N-H groups which had missing or non-direct maxima (see appendix for example), the following procedure was applied. Firstly, a sphere was constructed using a radius defined by the distance from the hydrogen atom and its closest point of the MEP surface, *A*. The coordinates of the HB donor and hydrogen atom were used to create a line, and the point of intersection, *B* between such line and the sphere was located, as shown in figure 5.3. This was done in order to locate a favourable location where a contact might take place (region *CD*), taking into account the relatively high degree of directionality typically present in O-H and N-H contacts. The point on the surface closest to the intersection point was identified and its mapped value was taken as energy associated with that hydrogen atom. This method was employed because taking the maximum energy point on the local surface of a hydrogen atom did not always lie within the area of creating possible non-covalent contacts. Some important C-H groups with no direct maxima, such as in case of GCV and FCV, where the known structurally determined forms contain HB with such contributors, were also processed using this method.

$$\alpha_p = 0.0000162MEP_{max}^2 + 0.00962MEP_{max} \tag{5.5}$$

$$\beta_q = 0.000146MEP_{min}^2 - 0.00930MEP_{min} \tag{5.6}$$

$$E = -\sum_{pq} \alpha_p \beta_q \tag{5.7}$$

$$\Delta E = E_{CC} - jE_{API} - kE_{COF} \tag{5.8}$$

The resultant molecular electrostatic potentials corresponding to the maxima, MEP_{max} and minima, MEP_{min} sites on the MEP maps were converted into their associated parameters, α_p and β_q.The free interaction energy, *E* is given in equation *5.7,* with the pairing of α_p and β_q performed in a hierarchal manner as suggested by Musumeci *et al.*[171] Furthermore, the propensity of a co-crystal formation was given in terms of the energy difference between the total interaction energy of a hypothetical co-crystal, E_{cc} and its pure constituents, E_{API}, E_{COF} (eq. *5.8).*

5.2.2.4 HBP model approach

HBP models were applied to predict the likelihood of all the possible interactions between each API and coformer components. The construction of each HBP model followed the method described in section 4.2.2.2.1, while retaining default settings for the HB definition due to the lack of additional information. Each molecule was dissected into representative sub-structures, all of which had to have sufficient amount of hits in order to proceed with the construction of the model. Furthermore, an adequate training dataset (over 2000 hits) and a satisfactory value for the area under the ROC curve were required to reduce the chance of producing inappropriate models. Two HBP models were constructed for each API – coformer pair to analyse the effect of including carbon as a HB donor.

5.2.2.4.1 Hierarchal approach

This approach was primarily introduced for hydrate formation prediction by Tilbury *et al.*, where they applied the principles of the hierarchal approach by Musumeci, but using the results of the HBP models.[171, 181] However, it was required to modify the procedure so as to cater for the complexity of having diverse lists of functional groups with different availabilities and HBC numbers. Whilst the construction of HBP models was performed in *Mercury,* the algorithm for the selection of interactions was conducted in *Python* (see appendix).[169, 243] After obtaining the list of interactions with their respective propensity values, this list was divided into two groups: pure and impure contacts. Each sub-list was analysed and the following refinements were implemented.

5.2.2.4.1.1 Selection of available HB participants

The reliability of a hierarchal approach relies on the selection of interactions present within the list, which in return is based on the HB participants available within both potential co-crystal components. Different routes were established based on diverse ways which can provide suggestions concerning such availability. The first route relied on the list of intramolecular bonding interactions produced from the model itself, with a propensity value larger than 0.5. It was ensured that all chosen intramolecular contacts were geometrically feasible. In case of having multiple highly probable intramolecular interactions, particularly when including CH as a HB participant, only the highest probable contact was taken into account. Only functional groups which participated in the selected intramolecular interactions and had no other donating or accepting abilities, were eliminated from both impure and pure contact lists. The HBC table was utilised to determine whether such groups were likely to still contribute in intermolecular bonding (contain an additional hydrogen or lone pair of electrons).

The rest of the routes were based on different combinations of API and coformer molecular conformations, similarly to previous approaches (table 5.1). The same HBP model was applicable for these routes because the algorithm is not dependent on the molecular conformation, hence producing the same list of propensities. However, in these particular routes, the elimination of HB contributors was based on their lack of availability in the present molecular conformation under study, rather than on the list of intramolecular contacts produced by the model. This implied that any HB group which was unavailable due to an intramolecular bond or steric purposes in the present conformation, was eliminated from both impure and pure contact lists.

5.2.2.4.1.2 HBC number

As opposed to the HBP model itself, the HBC propensities are affected by the inputted molecular conformations. In general, only minor changes in probabilities were observed, unless the molecular conformers differed significantly (ex. an intramolecular bond). The HBC table was produced for every inputted molecular combination and analysed to identify which HB contributors can participate more than once. By default, functional groups were assumed to participate once in HB, unless they were unavailable as described in the previous section. In these cases, their corresponding most probable interaction was chosen in each list, as per hierarchal policy, while eliminating all the other contacts involving them.

However, there were certain groups whose HBC number was likely to be more than one. For example, in case of a primary amine group with available hydrogen atoms, it was ensured that its two contacts with the highest propensity values were retained in the final lists of pure and impure interactions, respectively. The HBC table was particularly necessary for carbonyl oxygen atom because its most favourable HBC number was fairly sensitive with respect to changes in conformation. A functional group was also allowed to be present more than once if it could act both as a donor and acceptor.

5.2.2.4.1.3 Scoring function

After such process of refinement and elimination, each resultant list of pure and impure contacts was analysed. A multi-differential (MD) score was tailored to evaluate the extent of competition between pure (I_A and I_C) and impure (I_{AC}) interactions, thus ultimately reflecting the likelihood of co-crystal formation (figure 5.4).

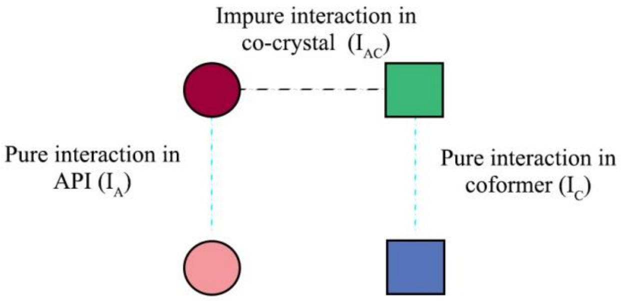


Figure 5.4: Generic graphic illustration of the competition between intermolecular pure (turquoise) and impure (black) interactions. Objects with the same shape belong to the same co-crystal component, whereas distinct colours represent different HB contributors

$$MD - score = \sum_{i=1}^{n} (\mathbb{P}[I_{A_i}] + \mathbb{P}[I_{C_i}] - 2\mathbb{P}[I_{AC_i}]) \qquad (5.9)$$

Each impure interaction in the final list was considered in comparison to the pure contacts of its respective components. The MD-score was evaluated as shown in eq. 5.9, where *n* represents the total number of impure interactions in the list. This score was computed for each route per API-coformer pair for comparison. A negative MD-score promotes the favourability of co-crystal formation.

5.3 Results and discussion

5.3.1 Experimental outcomes

LAG co-crystallisation was chosen instead of neat grinding due to the tendency of the latter method to produce incomplete conversions to co-crystals, thus requiring further purification steps.[64] Furthermore, a minimal addition of solvent to the grinding material seems to have the likelihood of accelerating co-crystal formation kinetics.[64] Given such crucial role, it was essential to ensure that both components were fairly soluble in the chosen solvents. Even though the amount of added solvent was fairly coherent amongst the samples of different antivirals, it was noticed that the consistency of the resultant paste varied, with those corresponding to FCV forming a slurry with even one/two drops of solvent. This effect was a clear consequence of the higher solubility of FCV in polar solvents, which might suggest that the role of the solvent was more active in such trials, beyond the initial stages of co-crystallisation.

Table 5.2: Experimental results of LAG co-crystallisation (Yes: co-crystal formed, No: no co-crystal formation). The results for GCV co-crystals were provided and analysed by another research group member. [318] Shaded results indicate a different outcome (or polymorph) due to the use of diverse solvent.

	Ganciclovir		**Famciclovir**		**Valacyclovir HCl**
Solvents	Water	DMF	Acetone	Methanol	Water
L-ascorbic acid	No	No	No	No	Yes*
Mellitic acid	Yes	Yes	No	No	Yes*
Fumaric acid	No	No	Yes	Yes	No
L-glutamic acid	No	No	No	No	No
Benzoic acid	No	No	Yes	Yes	No
L-tartaric acid	Yes	Yes	Yes	Yes	No
Salicylic acid	No	No	Yes	Yes	No
Trans aconitic acid	Yes	Yes	Yes	Yes	No
Glutaric acid	No	No	Yes	Yes	No
Methyl Paraben	No	No	No	No	No
Nicotinic acid	No	No	No	No	No
Urea	Yes	Yes	No	No	Yes
Caffeine	No	No	No	No	No
Nicotinamide	Yes	No	No	No	No
L-malic acid	No	No	NA	NA	No
Citric acid	Yes	Yes	Yes	Yes	Yes
Pyridoxine	No	No	No	No	No

*These results were extracted from literature.[319]

The experimental results of the LAG co-crystallisation for each antiviral are presented in table 5.2. In most trials, the heat produced from manual grinding was enough to provoke the evaporation of the solvent, thus leaving a dry powder sample behind. However, samples that remained wet, particularly those with hygroscopic components, as l-malic acid, were dried in a dessicator in the presence of a drying agent.

Nonetheless, in case of FCV and l-malic acid, the final resultant sample after LAG remained wet, despite several attempts for dehydration. The lack of proper crystallisation made it impossible to test the product appropriately, and thus any predictions related to this particular potential co-crystal were removed from the following results.

The number of successful trials for GCV and FCV was fairly close, with six and seven co-crystals respectively, while VCV·HCl trials resulted in the least positive outcomes. Each sample was characterised by PXRD and FTIR spectroscopy, as those corresponding to FCV-benzoic acid co-crystal in figures 5.5 and 5.6. Significant variations can be observed between the characterisation data related to the newly formed co-crystals and their constituent pure components. This was expected due to the changes in atomic positions and HB framework caused by the integration of two distinct molecules within one crystal lattice. All the data concerning successful co-crystallisation for FCV and VCV·HCl can be found in appendix.

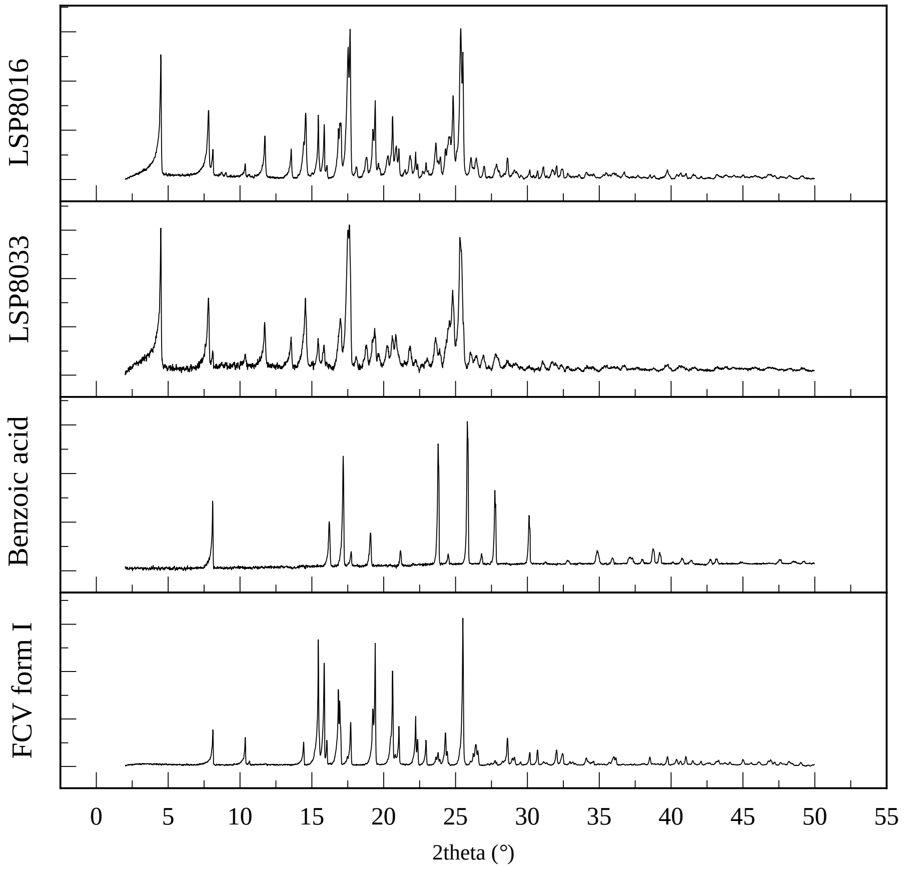


Figure 5.5: PXRD patterns of the starting compounds FCV form I and benzoic acid, together with the resultant co-crystals formed using methanol (LSP8016) and acetone (LSP8033).

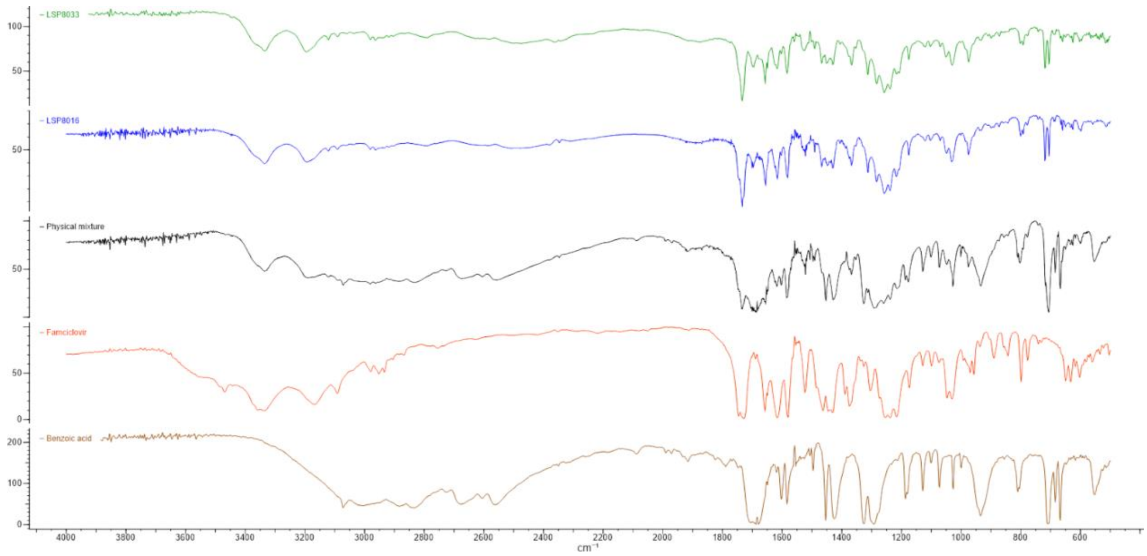


Figure 5.6: FTIR spectra of the starting compounds FCV form I (orange) and benzoic acid (brown), their physical mixture (black) together with the resultant co-crystals formed using methanol (blue) and acetone (green).

The use of polar aprotic and protic solvents did not cause any differences in FCV co-crystal results, with identical PXRD patterns and FTIR spectra, as seen in figures 5.5 and 5.6. Conversely, in case of GCV, there were two cases where co-crystal formation led to different polymorphs and a trial concerning nicotinamide which resulted in opposite outcomes. These instances (shaded in blue in table 5.2) suggested that for GCV trials, the interactions with solvent molecules did not always lead to the same molecular conformations or packing arrangements. Such effect might have been caused by the exposure of solvent molecules to cleavage planes with distinct polarities.

Citric acid was the most successful coformers, alongside urea, l-tartaric acid, mellitic acid and trans aconitic acid. These coformers are characterised by a wide range of molecular sizes, but they all constitute of multiple HB contributors. In fact, on comparison, it was observed how these coformers have the highest ratio between the number of HB contributors and the total amount of atoms (see appendix). This was also noted from the quantity of HB donors and acceptors as defined by Lipinski.

5.3.2 Computational approaches for prediction

Co-crystallisation is a mechanism which still requires further investigation to be understood exhaustively due to the interplay of various factors which influence its course.

Such process can be very sensitive towards the method of execution as well as the nature of the solvents used. Despite not fully comprehended, the role of solvent in LAG is generally regarded to facilitate molecular diffusion, whilst influencing the process of crystal formation.[320-321] However, the nature and extent of impact of solvents may vary according to the polarity and relative solubility of the components, as observed through the obtained experimental outcomes. Such impact highlights a possible interaction between the solvent and the solute components during grinding, which leads to a change in the molecular conformation prior to nucleation.

The design of each approach for co-crystallisation prediction was conducted with the premise that the final predictions will be validated against experimental results obtained through LAG co-crystallisation. Given that presence of the solvent molecules during this stage provides additional degrees of freedom, one should not assume that the molecular conformation during such process is necessarily similar to the one present within either the pure crystalline component or the resultant co-crystal, where the molecule exists within a rigid framework of long range intermolecular bonds.[320] This issue introduced another dimension to the typical prediction of co-crystallisation, especially when taking into account that the chosen approaches are all sensitive towards the molecular conformations to a varying degree. This happens due to the fact that these techniques all revolve around the availability of HB contributors, which in return is dependent on the specific molecular orientation. The *COSMO-RS* approach was the first one to be employed because it is able to generate low-energy conformers in solution. The testing of various conformations was done to potentially increase the coherency between experimental and computational paths, thus improving the quality of predictions.

Given the novel adjustments of the following methods, as well as the complexity to perform conformational analysis in the presence of multiple components, the prediction of VCV·HCl co-crystals was restricted only to HBP models. In addition, all the derived observations and conclusions are only applicable to case studies dealing with molecules with similar properties as those involved in this study.

5.3.2.1 COSMO-RS approach

COSMO-RS is a technique developed to evaluate thermodynamic equilibrium characteristics of liquids, based on the statistical physics of interacting molecular surface segments, thus making it ideal to deal with systems in solution.[128, 322] Quantum chemical continuum solvation calculations are used to assess the surface screening charge densities, which in return quantify the interaction energies that are required for the prediction of a wide range of properties.[134] A more detailed description can be found in various papers, written by Klamt.[128, 134] Given that co-crystallisation deals with solid-state products, both API and coformer are treated a virtually subcooled liquids, neglecting long range contacts present in their respective crystal structure. Co-crystallisation prediction is possible because the excess enthalpy of mixing and Gibbs free energy of mixing of the super-cooled melt (eq. *5.1* and *5.2*) of both pure components are correlated with their potential to form a co-crystal. The process of co-crystal formation is very often directed by short-range API – coformer contacts composed of HB and vdW interactions, all of which can be appropriately catered for by COSMO-RS theory.[132]

5.3.2.1.1 COSMO σ-profiles

The σ-profile acts as a fingerprint of the chemical nature of the constituent functional groups of each antiviral, presenting the charge distribution on the molecular surface as a histogram. Given that the σ-surface is sensitive towards the exposure of each functional group, it implies that the σ-profile is unique for each particular conformation. Such characteristic is featured in plots exhibiting the σ-profile of multiple conformers for each antiviral (see appendix). Histograms can be used to extract information about the polar nature of each molecule, as well as for comparison, as illustrated in figure 5.7. In fact, the profile for GCV extends to a wider range relative to that of FCV, indicating the presence of less diverse components in the latter antiviral. This property is also displayed through the maximum value for $p(\sigma)$ which belongs to FCV.

Peaks within the range of ± 0.008 e/Å^2 in σ-profiles represent groups which are less likely to form strong HB, primarily –CH groups.[151] The largest region within this bounded range is associated with the FCV conformer due to the presence of –CH groups along the branches. The lack of conventional HB donors is displayed in the σ-potential graph of FCV with a relatively smooth negative gradient on the positive x-axis. Conversely, the significant presence of HB donors in GCV is exposed through the presence of higher peaks with a wider range on the negative axis of σ-profile, and consequently a steeper negative gradient on the positive axis of the σ-potential graph. The peaks on the positive axis of the σ-profiles and the positive steep gradients of the σ-potential graphs on the negative axis suggest the comparatively equivalent presence of potential HB acceptors within the antivirals.

Figure 5.7: The σ-profiles (top) and σ-potentials (bottom) of a generated conformer for GCV (blue) and FCV (red). The profile of each graph can partially vary depending on the conformation of the molecule, especially in cases where any potential HB contributor would be shielded from the surface.

5.3.2.1.2 Selection of molecular conformations

As stated earlier, the molecular conformation present prior to nucleation which initiates co-crystal formation during LAG co-crystallisation is not necessarily equivalent to that within the unit cell. This is particularly more probable in cases where the conformation within the solid-state experiences intramolecular strain which is only compensated by the intermolecular network within the crystal or has very high solubility. Despite such possibility, it is a usual practice that only one conformation, representative of the solid-state is used as an input for co-crystal prediction. On the other hand, the low volume of solvent used during the procedure might not be enough for the solvent molecules to completely screen the pure components' molecules, and therefore, the dominant molecular conformation/s might not be equivalent to those thermodynamically stable in solution. Disparity between the actual and predicted molecular conformations, especially in relatively large API molecules is known to be one of the potential reasons of incorrect predictions.[181] In order to investigate such issue, the process of prediction was conducted using sets of conformers present in solution (one is based on a range of solvents – *kN* set and one is solvent-specific) and an extended conformation representing solid-state environment.

5.3.2.1.2.1 Conformer sets

The ReSCoSS framework was chosen because it allows descriptors other than free energies to drive the selection and ranking processes.[303] Furthermore, it does not depend on stringent cut-off limits as the *COSMOconf* workflow, hence decreasing the possibility of discarding relevant conformers, especially in cases where energy calculations are performed at a low level of theory.[303] In addition, this workflow, together with the selected conformer generators, do not utilise low level force-fields for the initial energy evaluation, which is often used as the underlying variable for the primary conformer selection process.[305] Instead, a semi-empirical method in a COSMO environment is utilised, with the intention of providing better energy values.

The choice of *k* and *N* was crucial because they control the balance between the influence of shape diversity and COSMO-RS free energy present within a particular set of conformers. A set with a higher value of *k* is expected to be more geometrically diverse than one with a smaller *k* value. The value for *k* was primarily determined by clustering on the standardised set of variables, ensuring convergence of the variation of cluster centres. Given that both GCV and FCV have structurally determined anhydrous forms, a wider range of *k* values were tested (*k* =1,3,5,8). This was done because examples given by Udvarhelyl *et al.* illustrated how sometimes conformations relevant to the solid-state were only selected in imbalanced *kN* combinations.[303] In fact, this happened in case of GCV, whereby the extended conformation most similar to that present in form I was only obtained in *k8N2* set. The greater extent of geometric diversity in sets with high *k* values was illustrated through the σ-profiles of the conformers, particularly for GCV (see appendix).

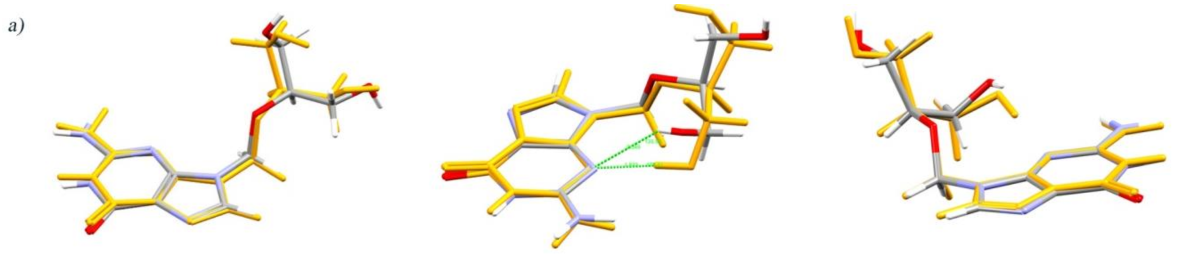
The ability to reproduce a conformer which represents the molecular geometric orientation within the crystal structure is considered to be a measure of assessing the appropriateness of specific *kN* set. However, as annotated earlier, one needs to take into account that a molecule might adopt a conformation with a lower intramolecular energy when present in solution.[303] Despite not reproducing the exact geometric orientations present in UGIVAI, UGIVAI01 and IMADEI, relatively similar conformations were still obtained (see table 5.3).

Table 5.3: Sets of conformers with different *k* and *N* combinations for each antiviral. Some molecular structures are illustrated in figure 5.8. Only GCV sets were reduced in size due to few conformer pairs with a RMSD < 0.1 and negligible Boltzmann's weights.

	k	***N***	**Number of conformers present**	**Conformer relevant to solid-state is present**
GCV	1	10	42	Form II (c_93)
	3	3	33	Form II (c_93)
	5	2	38	Form II (c_93)
	8	**2**	50	Form I (c_309) and II (c_86)
FCV	1	10	31	Form I (c_32)
	3	3	25	Form I (c_32)
	5	**3**	39	Form I (c_32)
	8	2	37	No

5.3.2.1.2.2 Dominant conformers in different solvents

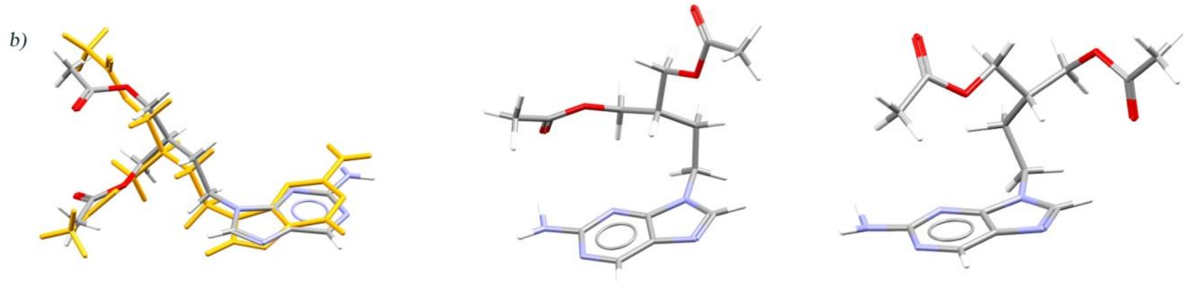
The assessment of different *kN* combinations was carried out by analysing the dominant conformations in solvents of different polarities. When a compound is inputted as a set of conformers in *COSMOtherm*, the results will be based on the properties of each individual conformer, but scaled according to their respective weight factor related to the Boltzmann's distribution.[309] The Boltzmann-weighting, w_j depends on the conformer equilibrium computed using the COSMO energy, E_i^{cosmo} and chemical potential of each conformer in the liquid state. While surveying the conformers generated and selected in sets, particularly for GCV, it was clear that the most relevant ones had a significant degree of folding, accommodating intramolecular contacts, not necessarily HBs.



Conformer 309
Set: *k8N2*

Conformer 93
Set: *k8N2*
Water: 14.54%
Cyclohexane: 45.15%

Conformer 86
Set: *k8N2*
Water: 28.92%
Cyclohexane: 0.86%



Conformer 32
Set: All excluding *k8N2*

Conformer 47
Set: *k5N3*
Water: 13.58%
Cyclohexane: 25.29%

Conformer 165
Set: *k5N3*
Water: 39.05%
Cyclohexane: 6.51%

Figure 5.8: a) Generated conformers (orange) with the smallest RMSD relative to the solid-state polymorphs of GCV, with conformers 86 and 93 being dominant conformations in various solvents and b) conformer 32 (orange) superimposed on the molecular conformation of anhydrous form of FCV, together with significant conformations in cyclohexane (47) and water (165). All dominant conformers are accompanied by their respective Boltzmann's weight percentage, corresponding to each respective solvent. Intramolecular HBs are visualised in light green.

GCV conformers 86 and 93 had the lowest RMSD values when superimposed onto the molecular structure of form II, but their intramolecular bond was characterised with a higher degree of directionality, probably leading to their comparatively substantial Boltzmann's weights (figure 5.8a). Despite an extended formation similar to UGIVAI01 would have enabled the participation of several functional groups in HB with the solvent molecules, the formation of an intramolecular bond was still favoured. This observation supports the conclusion made in chapter 4, stating that form I is more thermodynamically stable as a solid-state structure at ambient conditions, whereas the molecule of form II has a lower intramolecular energy and is much more geometrically similar to dominant conformations in both gas and solvation phases. The most relevant conformations in solvents with different polarities were all characterised by an intramolecular bond, with the only variation revolving around the position of the other hydroxyl group. For example, in set *k8N2*, this hydroxyl group was more accessible for HB in the presence of polar solvents (conformer 31 and 86), rather than in a relatively non-polar environment (conformer 93 and 316). The extent of folding and lack of sufficient groups to participate in HB when in solution might be a major contributor towards the low solubility of GCV in many solvents.

The branches of the most dominant FCV conformers in non-polar solvents were positioned in a way to maximise the exposure of non-polar groups within the molecule, despite having a degree of folding onto the sides. This can be observed in conformers 47 (*k5N3* set), 236 (*k3N3* set) and 178 (see appendix), whose ring systems are exposed, primarily to be available for any CH···π interactions with the solvent molecules. Similar effect can be noted for the relevant conformations in polar aprotic solvents as DMSO and acetone, which can utilise their methyl groups to form such contacts with the aromatic backbone of FCV, coupled with HB involving the carbonyl groups. Conversely, in polar protic solvents, as water and methanol, the most dominant FCV conformer (165) was characterised by a large degree of folding onto the aromatic system, most likely to stabilise intramolercularly through CH···π interactions with a distance of 3.372 Å between the its C-H and the aromatic plane.[282] Furthermore, the carbonyl groups in conformer 165 were still available to form strong HB with solvent polar functional groups. Despite these geometrical differences, the σ-potential graphs of FCV conformers (47 and 165) were relatively equivalent, except for those associated with fully extended conformations (32) which had steeper negative gradients (accessible HB donors). Such uniformity amongst the σ-potential graphs might have been caused by the presence of longer carbon chain and terminal methyl groups whose changes did not necessarily impact the exposure of the HB contributors.

5.3.2.1.2.3 Solubility prediction

Another method applied to identify the reliability of each set of conformers was based on solubility prediction, with the objective of utilising the discrepancy between the predicted and experimental solubility values as a measure of assessment. The solubility values of each antiviral in a variety of solvents were computed in *COSMOtherm,* at room temperature. The predicted results for each *kN* combination set were compared to the ones obtained by fitting appropriate regression models on clear points detected by crystallisers. The individual input set included different *kN* combination sets, as well as a conformer set composed of ten structures generated through the default algorithm in *COSMOtherm,* starting from a SMILES code, for comparison purposes. The majority of the clear points were validated through repeated runs, using the same temperature profile and stirring speed, as well as using particle viewer equipment.

One of the main disadvantages revolved the lack of experimental data related to the antivirals, particularly the Gibbs free energy of fusion, which was required in order to take into account the energy change from a super-cooled liquid to an ordered crystalline structure. Therefore, quantitative structure property relationship method was utilised to estimate this thermodynamic quantity for FCV, leading to less accurate predictions.[150] Given that the experimental aqueous solubility of GCV (2.6 mg/ml) was compatible with the literature, this property was used as a reference for the estimation of the Gibbs free energy of fusion for GCV.[323] The iterative algorithm was chosen in attempt to improve the accuracy in the predicted absolute solubility values. Literature only provides minimal solubility data, many of which vary too drastically to be used as a reliable reference.

In case of GCV, its marginal solubility made it very difficult to obtain any clear points, even when small concentrations (<1 mg/ml) were utilised. In such cases, due to the small amount of solute particles, the instrument was unable to detect the point of dissolution (100% transmissivity with solute particles present). Apart from the aqueous solubility which was compatible with the literature, other values were close to 0 mg/ml, highlighting the lack of solubility of GCV, even in polar solvents as ethylene glycol.

The results obtained from the conformer sets chosen earlier were similar, with few non-significant discrepancies, probably due to the fact that the dominant conformations within each set had similar σ-potential graphs (see appendix for results). The predicted solubility values for antivirals associated with the *kN* sets demonstrated an overall higher extent of compatibility with the experimental results, compared to those produced by the sets of automatically-generated conformers. This observation supported the application of the ReSCoSS framework. Nonetheless, there were still considerable discrepancies, caused by both experimental and prediction types of errors, particularly for the solubility of FCV in 2-propanol. It is worth noting that when QSPR approach was used for estimation, without the aqueous solubility value as reference, the predicted solubility of GCV in water using the automatically- generated conformers was 37.016 mg/ml. This significant disparity shows the importance of integrating experimental data.

The relevant presence of both non-polar and polar groups, coupled with the high degrees of freedom in solution, created a more difficult task to predict solubility. The use of a wider range of solvents in conjunction to more diverse concentrations would have provided a better picture of the solubility profile for each antiviral. A more suitable approach of testing the appropriateness of the chosen set is by comparing the experimental NMR chemical shift of each antiviral in solution with that calculated for the most dominant conformation in a particular solution within the set. However, one of the main problems with such method would be the low solubility of the antivirals, particularly GCV in solvents suitable for NMR spectroscopy.

5.3.2.1.3 Co-crystal formation predictions

The conformer sets in solution based on multiple solvents (ReSCoSS) were chosen based on their respective dominant conformations in solvents of different polarities, as well as on their ability to include a conformer relevant to the solid-state. The results displayed in the following section relate to sets which satisfied both criteria. For exhaustiveness purposes, the remaining sets were also used for prediction, and the results were relatively similar. Due to the assumption that the interactions in the crystal are similar to those present in virtual supercooled liquid, then the strengths of such contacts in the co-crystal can be approximated via the excess enthalpy of mixing, H_{ex} (kJ/mol), at a given stoichiometry.[150] It was exhibited in several papers how the parameter of H_{ex} is a more suitable indicator for co-crystal formation prediction than the Gibbs free energy of mixing, G_{mix}.[181, 324] The H_{ex} is known to serve as a ranking system, rather than to provide the absolute values for co-crystal formation.[132] Nonetheless, the likelihood of co-crystallisation decreases exponentially with positive excess enthalpies.

5.3.2.1.3.1 Input single conformer or sets

On comparison, the predictions for the two antivirals produced using the conformer sets generated by the ReSCoSS framework and those selected based on the Boltzmann's weights in specific solvent only, were almost identical (see appendix). Given such similarity, it was not possible to deduce the optimal selection of API input for these cases of co-crystal prediction. Further research, including a larger and more diverse sample of targets is required for clearer indications. However, this evidence demonstrated that *kN* sets contained relevant conformations for each individual solvent, despite following a novel route for selection, characterised by clustering. The consistency in predictions was particularly evident for FCV co-crystal results (see figure 5.12), probably due to having various conformers with highly uniform σ-potential graphs. Such level of coherency can be also noted amongst results obtained from routes using an individual extended conformer and a set of conformers as inputs. With regards to GCV co-crystal predictions (figure 5.10), the consistency in ranking resided solely amongst the top half ranks of the predictions produced using an extended conformer and a full set as inputs. This degree of compatibility mirrors the intermediate level of dissimilarity in the σ-potential graphs of GCV conformers.

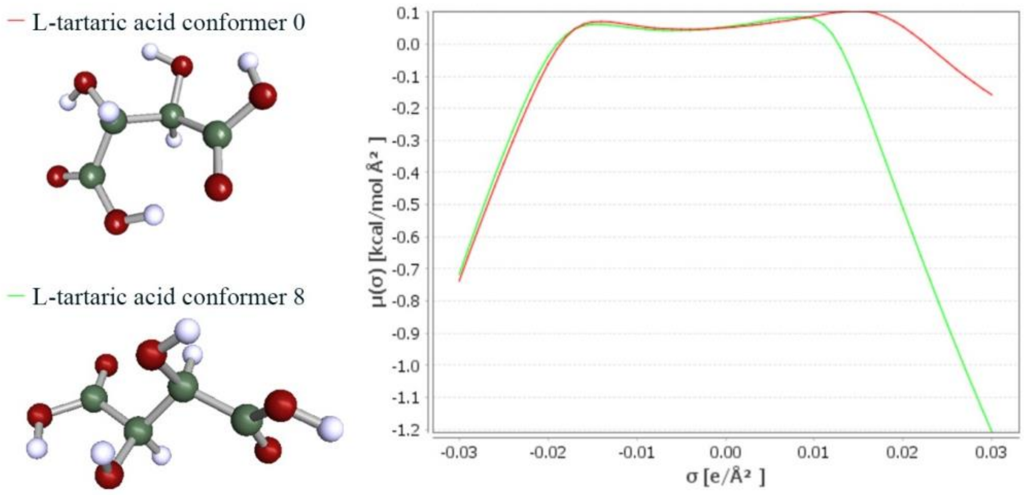


Figure 5.9: The molecular conformation of l-tartaric acid conformers 0 (red) and 8 (green), together with their respective σ-potential graphs.

The utilisation of a conformer set had higher impact when the σ-potential graphs were fairly different, as in the case of l-tartaric acid and l-glutamic acid. For example, the H_{ex} for potential co-crystallisation of GCV (*k8N2* set) with the conformer set of l-tartaric acid is -4.619 kJ/mol, whereas using only the coformer conformation 0 increased H_{ex} to 0.523 kJ/mol. This is caused by the folded structure of conformer 0, which enabled the formation of multiple intramolecular HBs, leading to the lack of HB donors which can be available to interact with the solvent molecules, as shown in figure 5.9. Conversely, conformer 8 had an extended orientation, hence exposing more potential HB participants and resulting in a more favourable H_{ex} value (-3.489 kJ/mol). The structures with the most distinct σ-potential graphs corresponded to molecules populated with a considerable number of polar functional groups.

5.3.2.1.3.2 Rankings

The predictions for GCV co-crystals are illustrated in figure 5.10, whereby coformers are ranked (rank 1 being the most favourable) based on the H_{ex} values produced by each individual route (inputted conformation/s), thus enabling a clear comparison of results. The coformers are listed on the y-axis according to the ranking of the first route. The experimental results for co-crystal screening are also shown through first set of circles next to the left margin. This description is valid for all the following parallel plots.

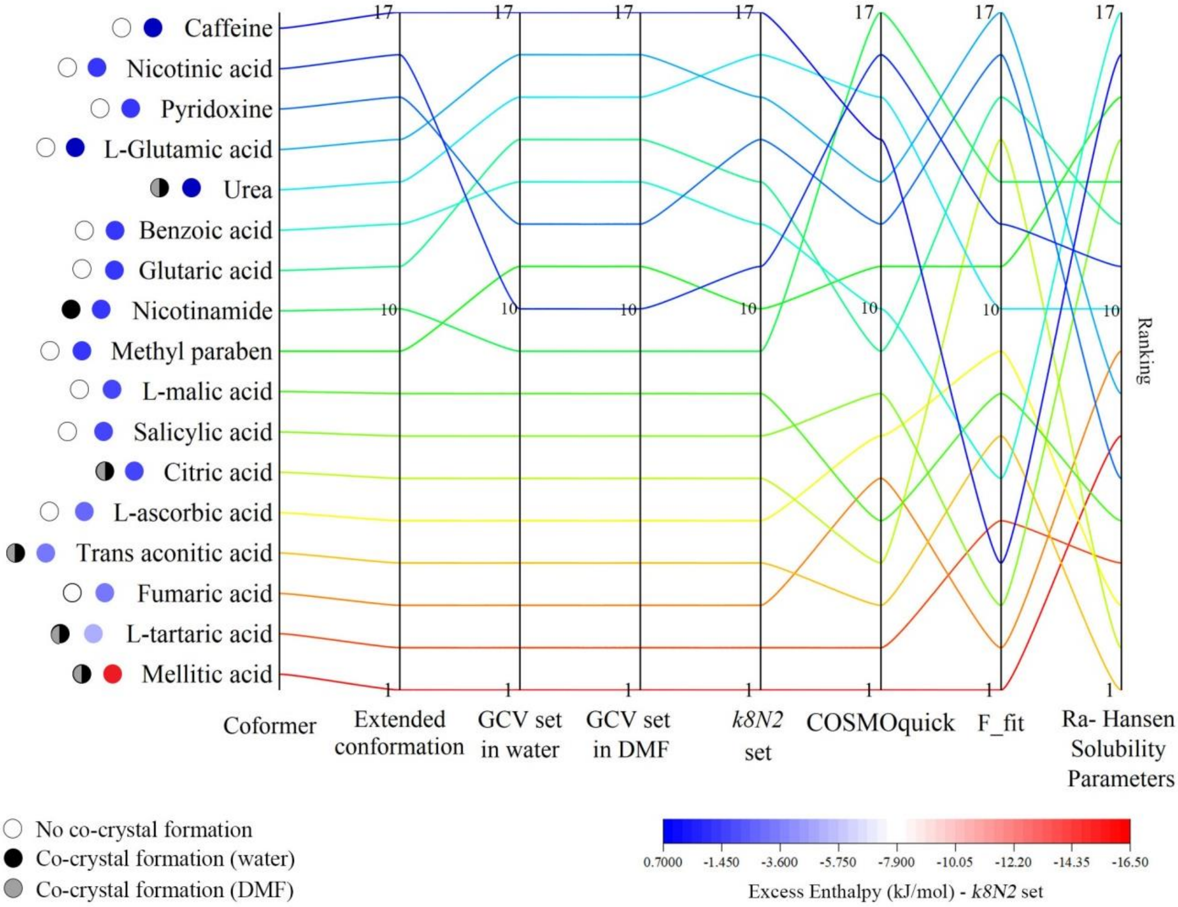


Figure 5.10: A parallel plot displaying the predicted rankings based on the H_{ex} values (various routes), f_{fit} model and Hansen distance, for the appropriateness of each potential coformer for co-crystallisation with GCV (1:1). The x-axis presents different routes employed for comparison (route 1: Extended conformation till route 7: Ra- Hansen Solubility parameters).

The experimental results demonstrated the formation of six GCV co-crystals (1:1), four of which were ranked amongst the top positions by the majority of the routes presented in figure 5.10, thus supporting the appropriateness of this methodology. However, evidence showed that the COSMO approach should be restricted to 1:1 stoichiometry because experimental results of other ratios were not compatible with predictions (see appendix).[318] Upon comparison, the predictions by *COSMOtherm* only yielded a small improvement relative to the ranking based on the H_{ex} computed by *COSMOquick*, which constructed each antiviral using different fragments from *COSMOfrag* database. This outcome exhibited how *COSMOquick* can provide a good indication of the favourability of co-crystallisation at a much lower computational cost, as opposed to the utilisation of sets which did not yield a significant improvement in predictions, except for the ranking of nicotinamide. The F_{fit} model (eq. 5.4), and Hansen distance, Ra provided the most relative diverse set of rankings. The latter approach utilises the difference (computed as Ra) between the Hansen solubility parameters of the pure components to mirror their degree of compatibility. A high extent of compatibility would add evidence in favour of the suitability of that specific candidate to act as a co-former for co-crystallisation.[312]

The likelihood of co-crystal formation was predicted to be fairly equivalent in water and DMF, and this was supported experimentally by the fact that the majority of the successful coformers were adequate in both solvents. It is noteworthy that having a considerable propensity to form a co-crystal in both solvents does not have implications on whether the same polymorph will form or not. Given the similarity in the predictions produced by different approaches and their respective accuracy, it is difficult to derive any conclusive observations regarding the process of LAG at a molecular level. The discrepancies in the rankings predicted by f_{fit} reflected the influence of NRot on the final predictions. For example, in case of citric acid, its molecular flexibility was overestimated by NRot, hence misleading its predicted favourability as a suitable coformer, despite having a comparatively small H_{ex} value. This variable was the factor which caused f_{fit} model to produce the least accurate rankings when compared to experimental results.

The predicted rankings revealed the substantial favourability of mellitic acid, fumaric acid, trans aconitic acid and l-tartaric acid as coformers, with the majority of the approaches in agreement (figure 5.10). Similar observations were noted for the ranking of FCV co-crystals in figure 5.12. Given that HBs are utilised as synthetic vectors for co-crystal formation due to their strength and directionality, the explanations in relation to the mechanism behind the ranking will revolve around the optimisation of such intermolecular contacts. The advantage of σ-profile resides in illustrating the σ-charge density exposed to the surface, hence integrating both polarity and availability. The plot in figure 5.11 illustrates the σ-profiles of some of the potential coformers, together with that of GCV conformers for comparison. The high peaks (dark blue) on both sides of the x-axis indicate the abundance of HB contributors in mellitic acid which ensure significant interaction with the antiviral that leads to an exceedingly negative excess enthalpy (-16.485 kJ/mol), correctly exposing its tendency to form a co-crystal. The extent of the negative H_{ex} value might suggest the presence of having sufficient interactions between the API and coformer in solution, so as to reduce the activity of the antiviral relative to its pure state, thereby improving its solubility.[132]

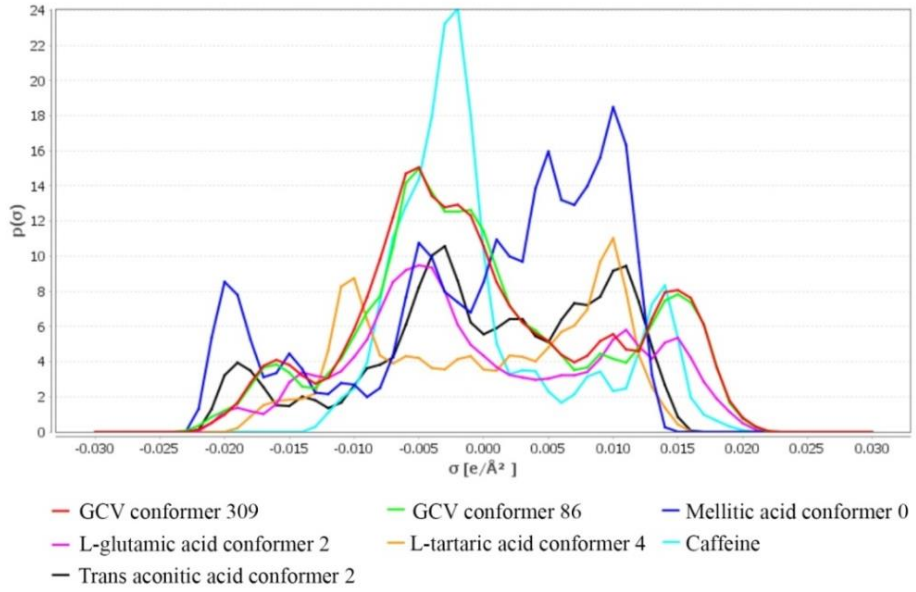


Figure 5.11: The σ-profiles of the two GCV conformers (309: extended and 86: folded), and some coformer conformers which were ranked on the extreme ends on the range.

The favourability of l-tartaric acid and trans aconitic acid as adequate coformers, supported experimentally, can be extracted from their respective σ-profiles, which are characterised by similar patterns as the one for mellitic acid, but with smaller peaks. The considerable number of potential HB participants in l-tartaric acid compensated for the unavailability of some of them which might have been utilised in intramolecular bonding. The low ranks of caffeine and l-glutamic acid can be attributed to the major conformity present amongst their σ-profiles and those corresponding to the antivirals, with large non-polar areas.

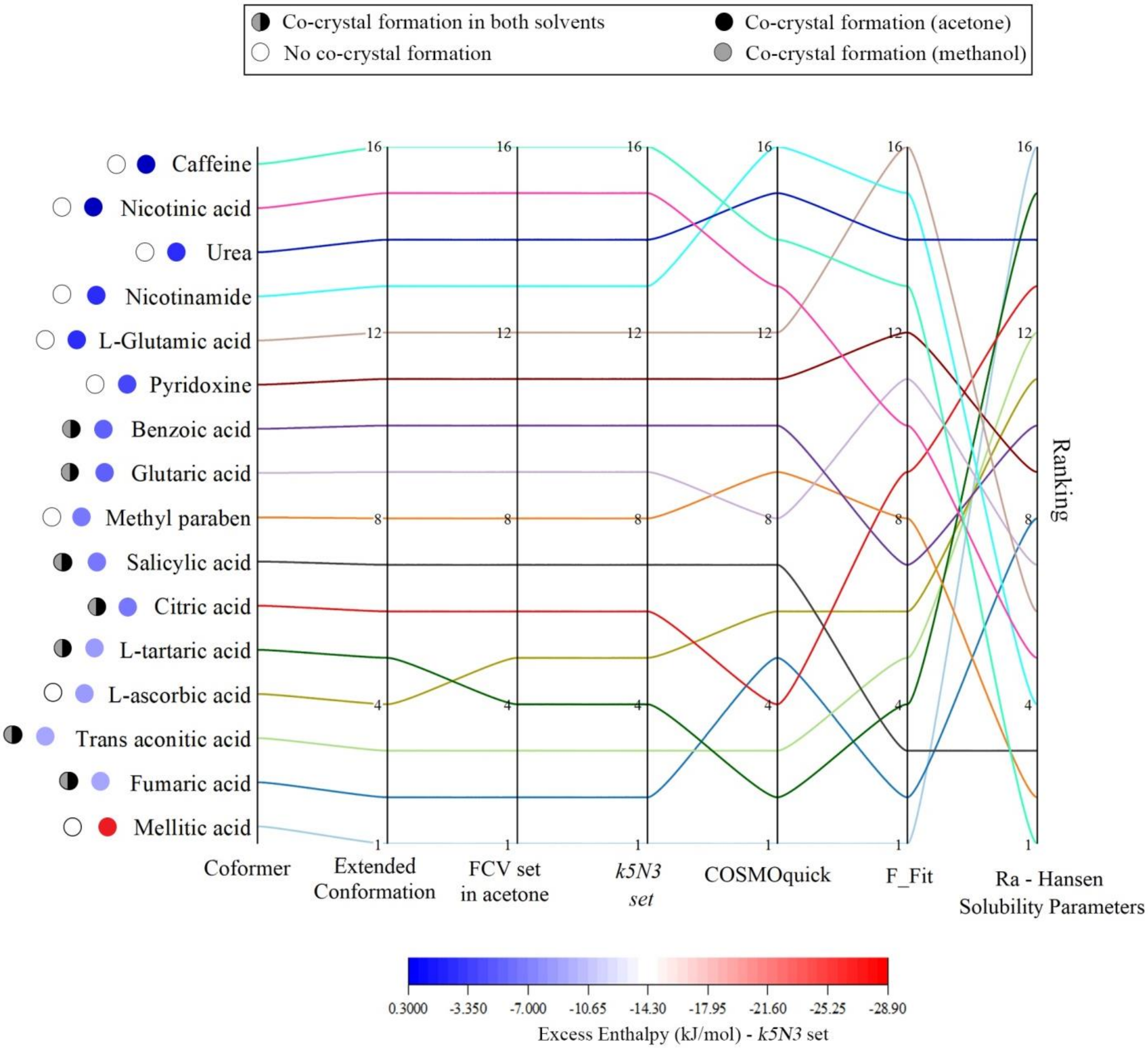


Figure 5.12: A parallel plot displaying the predicted rankings based on the H_{ex} values(various routes), f_{fit} model and Hansen distance, for the appropriateness of each potential coformer for co-crystallisation with FCV. The x-axis presents different routes employed for comparison (route 1: Extended conformation till route 6: Ra- Hansen Solubility parameters). Sets in acetone and methanol solvents produced the same results.

The predictions of FCV co-crystals promote the ability of the COSMO-RS approach, given that the seven successful co-crystal formations were classified amongst the first ten rankings. Furthermore, equivalent rankings were produced for the sets in acetone and methanol, in agreement with the experimental outcomes. This observation does not hold for the Hansen distance, which produced the most inaccurate results, with caffeine (1) and mellitic acid (16) being placed on the opposite ends of the rank. Better compatibility was observed between the predictions associated with the *kN* set, *COSMOquick* (H_{ex}) and

f_{fit}. In certain examples, the influence of NRot on the sequence of ranks was slightly minimised by the presence of more negative H_{ex} values.

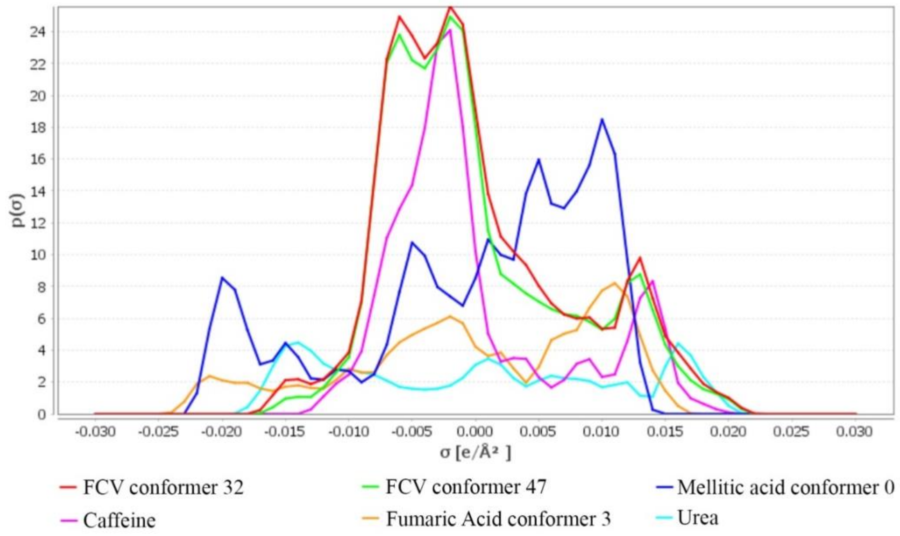


Figure 5.13 : The σ-profiles of the two FCV conformers (32: extended and 47: folded), and some coformer conformers which were ranked on the extreme ends of the range.

Similarly to fumaric acid, both trans aconitic acid and l-ascorbic acid were classified relatively high for FCV co-crystals, due to their strong available HB donors. The presence of intramolecular bonds led to the decrease in the ranking for salicylic acid and pyridoxine. The requirement of strong or abundant HB donors was not the sole factor governing the process of co-crystallisation. This can be illustrated through the unsuccessful FCV trials concerning urea and mellitic acid coformers, as opposed to GCV runs. The introduction of several FCV HB acceptors to mellitic acid molecules would increase the competition for a limited amount of HB donors, thus making the formation of API-coformer interactions not feasible. Conversely, this effect might be reduced in case of GCV due to its balanced ratio of HB donors and acceptors, which enables it to contribute towards impure interactions more holistically.

5.3.2.1.3.3 Flexibility score example: GCV- citric acid co-crystal

The f_{fit} parameter incorporates the NRot within its equation, in attempt to account for the individual molecular flexibility. However, on comparing the quantities given by NRot with *Mogul* torsion angle graphs, it was evident that as shown earlier, the NRot overestimated the molecular flexibility of the coformers. Such misrepresentation was not only caused by the discrete nature of NRot, but also due to the loose criteria of this parameter in *COSMOquick*, which also included terminal torsional angles. This led to inaccurate scores, particularly for coformers with a higher degree of flexibility, as citric acid (NRot = 9). When the quantification of flexibility was executed by the new method suggested in chapter three, the flexibility score for citric acid was reduced to 5.456, whilst still taking into account nine torsion angles for comparison purposes (see appendix). If such approach was applied to every potential co-crystal, there would be an overall decrease in f_{fit} scores, but more markedly for coformers with large number of potential rotatable bonds, thereby affecting the final ranking. In case of GCV-citric acid co-crystal, the new f_{fit} score improved its position in ranking as it resulted to be lower in value than some others, such as GCV-nicotinamide and GCV-nicotinic acid hypothetical co-crystals. This outcome supports the validity of the novel method for the flexibility quantification.

5.3.2.2 MC approach

The rationale of this approach relies on the assumption that comparable molecules are more likely to interact and pack efficiently, thus have a higher likelihood of forming co-crystals. This idea was derived from the typical efficient close packing of similar / identical molecules, which yield dense, thermodynamically stable crystal structures.[179] This premise was supported by the positive correlation coefficients between co-crystal formation and similar molecular descriptors.[178] The default five descriptors in *Mercury* were utilised for the analysis, three of which relate to shape (short (S), medium (M), long (L) axes), while the remaining deal with polarity (dipole moment (DM) and fraction of nitrogen and oxygen, (FNO)).

The choice of inputted molecular conformation is even more crucial for this approach because the likelihood of co-crystal formation depends on the level of molecular similarity between the API and coformer, rather than on the degree of opposing features displayed through the σ-potential graphs. The combinations of tested API-coformer conformations were defined in table 5.1. The solid-state molecular conformations of FCV and GCV form I were represented by their respective extended molecular conformers produced through the ReSCoSS framework. Besides the fact that the structural differences were not significant, with the same accessible HB functional groups, this was also necessary to enable fair comparison between approaches. Moreover, this was done with the intention of employing these routes for prediction to similar APIs with structurally undetermined crystal forms, particularly VCV·HCl, in future work.

5.3.2.2.1 Overall performance

The performance of MC approach with default settings proved to be rather inconsistent for small molecules.[302] Furthermore, the quality level of predications resulted to be unsatisfactory for relatively larger molecules, beyond the typical conservative parameters which characterised its original training set.[179, 302] Published studies highlighted the requirement for the customisation of the criteria in MC approach so as they can be coherent with the nature of data under analysis, particularly for those associated with the geometric compatibility.[155] Given that both GCV and FCV did not lie within such restrictive parameters due to their larger molecular weight, adjustments were employed to potentially improve the outcomes.

Sarkar *et al*. proposed the reduction of the number of passes required for a positive outcome from five to three.[179] When this suggestion was tested for GCV and FCV co-crystals, it did not discriminate efficiently between the potential coformers, resulting in many false positives. In the same research work, an adjustment of the acceptable difference between the length of S-axis resulted in an average higher success rate, presumably since they were dealing with molecules on the larger part of the spectrum of APIs.[179] Such adjustment was also applied, but only a decrease in the tolerance yielded an overall improvement in the success rates (table 5.4), particularly for GCV trials. The optimal acceptable difference for the length of S-axis was set at 2 Å (larger limits provided minor improvements, while tighter criteria resulted in the rejection of most trials).

Overall success rates indicated that the most successful route (68.97%) for MC is the one where all inputs represented the molecular structure within the solid-state. However, more in-depth analysis is required to determine whether such result is applicable to individual GCV and FCV co-crystal outcomes. The improvement in predictions can be attributed to the reduction in false positive results, caused by the tighter criteria with respect to the allowed variation of S-axis values. Such adjustment should only be applied to datasets entailing molecules with similar characteristics as those involved in this study. Furthermore, a larger training set with chemically similar molecules would have been useful to refine the customisation of the criteria. All results presented in the following section were obtained through the use of adjusted boundaries (S-axis).

Table 5.4: GCV and FCV co-crystal prediction success rates (based on true positives and true negatives) using the five descriptors produced by *Mercury*. Only experimental results related to 1:1 stoichiometry were considered. Detailed results can be found in appendix.

Routes	API	Dominant conformation in solvent	Extended conformation	Extended conformation	Dominant conformation in solvent
	Coformer	Dominant conformation in solvent	Dominant conformation in solvent	Solid-state conformation	Solid-state conformation
Default settings	**Success rate (%)**	53.03	54.00	62.07	58.14
Adjusted settings	**Success rate (%)**	57.58	64.00	68.97	58.14

5.3.2.2.2 Individual API predictions

In order to deal with the other potential inadequate cut-offs or borderline values, a ranking system was employed, which enabled the interpretation of results beyond the typical binary response (see appendix). In particular, such function is useful to highlight cases which are classified incorrectly due to small discrepancies outside the acceptable criteria. This ranking system involves a normalisation function, proposed by Costa *et al.* (*eq. 5.10*), based on the five descriptors and their respective bounds.[155] The function was also altered to mirror the adjusted criteria applied.

$$MC\ score = \frac{\Delta M/L}{0.31} + \frac{\Delta S}{2.00} + \frac{\Delta S/L}{0.275} + \frac{\Delta DM}{5.94} + \frac{\Delta FNO}{0.294} \qquad (5.10)$$

The parallel plot in figure 5.14 exhibits the ranking for GCV co-crystals, using diverse inputs as described earlier (indicated by boxes), along the x-axis, together with each respective success rate. The dominant conformation of GCV in solution was the same in both water and DMF (c101), but some coformers had different conformations. The results based on the molecular solid-state representatives for coformers excluded trans aconitic acid and pyridoxine because CSD (last updated March 2022) did not contain their respective pure crystal structure. Similar to predictions provided by the conformer set inputs in the *COSMO* approach, citric acid, l-tartaric acid and l-ascorbic acid were also categorised as adequate coformers. The majority of the correct predictions involved true negatives, with only one route providing four out of six correct true positives.

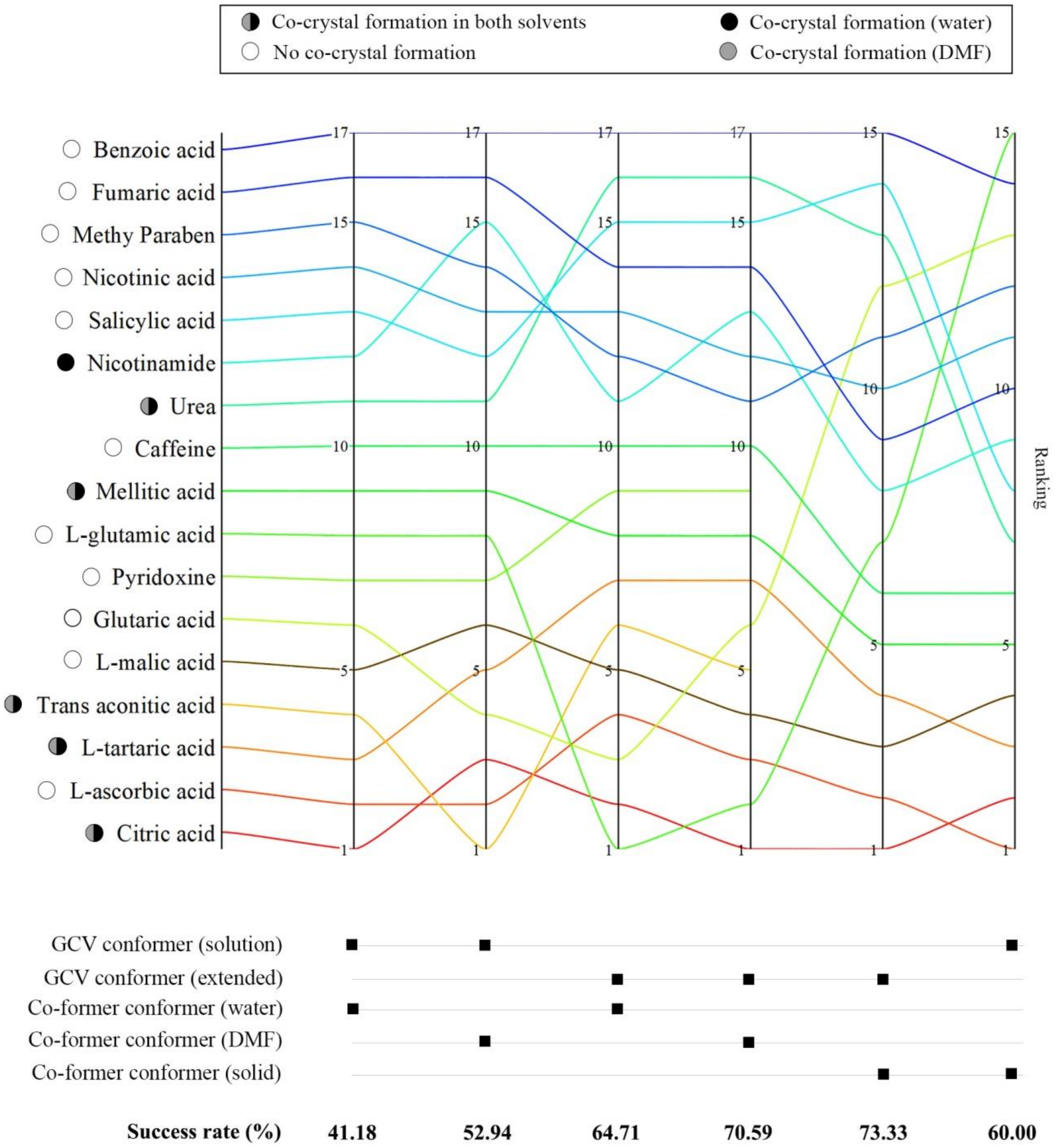


Figure 5.14: A parallel plot illustrating the ranking of the 17 coformers (similar setup to figure 5.10), based on the MC scores obtained using *eq.5.10*, for potential GCV co-crystals. The x-axis presents the six different routes applied, with descriptions of the inputs used for each route.

The advantage of this system can be illustrated through the example of mellitic acid, which despite being wrongly classified as inadequate by the several binary response routes, it was ranked amongst relatively favourable positions. Similarly, this ranking method exposed how the favourability of GCV-nicotinamide co-crystal was higher when the inputs were related to dominant conformers in water, as opposed to DMF. This observation is in agreement with the experimental formation of GCV-nicotinamide co-crystal, which was only noted when water was used as solvent. Even though it did not have the highest success rate, the route using the dominant GCV conformation in solution and solid-state representatives for coformers was able to produce improved results for urea and l-tartaric acid co-crystals. This contrasts the predictions obtained by the *COSMO* approach, where none of the inputs employed suggested the suitability of urea as a coformer. The significant favourability of citric acid as a coformer was exposed unanimously amongst all routes, both through the binary response and rankings.

The percentage rates in figure 5.14 indicated that the most successful route for GCV co-crystals was the one where all inputs represented the molecular structure within the solid-state, in line with the overall success rate in table 5.4. This might suggest that during LAG, the solute molecules might not be completely surrounded by solvent molecules, thereby more likely to retain their solid state conformation while interacting with each other. However, this premise might not be applicable for FCV co-crystals because the highest success rate corresponded to routes set in solution environment (figure 5.15), with a significant number of true negatives. Such result might be due to the higher solubility of FCV (a loose slurry was immediately obtained during grinding – not a paste as for GCV trials despite using the same volume of solvent), which might have enabled it to adopt a stable conformation in solution due to a more energetically feasible solvent screening. Further examples need to be tested in order to confirm the correlation between the most optimal routes and solubility, especially due to the similar success rates obtained.

The overall level of coherency amongst the binary results of FCV predictions was more significant relative to the outcomes of GCV trials. This observation mirrored the same trends noted from the predictions produced through the *COSMO* approach. In case of FCV, routes involving different dominant conformations in solution (c47 for acetone and c165 for methanol), produced identical binary responses (68.75%), probably due to their geometric similarity (see appendix). These predictions are in agreement with the fact that the same FCV co-crystals were formed experimentally, irrelevant of whether acetone or methanol was employed as solvent. The only few differences amongst the binary outcomes were caused by employing an extended conformation of FCV instead of the one dominant in solution. Nonetheless, the degree of variance between the five molecular descriptors of all FCV conformers was enough to generate some discrepancies in their respective ranking scores, with the route set in solution being the most accurate.

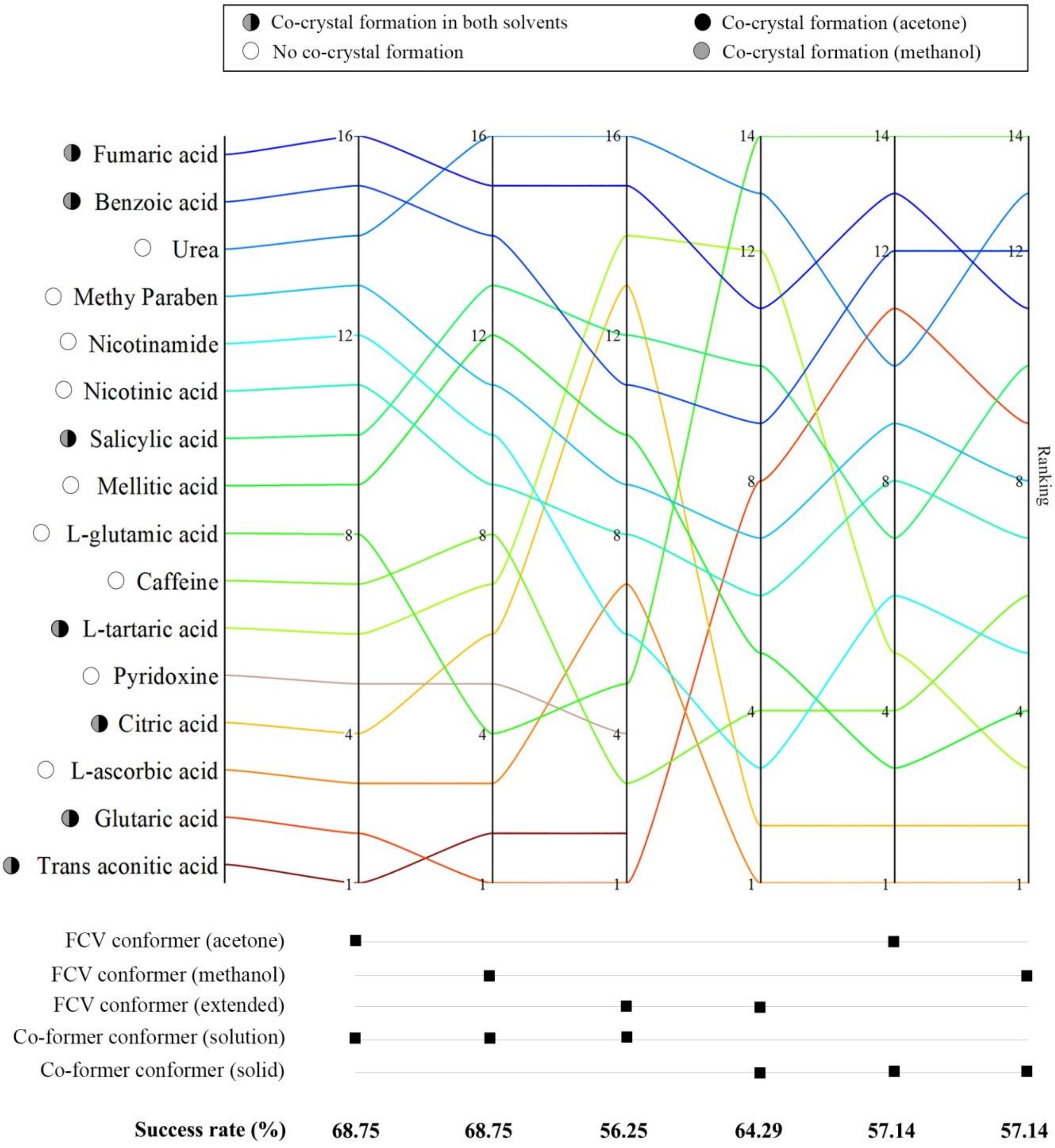


Figure 5.15: A parallel plot illustrating the ranking of the 16 coformers (similar setup to figure 5.10), based on the MC scores obtained using *eq.5.10*, for potential FCV co-crystals. The x-axis presents the six different routes applied, with descriptions of the inputs used for each route.

The integration of results obtained from both binary response and rankings provided improved indications. While the binary responses relied mostly on the adjusted bound of S-axis, whereby most routes resulted in true negative outcomes, the ranking scores were influenced by all descriptors. This enabled the correct result for FCV-glutaric acid which, opposed to the false binary response, it was ranked as a highly favourable coformer amongst the routes set solely in solution. Conversely, the indications for fumaric acid and benzoic acid were improved when an extended conformation of FCV was employed. This demonstrates that even though the route set in solution was the most accurate overall, a larger sample size is required for further validation, particularly to test its ability to predict true positive outcomes.

5.3.2.3 HBE approach

The use of extremes on MEP maps for virtual co-crystal screening was first introduced by Musumeci *et al.*, whereby the favourability of a co-crystal is judged based on the collective energy gain from potential intermolecular interactions between the parental components (*eq. 5.8*).[171] Parameters α_p and β_q (*eq. 5.5* and *5.6*) are used to represent the strength of specific interacting sites, which inherently depend on the availability and chemical nature of the functional groups present. This implies that predictions are sensitive towards the molecular conformation used, and the level of theory employed to construct the MEP map. In similar applications very often only the most stable conformer in vacuum will be considered, with a varying degree of success.[171, 317, 325] However, due to the same reasoning stated earlier for the MC approach, all molecular combinations in table 5.1 were tested. Possibilities as intramolecular bonding highlight the importance of using the correct conformation, because they can significantly impact the predictions.

Another crucial decision within this approach concerns the selection of interacting sites, which very often are composed of conventional HB groups. In similar works, it is accustomed that only such traditional sites are taken into account, excluding other extreme points specifically related to carbon HB donating abilities.[171, 317] Due to the role of –CH groups within GCV and FCV crystal structures, as demonstrated in chapter four, maxima related to such groups were also taken into account. These maxima were predominantly present on the MEP maps of the antivirals. In order to compare and determine the effect of such considerations, the hierarchal approach of pairing interacting sites was done with and without –CH contributions. Cases of intramolecular bonds were only considered if the HB angle was larger than 120° and distance was smaller than the sum of their respective VdW radii. Participants in intramolecular bonding were excluded from the list of potential intermolecular contributors, excluding oxygen atoms (particularly carbonyl oxygen atoms) whose position was likely to enable bifurcated HB.

5.3.2.3.1 Overall performance

The ranking system of this approach is based on ΔE (*eq.5.8*), which is derived from the difference in energy of interacting pairs constituting the pure components and the potential co-crystal. The worst possible value for ΔE is zero; where the interaction pairs within the co-crystal are identical to those present within the parental components (no gain in energy). Given such system, a cut-off value is required to indicate which co-crystals are favourable to form. Musumeci *et al.* concluded through thermodynamic analysis, that $\Delta E < -11$ kJ/mol would indicate that the probability of co-crystal formation is higher than 0.5.[171] The restraint in confidence is due to the underlying assumption that all possible interactions need to occur within the new co-crystal structure, and that pairings will be governed by the strength of each respective participant in a hierarchal manner.

Table 5.5: GCV and FCV co-crystal prediction success rates (based on true positives and true negatives) based on ΔE values (*eq. 5.8*). Only experimental results related to 1:1 stoichiometry were considered. Detailed results can be found in appendix.

Routes	API	Dominant conformation in solvent	Extended conformation	Extended conformation	Dominant conformation in solvent
	Coformer	Dominant conformation in solvent	Dominant conformation in solvent	Solid-state conformation	Solid-state conformation
Inc. CH	**Success rate (%)**	65.15	66.00	51.72	46.51
Exc. CH	**Success rate (%)**	62.90	74.47	48.15	42.86

Table 5.5 exhibits the overall success rate per route, while categorising results which included and excluded carbon as HB donors. Single differential values measuring the variance between the highest pure and impure interactions overall resulted in less accurate results (ex. for urea co-crystals), and therefore were excluded from the final predictions. Caffeine is an example where carbon H donors exhibit their full ability to participate in HB, because they are the sole type of HB donors present in the molecule. This implied that in cases where the carbon HB donors were excluded, it was not possible

to reproduce the relative strength of the intermolecular framework of caffeine, and hence it was not possible to evaluate the probability of co-crystal formation without any bias.

On comparison, the most successful route entailed the extended conformation of the antiviral and the dominant conformer in solvent for the coformer. This route had the second highest success rate for the MC approach (64.00%). In addition, this combination of molecular conformers performed better when carbon was excluded from HB interactions. In general, the cut-off point value of -11 kJ/mol proved to be rather conservative for GCV and FCV trials, with the majority of correct responses being true negatives. As noted for the MC approach, an analysis of the individual API-coformer ranking is required to understand the results better, particularly regarding the ability of identifying suitable coformers.

5.3.2.3.2 Individual API predictions

The ranking of coformers in the following parallel plots was done based on ΔE values, as suggested in literature. Given the numerous routes tested, a ranking score inspired from the research work done by Costa *et al.,* was devised in attempt to identify which route provided the best predictions, based on the experimental results.[155] The ranking score is as follows:

$$Ranking\ score = \sum_{i=1}^{n} r_i \qquad (5.11)$$

where n is the maximum number of experimentally formed co-crystals and r is the respective rank of each coformer, as produced by that specific route. A low ranking score would support the appropriateness of the respective route. Such score was also introduced so as to have an alternative measure which is not influenced by any cut-off point value. In order to facilitate the comparison of rankings, in particular between results associated with settings excluding and including carbon HB donors, caffeine was removed from the list. Furthermore, the scores produced from solid-state coformer molecular conformations for trans aconitic acid and pyridoxine, were evaluated by taking an average of their scores obtained from other settings. This was done so as to ensure an equitable ranking score comparison which require sets of an equal size. In case of GCV trials, one should note that the ranking scores have to be compared only between settings of the same solvent, given that GCV-nicotinamide co-crystal only formed when water solvent was used.

The rankings based on different inputs were relatively diverse, as opposed to what was observed for the *COSMO* approach. However, for both antivirals, the variance between rankings can be majorly attributed to which interactions were taken into account (see figure 5.16 and 5.17). Such influence was observed more markedly in cases where the CH of either components contributed to a different extent in HB interactions within the co-crystal relative to pure coformer. For example, in case of GCV-mellitic acid co-crystal, the carbon donors of the API were all engaged in contacts with the hydroxyl oxygen atoms of mellitic acid, as opposed to their weaker extent of involvement in HB network of pure GCV. Despite having several similar scores, the best settings were attributable to the route involving the extended conformation of GCV coupled with dominant coformer conformations in solution (water/ DMF), taking into account only conventional interactions, in agreement with the best overall success rate. This route classified five out of six experimentally formed co-crystals within the top seven ranks, and was supported by both binary response and ranking score. Similar to MC approach, the ranking based on ΔE values illustrated how various routes (2 and 4) exposed the favourability of correct co-crystals, even despite having lower success rates.

Results based on the cut-off value in figure 5.16 demonstrated how a higher number of potentially successful co-crystals were obtained when using the extended solid-state molecular representative for GCV. Conversely, given that the dominant conformation for GCV in solution was folded, less HB participants were accessible, and this led to an overall increase in the energy value associated with the potential co-crystals, E_{cc} (see appendix). This effect was of a relatively lesser extent for urea, thereby resulting in a considerable increase in its respective co-crystal ranking when employing such GCV conformer (dominant in solution). The ranking for urea was also high for FCV trials, mostly due to its comparatively strong large number of hydrogens which can contribute in intermolecular contacts. Furthermore, the intermolecular energy in pure urea coformer, E_{COF} was rather small due to the presence of only one HB acceptor, thus leading to a highly negative ΔE. This example highlights a limitation of this approach, which does not always taken into account the most favourable HBC number, thus leading to a potential underestimation of energy values. Predictions associated with the formation of GCV-nicotinamide co-crystal did not vary when using dominant conformations in diverse solvents, as opposed to experimental evidence and MC predictions.

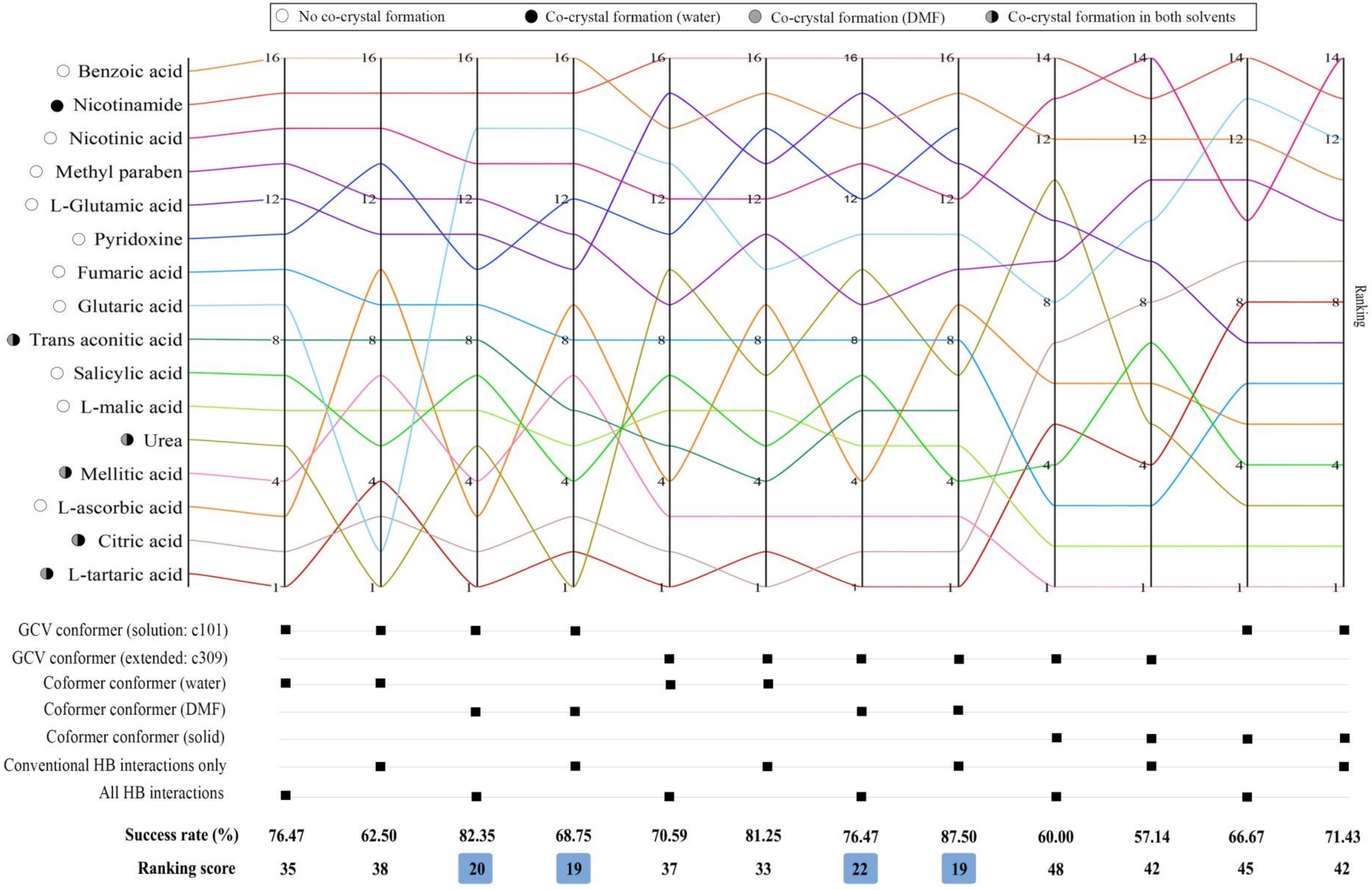
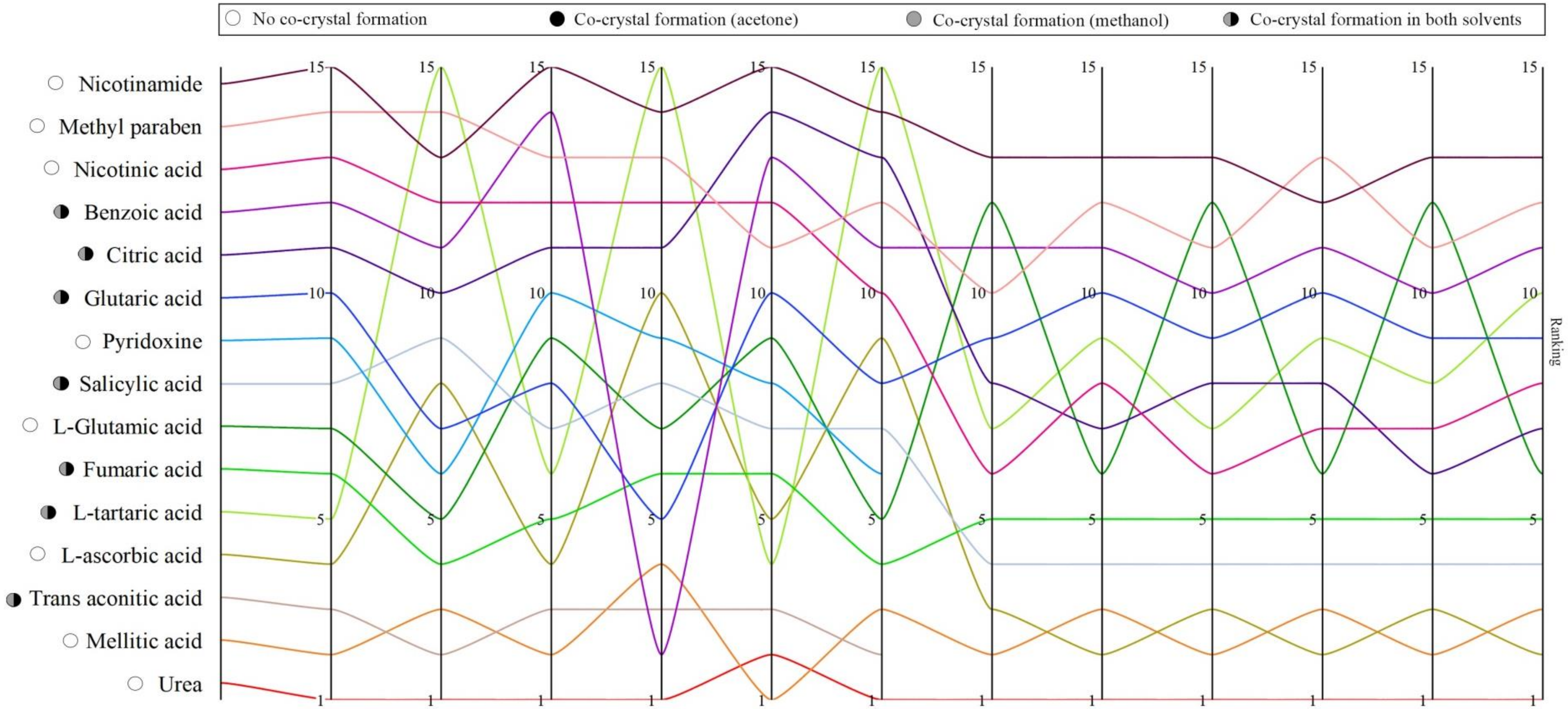


Figure 5.16: A parallel plot showing the ranking for the potential GCV co-crystals (1:1) based on ΔE values, with rank one being the most favourable (similar to previous parallel plots). Circles adjacent to the coformer names illustrate experimental results with the legend on top. Ranking scores shaded in blue correspond to routes in DMF solvent, and these values should not be compared with the rest of the scores (five co-crystals only).



FCV conformer (acetone: c47)	■	■							■	■		
FCV conformer (methanol: c165)			■	■							■	■
FCV conformer (extended: c32)					■	■	■	■				
Coformer conformer (solution)	■	■	■	■	■	■						
Coformer conformer (solid)							■	■	■	■	■	■
Conventional HB interactions only		■		■		■		■		■		■
All HB interactions	■		■		■		■		■		■	
Success rate (%)	**50.00**	**53.33**	**50.00**	**66.67**	**50.00**	**53.33**	**42.86**	**38.46**	**35.71**	**28.57**	**35.71**	**28.57**
Ranking score	**55**	**58**	**54**	**50**	**57**	**60**	**57**	**59**	**57**	**60**	**56**	**59**

Figure 5.17: A parallel plot showing the ranking for the potential FCV co-crystals (1:1) based on ΔE values, with rank one being the most favourable (similar to previous parallel plots). Circles adjacent to the coformer names illustrate experimental results with the legend on top.

Figure 5.17 displays the rankings for potential FCV co-crystals, based on same routes described earlier. The results associated to the extended conformation of FCV can be attributed to both solvents given that the dominant conformations in solution for coformers were the same in both methanol and acetone. Given the absence of abundant HB donors, the intermolecular energy, E_{API} of FCV in all conformations was very small compared to the values for GCV.

The overall outcome from the assessment of routes with respect to experimental results exhibited the general difficulty of producing correct predictions for FCV co-crystals. Such conclusion was drawn from the high ranking scores, all of which were considerably distant from the perfect score of 28, and low success rates, even relative to those obtained for GCV trials. Such difficulty might be related to the lack of conventional HB donors in FCV, which makes it more complex to predict the pairings of HB participants. The choice of interacting sites caused significant alterations in the rankings, particularly for l-ascorbic acid, l-tartaric acid, l-glutamic acid and benzoic acid.

On comparison, the best route for FCV co-crystals involved the dominant molecular conformations in solution for both API and coformer, whereby combining the outcomes of both trials in acetone (53.33%) and methanol (66.67%) generated a success rate of 60%. Such route has the same settings as those which parameterised the optimal route for FCV trials based on MC approach. This observation supports the premise regarding the correlation between solubility and dominant molecular conformations during LAG, described earlier. This route excluded the integration of carbon as a HB donor, which was rather unexpected especially due to its contribution in the pure FCV polymorphs. The favourability of this route was mostly increased by the proper ranking of benzoic acid coformer. Routes involving the solid-state molecular conformations for coformers performed the most poorly, with their top three ranks containing unsuitable coformers (urea, mellitic acid and l-ascorbic acid).

5.3.2.4 HBP models approach

In general, HBP model approach is utilised for the assessment of the risk of polymorphism, as displayed in chapter four, however its method of application enables it to be employed also for the assessment of the favourability of co-crystallisation. This is possible because the procedure deals with sub-structure functional groups rather than the inputted molecule as a whole. The model generates a propensity value for specific HB pair, irrelative of whether they are located within the same molecule or otherwise. As opposed to other approaches, this procedure provides the likelihood of every possible contact, hence allowing the inclusion of every available contribution towards the HB network. This feature offers the user more control with respect to the selection of interactions viable within the multicomponent system.

The details of each model per API-coformer combination, both including and excluding carbon as a HB donor, can be found in appendix. In some cases, larger training sets were required to ensure that each sub-structure was represented appropriately. Every acceptable model had an area under the ROC curve value above 0.77, to verify that the model was able to distinguish between true and false hits. The use of different conformations as inputs still led to almost identical models, with few differences in the HBC number propensities. An example regarding the calculation of the MD-score, detailing which groups were not accessible, is also displayed in appendix.

The elimination of fragments based on their accessibility for HB was at the centre of this procedure. However, this selection was not always very clear, especially for groups which were only partially available for HB, meaning that their orientation was not likely to allow the formation of a highly directional interaction. These cases were not treated differently by the model because each HB contributor was dealt with separately without considering the local chemical environment. In such scenarios, the molecular electrostatic potential map produced for the HBE approach, together with the extracted α and β parameters were consulted. Full interaction maps of specific conformers were also constructed to provide more information about the accessibility of such groups.

5.3.2.4.1 Overall performance

The application of this method for co-crystallisation is very often limited to the evaluation of the single differential (SD) score, which only takes into account the strongest pure and impure interactions (exc. CH donor). The hierarchal approach employed in this research study was aimed to provide improved predictions relative to the single differential scores, and hence for comparison purposes, the same cut-off point value of zero was used. Examples where the score fell exactly on the cut-off point, were regarded as inconclusive and hence excluded from the final success rates. However, it was still possible to include these results in the ranking scores. The route (no.4 in table 5.1) involving the dominant conformers in solution for the API and solid-state molecular conformations for coformers was excluded from this procedure. This was done because such route performed poorly in both previous approaches.

The routes considered for this approach, described in table 5.6, vary with respect to the list of unavailable HB participants. In case of following only model suggestions, the groups participating in highly probable intramolecular bonds were eliminated, whereas for other routes, such decision was solely based on the geometric orientation of each specific conformer. The list of available HB contributors based on the extended conformation of the API was also valid for their respective solid-state molecular conformation. Conformers with the same list of accessible HB participants produced equivalent outcomes, despite having discrepancies amongst their geometric orientation.

Table 5.6: GCV and FCV co-crystal prediction success rates (based on true positives and true negatives) based on SD/MD-scores with a cut-off value of zero. Only experimental results related to 1:1 stoichiometry were considered. Detailed results can be found in appendix.

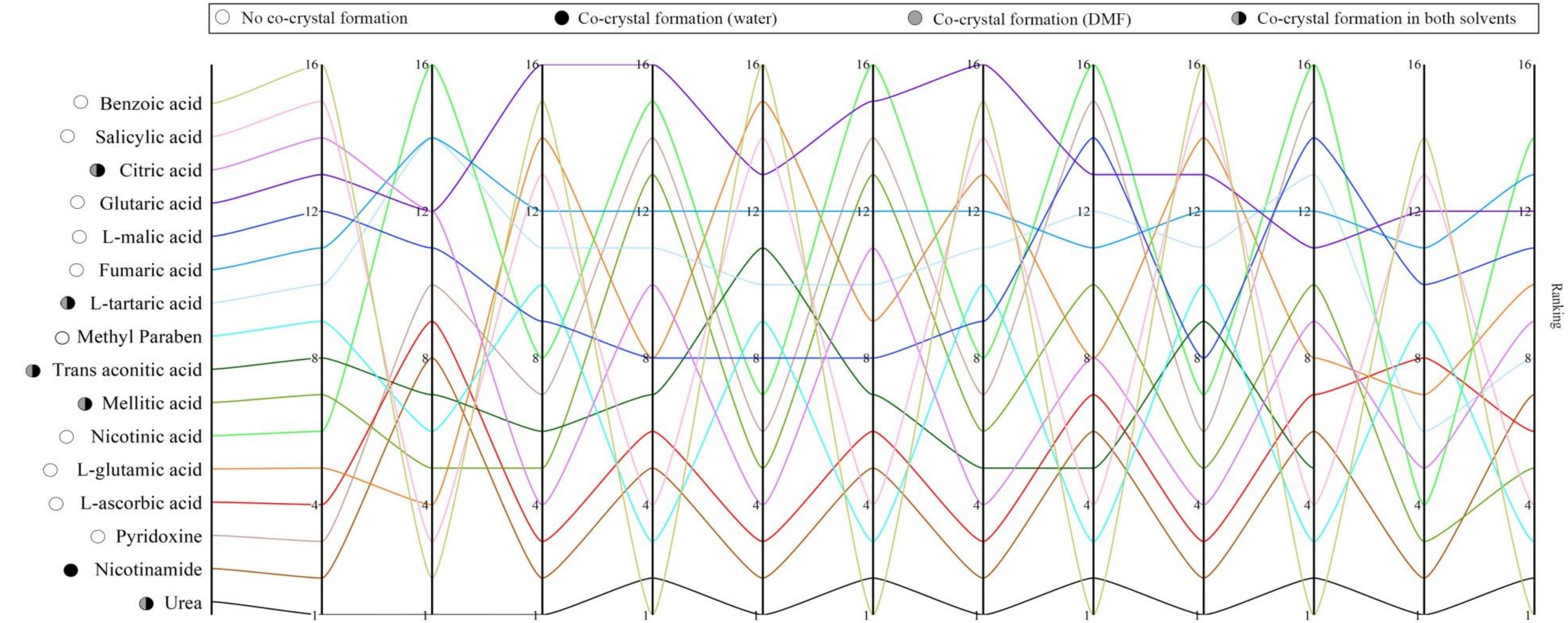
			MD-score			
Routes	API	**SD-score**	Model settings	Dominant conformation in solvent	Extended conformation	Extended conformation
	Coformer			Dominant conformation in solvent	Dominant conformation in solvent	Solid-state conformation
Inc. CH	**Success rate (%)**	/	50.00	49.23	54.00	58.62
Exc. CH	**Success rate (%)**	54.84	51.61	57.63	59.09	65.38

The scores presented in table 5.6 exhibited the overall less-than-average performance of this approach relative to the previous ones, with the highest success rate of 65.38%. The most successful route involved the extended solid-state molecular representatives of both co-crystal components (same settings of the best MC route), and excluding the contribution of carbon HB donor. The exclusion of such HB participant was supported by the success rates of all routes, as well as in agreement with most successful route for the HBE approach. It is noteworthy that the outcomes of the best route had a greater degree of accuracy compared to the standard calculation of the SD-scores, which contained multiple equivalent results. This occurred because several coformers shared multiple sub-structures, which also led to common strongest pure and impure interactions. Such outcome indicated that the integration of multiple interactions can provide improved predictions, as well as it is able to distinguish better between coformers. Furthermore, such route also performed better than the settings proposed by the model through the predicted intramolecular interactions.

5.3.2.4.2 Rankings

The most evident observation in the parallel plots for GCV (figure 5.18) and FCV (figure 5.19) trials is surely the alternating pattern of several coformer ranks. This pattern mirrored the choice of including or excluding carbon as a HB donor. Such outcome was also noticeable amongst the predictions of the HBE approach, but it was to a lesser extent. The effect of this choice prevailed in examples where the inclusion of CH donor enabled the integration of a significant number of interactions. For instance, in case of FCV-benzoic acid as per model suggestions, such inclusion increased the number of impure interactions from three to seven contacts, thereby augmenting the strength of the intermolecular framework of the potential co-crystal. The resultant impact of the increase in the number of contacts did not always yield a decrease in MD-score (depends on the competition between pure and impure contacts), as seen for FCV-fumaric acid.

The ability of HBP models to treat adequately CH fragments was already questionable after the analysis of their performance with respect to the risk of polymorphism in chapter four. Similar issues were also noted through discrepancies between *IsoStar* contour plots and full interaction maps related to carbon as a HB donor. During this application, evidence of overestimated propensities related to contacts involving CH donor frequently concerned intramolecular interactions. For example, this occurred in GCV-l-glutamic acid and GCV-glutaric acid trials, whereby even some contacts which were not geometrically feasible, had considerable probabilities. Similar overestimation was also noticed specifically in the intermolecular contacts involving the sub-structure of CH in an aromatic environment. This effect led to the ranking of benzoic acid, salicylic acid and methyl paraben coformers amongst the top spots for particular routes (included CH) related to both antivirals.



GCV conformer (solution: c101)			■	■	■	■						
GCV conformer (extended: c309)							■	■	■	■	■	■
Coformer conformer (water)			■	■			■	■				
Coformer conformer (DMF)					■	■			■	■		
Coformer conformer (solid)											■	■
Model suggestions	■	■										
All HB interactions		■		■		■		■		■		■
Conventional HB interactions only	■		■		■		■		■		■	
Success rate (%)	**56.25**	**52.94**	**73.33**	**52.94**	**60.00**	**41.18**	**73.33**	**52.94**	**60.00**	**47.06**	**92.31**	**60.00**
Ranking score	**42**	**47**	**29**	**48**	**31**	**43**	**29**	**43**	**30**	**39**	**27**	**40**

Figure 5.18: A parallel plot showing the ranking for the potential GCV co-crystals (1:1) based on MD-scores, with rank one being the most favourable (similar to previous parallel plots). Circles adjacent to the coformer names illustrate experimental results with the legend on top. Ranking scores shaded in blue correspond to routes in DMF solvent, and these values should not be compared with the rest of the scores (five co-crystals only).

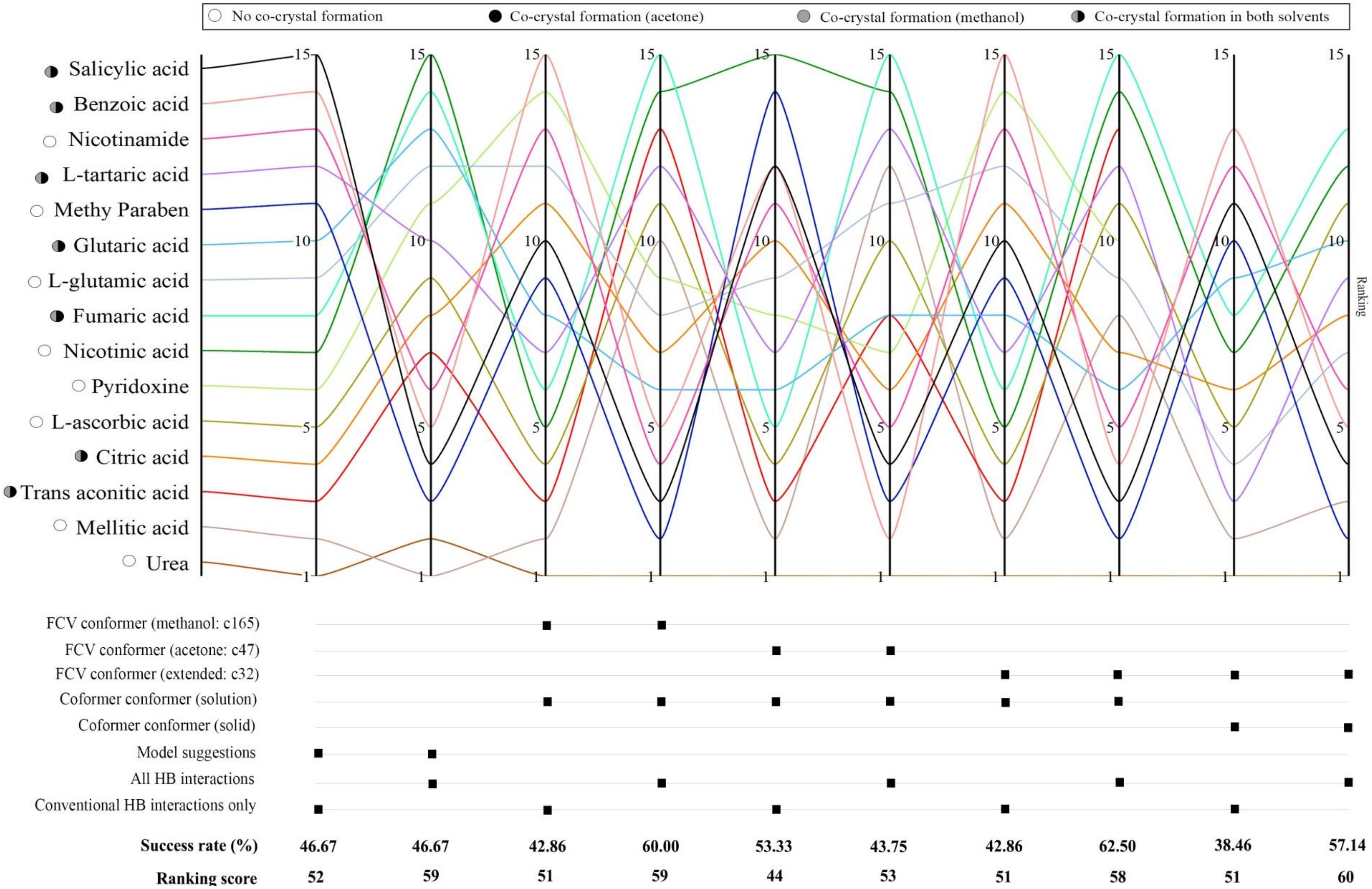


Figure 5.19: A parallel plot showing the ranking for the potential FCV co-crystals (1:1) based on MD-scores, with rank one being the most favourable (similar to previous parallel plots). Circles adjacent to the coformer names illustrate experimental results with the legend on top.

However, discrepancies between scores demonstrated that CH variable was the strongest but not the sole cause for variances between routes. These differences were caused by dissimilarities between the lists of accessible HB groups. In particular, this was evident in FCV trials, such as pyridoxine, whereby most rankings were not coherent even within the same CH settings sub-group. This was also noted for GCV examples as trans aconitic acid, l-malic acid, glutaric acid and mellitic acid. Given the lack of CH fragments and the relative abundance of strong HB donors, urea was the coformer whose ranks varied the least with the choice of interacting sites (CH).

As opposed to other approaches, this was the only one whereby the high favourability of urea as a coformer was exposed by all routes, with considerably negative scores. The case of urea was rather unique because even though it has only one type of pure interactions (N-H···O), such contact occurs more than once due to the substantial HBC number of the carbonyl oxygen atom. Structurally determined co-crystal structures containing urea as MOWKUF and OTIYUM, also exposed the significant capability of such oxygen atom to have multiple contacts simultaneously, with a minimum of two interactions. The HBC number of this oxygen atom within the potential co-crystals was adjusted according to the propensity values in the HBC table. However, this method was not able to distinguish between the true GCV-urea co-crystal formation and the false FCV counterpart.

Given the requirement to compare multiple routes, the same ranking score (*eq.5.11*) was employed. The considerations discussed in section 5.3.2.3.2, regarding trans aconitic acid, pyridoxine and caffeine, were also applied in the ranking score of this approach. Both ranking scores and binary responses exhibited how this approach performed much better for GCV rather than for FCV, particularly due to the presence of several false positive predictions related to the latter antiviral. Some routes generated the same scores for coformers whose conformations had the same accessible functional groups or which were composed of multiple identical sub-structures. For example, in case of citric acid, despite changing the list of unavailable groups according to each conformation, such HB participants were identical, and consequently, there was no difference in MD-scores.

The route with the highest success rate for GCV co-crystals had an excellent success rate of 92.31%, together with a ranking score of 27. This was in agreement with the overall results exhibited in table 5.6. This route ranked five out of six successful coformers within the top six ranks. The route involving the extended conformation for the API and the dominant conformation in solution (best overall route for HBE approach) was associated to the second best performance. Similar to the predictions produced by HBE approach, the routes in water and DMF environments generated equivalent outcomes for GCV-nicotinamide co-crystal.

The selection of accessible HB participants for FCV conformations presented to be a more complex task, relative to that encountered for GCV, especially when dealing with CH fragments. This occurred due to the bulky side chains of FCV, which hindered to an extent the accessibility of several groups. Despite following information extracted from the HBE energy parameters, as well as full interaction maps, some element of subjectivity might have introduced a bias within the procedure. On comparison, one can notice that certain routes, particularly those which excluded carbon as a HB donor, yielded many negative MD-scores, which led to several false positive results. This occurred because the exclusion of such fragment caused the reduction of pure interactions within FCV, thereby underestimating the strength of its intermolecular network. The advantage of including CH was exhibited through the success rate (62.50%) of the route involving the extended solid-state molecular representative for FCV and dominant conformer in solvent for coformer. However, the lowest/ best ranking score was associated with the route whereby all inputted conformers were dominant in a solution environment and excluding CH as HB donor, similar to the observations for FCV in MC and HBE approaches. These results led to an inconclusive outcome for FCV trials, particularly with respect to CH inclusion / exclusion.

5.3.2.4.3 VCV·HCl co-crystal predictions

Given the sensitivity of the procedure of prediction, as well as the significant contribution of a chloride ion towards HB network (shown in chapter four), it was crucial to include its presence within the process of evaluation. However, the conformational analysis of a salt involves a more complex task than when dealing with neutral single molecules, because it needs to account for the charges and how both components can influence the conformation of each other. Even though, as described earlier, the conformational differences for VCV·HCl relied on VCV molecule, the presence of the chloride ion still has an impact on its energetic stability. Conformer generators employed for the other antivirals, as RDkit and the CSD generator are not equipped (or at least by default in case of RDkit) to handle multicomponent systems. A study on a good sized training set is required to examine different routes of conformational analysis for VCV·HCl. This work goes beyond the primary objective of this project, and therefore only one conformation of VCV·HCl was employed for co-crystal prediction.

The experimental co-crystal screening for VCV·HCl was performed in an aqueous environment, hence implying that the dissociation of the salt was retained in solution, as well as in the obtained co-crystals. Given the additional complexity, coupled with the lack of published assessments, the co-crystal prediction for VCV·HCl antiviral was conducted only using HBP models and the conformation in figure 4.40 as an input. This method was chosen because of its ability of deal with sub-structures rather than whole molecules. Moreover, given that no information was available regarding the actual orientation in solvent or in solid, the availability of each fragment was allowed to be determined by the model, through the propensity values for intramolecular bonding and the HBC table.

Despite that the salt is composed of two distinct components, it was still treated as one component, whereby pure interactions also included contacts between the chloride ion and VCV groups. This was done so as to employ the same procedure devised earlier. In order to do so, it was required to assume that neither of the salt components would form a co-crystal without integrating each other in the crystal form. This was ensured by assuming that in every potential scenario, the N^{+}-H···Cl^{-} interaction would always be present. This assumption was supported by the large propensity of the N^{+}-H···Cl^{-} interaction ($\mathbb{P} \sim 1$) in all trials, which suggested that the formation of such contact was inevitable even in the presence of a coformer. Even though the chloride ion can form more than one interaction with charged amino group, this likelihood decreases significantly in the presence of other viable donors as NH. In fact, only 24.9% of hits composed of both neutral and positively charged NH groups, contained a chloride ion with two or more contacts with the charged counterpart. This tendency may decrease further with the introduction of other competing donors present in the coformer. Due to this information, the N^{+}-H···Cl^{-} interaction was assumed to be present always once, and hence removed from the list of potential contacts and consequently omitted from the MD-score. Nonetheless, the HBC numbers of the charged amino group and chloride ion enabled them to participate in other pure and impure interactions.

The predictions of co-crystallisation are presented in table 5.7, together with the corresponding success rates and ranking scores. The favourability of mellitic acid and urea as coformers was highlighted in multiple routes. The highest success rate (81.25%) corresponded to the employment of the SD-score, primarily due to its true negative outcomes. The positive SD-scores occurred because the pure interaction of N-H···Cl$^-$ had the largest propensity value ($\mathbb{P} \sim 0.97$). In fact, only coformers with competing amine groups, namely urea and nicotinamide, were considered as favourable by the SD-score. It is evident that the outcomes were highly influenced by the propensity values concerning the chloride ion. The presence of such sole dominant contributor might have led to the satisfactory performance of this route (SD-score) for VCV·HCl, given that it deals only with the strongest interactions. On the other hand, this measurement of favourability was not suitable to cater for the competition of multiple similar HB contributors in FCV and GCV trials.

Table 5.7: Predictions of the formation of VCV·HCl co-crystals using HBP models.

		Model settings		
		Exc. CH	Exc. CH	Inc. CH
Experimental outcome	**Coformers**	**SD-score**	**MD-score**	**MD-score**
✓	L-ascorbic acid	0.02	-2.2	-0.39
✓	Mellitic acid	0.02	-1.45	-1.5
×	Pyridoxine	0.04	-0.85	-0.12
×	Fumaric acid	0.04	-0.73	-0.56
✓	Urea	-0.01	-0.61	-0.28
×	Nicotinic acid	0.04	-0.55	0.54
×	Trans aconitic acid	0.04	-0.53	0.25
✓	Citric acid	0.05	-0.52	-0.25
×	Glutaric acid	0.03	-0.51	-0.46
×	L-malic acid	0.04	-0.49	-0.36
×	L-tartaric acid	0.04	-0.39	-0.28
×	Nicotinamide	0	-0.32	-0.42
×	Methyl Paraben	0.05	-0.26	-0.09
×	Benzoic acid	0.04	-0.21	-1.03
×	Salicylic acid	0.04	-0.17	-0.53
×	L-glutamic acid	0.02	-0.07	0.26
×	Caffeine	0.28	NA	-0.72
Success rate (%)		**81.25**	**31.25**	**41.18**
Ranking Score			**16**	**28**

Conversely, the integration of other contacts led to the negative MD-scores, which yielded poor success rates. These scores resulted in negative values because in several trials, the propensity of an impure contact involving the chloride ion out-competed the likelihood of the associated pure interaction within the coformer. However, similar to previous cases, the poor behaviour of the hierarchal routes suggested by success rates might have been caused by the utilisation of an inappropriate cut-off point value. Indeed, this premise was supported by the relatively small ranking score associated with the hierarchal route excluding CH fragments, which classified the four successful coformers amongst its top half ranks. The ranking score of the SD-scores was not computed due to the large amount of ties present. Nonetheless, even by rough consideration, the unfavourable rank of citric acid would have increased its ranking score by a significant extent.

The performance of this approach for VCV·HCl might have been hindered by the assumptions made during the formulation of the methodology. Further work is required to investigate the potential impact of using specific input conformations, as conducted for the other antivirals. However, the obtained outcomes show considerable promise for the application of this method for co-crystallisation prediction.

5.4 Outcomes

This chapter revolved around the co-crystallisation of GCV, FCV and VCV·HCl with potential coformers, whereby computational predictions were validated against experimental outcomes. These procedures were also performed for VCV (without HCl), but experimental data is required to assess the quality of predictions. Lab work results demonstrated how the molecular size of the coformer was not a crucial determinant factor with respect to the favourability of formation. Conversely, the relative population of HB contributors with respect to the total number of atoms, remained an influential force, with the successful co-crystals of mellitic acid, l-tartaric acid and urea as evidence. This analysis also demonstrated how the strong pure interactions within VCV·HCl, particularly those involving the chloride ion, might have reduced its likelihood of form co-crystals.

Four approaches were applied while employing diverse routes in attempt to identify the optimal conditions for prediction. The HBE approach resulted to have the highest overall success rate of 74.47%, relative to MC and HBP model approaches, which yielded a rate of 68.97% and 65.38%, respectively. The use of an extended solid-state molecular representation as an input for GCV trials was supported by all approaches, coupled with the exclusion of carbon as a HB donor. However, it was also shown how, even though the highest success rate favoured particular routes, it might not imply that such route is the optimal choice for all types of molecules. In fact, it was exposed by MC and HBE approaches, how in case of FCV, better results were generated when the dominant molecular conformations in solution were employed as inputs. Furthermore, more research work is required to have a concrete deduction regarding the impact of the inclusion of CH fragments with respect to FCV trials.

Despite not being able to determine the optimal type of input in *COSMOtherm*, the importance of using a set was exposed to be correlated with the diversity in σ-potential graphs of the conformers of components. Such observation was made through the comparison of results between GCV and FCV, as well as through the analysis of particular conformations of coformers, as l-tartaric acid. Further work should include the application of similar routes, but focusing more on the coformers, rather than only on the API conformers. The rankings based on H_{ex} values produced by *COSMOquick* resulted to be fairly comparable with those corresponding to more intensive methods. The predictions obtained from the COSMO-RS approach were adequate in quality, particularly for FCV co-crystals, excluding results based on f_{fit} or Hansen parameters.

Minimal differences in molecular conformations yielded discrepancies in the rankings produced by MC approach, as opposed to the *COSMO* approach which was much less sensitive. Given the adjusted criteria, most of the routes provided true negative outcomes, but the ranking system aided to expose the favourability of suitable coformers (true positives). The integration of the results from both methods of route assessment led to a more accurate view of the likelihood of co-crystallisation, with the binary response highlighting true negatives and ranking system exposing true positives. Such observation was also noted for the HBE approach.

Even though the HBP models approach resulted to be the weakest amongst the other approaches, the new considerations applied in the hierarchal procedure related to specific conformations, yielded improved predictions. This was particularly evident with respect to the single differential score for GCV and FCV, which was a standard method for co-crystal prediction applied so far in literature. Furthermore, given the integration of multiple interactions, the hierarchal method reduced the likelihood of generating equal results for multiple co-crystal systems, as opposed to the outcomes of the SD-scores. This enabled a clearer ranking of coformers. The utilisation of the SD-score was only favourable in case of having one significantly dominant HB contributor, as in the case of VCV·HCl.

The validity of the implications derived from this study are bounded to the set of specific parameters applied, particularly the selection of coformers which have similar functional groups and chemically related API molecules. Moreover, the experimental results were only obtained from one experimental route for co-crystal formation. This means that further investigation including a wider range of chemically diverse molecules and different experimental techniques is required to standardise the conclusions of this study. In addition, a larger sample size would have enabled the application of statistical analysis so as to validate quantitatively the performance of each approach.

Chapter 6

CONCLUSION

6.1 Objectives and outcomes

This study showcased the application of various techniques, both computational and experimental to characterise the crystallographic behaviour of three prominent antivirals, namely GCV, FCV and VCV·HCl. The use of *in silico* methods was intended to provide satisfactory predictions and also to assist in formulating a better understanding of specific mechanisms which are commonly encountered in the crystallographic world. These mechanisms concern molecular flexibility and its subsequent quantification, polymorphism and co-crystallisation. This work incorporated a combination of statistical and thermodynamic approaches, all of which provided collectively a more reliable insight of the investigated processes, thereby complementing experimental work. The development of novel and modified methodologies was aimed to either address gaps in literature or to improve and broaden the applicability of published procedures. On the other hand, experimental work was crucial for the characterisation of formed crystalline materials and for the assessment of computational techniques.

6.1.1 Exploring flexibility

This work revolved around the exploration of molecular flexibility and its adequate quantification, beyond the utilisation of the NRot variable. This study was required because the NRot parameter is considered as a standard measure for flexibility, and employed in various applications. However, this conformational analysis highlighted how even though the NRot has a considerable impact on the extent of molecular flexibility, it is still not accurate enough to represent such property exhaustively. This was exhibited through two novel approaches which analysed the molecular flexibility of the antivirals, qualitatively and quantitatively.

The investigation showed how, as opposed to the assumption employed by the NRot, each rotational bond has its own distinct impact on the conformational space of the molecule. This impact is influenced by the nature of the components involved in the torsion angle and also by long-range chemical environment. Such variation of flexibility was firstly observed through the different profiles exposed by the torsional angle histograms, extracted from the CSD. These histograms led to the evaluation of flexibility scores, whose relative differences were not in agreement with those related to the respective NRot. Individual descriptors and flexibility scores demonstrated that FCV is not characterised by a greater extent of molecular flexibility relative to VCV, as suggested by its NRot. Moreover, the multivariate test for the homogeneity of variances proved how the discrepancy between the extent of flexibility of VCV and FCV is statistically non-significant, despite having different NRot. This result was also supported by the fact that the histograms of the torsional data showed how in fact both VCV and FCV have seven considerably flexible bonds each. On the other hand, the additional rotational bond of GCV relative to ACV, induced a significant extension in its accessible conformational space.

The validation of the scoring system requires a more diverse and larger training set to be adequate and complete. However, the obtained scores were in line with the outcomes extracted from the Hirshfeld surface descriptors, the average distances from data points to median per antiviral and also with the Kier flexibility index. The diversity of this collection of agreeing sources promotes the robustness of the developed flexibility scoring system. The advantage of this scoring system in relation to other existing flexibility measures revolves around the integration of data from structurally determined crystal structures, hence increasing the favourability of the results. Moreover, it provides a definite measure of flexibility per rotatable bond and is able to differentiate between regio-isomers. The continuous growth of the CSD projects better future results as more precise hits can be obtained, thus reducing bias and increasing certainty.

6.1.2 Polymorphism

This chapter presented a binary study of the administered forms of the three antivirals, whereby experimental and computational outcomes were utilised to provide a detailed characterisation, particularly in relation to thermodynamic stability. Solubility screening and thermal analysis exhibited the thermodynamic stability of GCV form I at room temperature, in line with literature. Computational work supported this claim through the outcomes from crystal structure analysis and energy calculations. Energy frameworks highlighted the extensive supramolecular network in form I, characterised by contacts with strong extent of directionality, yielding a higher degree of compactness and lower enthalpy compared to form II. The variance in the strength and nature between the intermolecular bonding framework of forms I and II was shown to be a significant contributor towards their diverse thermodynamic behaviours, thereby leading to the manifested conformational polymorphism.

The lack of newly discovered forms and the confirmation of the thermodynamic stability of form I at room temperature indicated that the likelihood of the formation of other anhydrous GCV forms is minimal. The prediction of the risk of polymorphism for GCV was conducted using two approaches with distinct setups, namely CSP and HBP model. Both methods predicted correctly the high feasibility of forming polymorph II experimentally, relative to other potential forms. However, they both failed to represent faithfully and capture the stability provided by the collective contributions of all non-bonding interactions in form I. Furthermore, on comparison with experimental work, the CSP approach yielded better results, whereby the considerable energy gap between form II and most of the hypothetical forms reinforced the claim of having minimal risk of polymorphism. However, these results would have much greater validation if form I was generated and ranked during the CSP procedure.

The thermodynamic stability of FCV form I was exposed through its dominant presence after re-crystallisation from various solvents, both protic and aprotic. In addition, such thermodynamic stability was also supported by the thermal analysis results, whereby all analysed forms transformed into form I with heating. Crystal structure analysis demonstrated how the non-bonding framework of form I is optimally constructed to maintain such stability, with the integration of C-H···X contacts and $\pi\cdots\pi$ interactions. It was also shown how the orientation of bulky branches led to a more spacious packing so as to reduce intramolecular strain. However, the thermodynamic stability of form I was not predicted correctly by the HBP model approach for the risk assessment of polymorphism. Similar to the case of GCV form I, the degree of stability provided by non-conventional contacts, particularly those involving carbon as a HB donor, was not taken into account by this approach. This prevailing limitation clearly indicated that the application of this approach should be restricted to molecules whose HB networks are likely to be governed by conventional HB interactions, as such VCV·HCl.

The solubility screening of VCV·HCl and the subsequent analysis of the produced samples exhibited the presence of various mixtures, as well as the expected characteristic of low crystallinity. PXRD data and STA results proved the formation of a new VCV·HCl hydrate form, which similarly to other hydrates, transformed into form I with heating. Experimental work and literature indicated that despite being sensitive to certain conditions of crystallisation and relative humidity above 75%, the thermodynamic stability of pure form I is considerable, particularly relative to anhydrate form II and hydrates. Nonetheless, a more exhaustive screening is required for confirmation, especially in relation to other patented VCV·HCl forms with less than 2% water content (considered as anhydrates), thereby making its risk of polymorphism still relevant.

The propensity chart for VCV·HCl was populated by a large number of hypothetical forms within the optimal region. This occurred because VCV·HCl is characterised by a fairly balanced ratio of HB acceptors and donors, coupled with HB participants with high HB coordination number and considerable molecular flexibility. These characteristics enabled the formation of multiple variations of HB frameworks which provided a significant degree of stability. The presence of highly similar generated HB networks also implied that some forms might have comparable energies, thus increasing the likelihood of polymorphism. Moreover, such feature is even more heightened in the presence of water due to the inclusion of more HB contributors, as proved through the formation of multiple distinct hydrates.

The investigation of these three antivirals displayed how polymorphism is governed by the present degree of stability, which in return is dependant on the extent of intramolecular strain and the intermolecular non-bonding network. Subsequently, these factors are influenced by the HB donors to acceptors ratio, HB coordination number per group and the strength of other non-bonding contacts, beyond HB. An increase in molecular flexibility might lead to a larger group of hypothetical polymorphs, but it does not dictate the number of forms which are energetically feasible and pose an actual risk of polymorphism. Conversely, such likelihood is enhanced in cases where either an improved or alternate feasible polymorphic arrangement is achievable, without significant energy barriers, as suggested for VCV·HCl.

6.1.3 Co-crystallisation

The concept of using various approaches for co-crystal screening has been applied in multiple case studies, but the methodology presented in this work was modified to take into account further possibilites.[155, 180, 317] These modifications were necessary due to the weaker performance of standard procedures when applied to larger API molecules. Besides obtaining accurate predictions for co-crystal screening, it was also vital to analyse how each approach and its respective results behave with different conformer inputs. This analysis was required because the ranking of coformers did not only depend on the number and nature of HB contributors, but also their respective availability. Four approaches with different frameworks were applied, each of which employed diverse routes in attempt to identify the optimal conditions for prediction.

The ranking based on the excess enthalpy evaluated through the COSMO approach, resulted to be an adequate indicator of the suitability of coformers, especially for FCV co-crystals. However, this approach seems to be much less appropriate to provide correct predictions for ratios other than 1:1. The utilisation of a conformer set as an input provides assurance that conformations other than the one with the lowest energy are taken into account, hence increasing the probability of including the most stable conformer in solution. It was noted that the importance of such input is correlated with the diversity in the σ-potential graphs of the conformers of pure components. As opposed to Hansen solubility parameters and f_{fit} values, *COSMOquick* was able to produce relatively adequate rankings, at a much lower computational cost.

The predictions concerning mellitic acid, not only showed favourable co-crystal formation, but the relatively large extent of negative excess enthalpy also illustrated the favourability of producing a co-crystal whose solubility is higher than that of the respective API. This was confirmed experimentally by the work done by another research group member.[318] The NRot parameter was proved to be a misleading variable with regards to its inclusion in co-crystallisation prediction through the formation of GCV-citric acid co-crystal, due to its overestimation of the flexibility of coformer. An improvement in prediction was obtained when the novel flexibility score replaced the NRot variable.

The highest overall success rate was associated to the HBE approach at 74.47%, followed by the MC (68.97%) and the HBP model (65.38%) procedures, all of which performed better for GCV trials. The resultant observations displayed the advantage of customising the settings of each approach according to the molecule under study. While the use of an extended solid-state molecular representation was favourable for GCV trials, improved predictions were generated when dominant conformers in solution were employed as inputs for FCV runs. This outcome suggested that opposed to GCV, it was more energetically feasible for FCV to adopt a stable conformation in solution, which is a direct implication of its high solubility. This displayed how the extent of solubility and the nature of solvents used, might have the potential to guide the process of selecting the optimal inputs for the prediction of LAG co-crystallisation.

Besides the testing of different inputs, it was also necessary to investigate the impact of integrating interactions other than the strongest pair, in attempt to capture a more representative view of the competition present during potential co-crystallisation. Such integration yielded a fairly good success rate for the HBE approach, and improved predictions for the HBP model methodology. The only exception where the SD-score performed better concerned cases with a distinctively significant dominant HB contributor, namely the chloride ion in VCV·HCl trials. Furthermore, routes were also set to analyse the effect of taking into account CH participation in HB. The exclusion of this HB donor for GCV runs was supported by all approaches, but the indications for FCV led an inconclusive outcome.

The utilisation of a ranking system proved to be beneficial to remove the impact of inappropriate cut-off values employed in binary responses. This was evident in both MC and HBE outcomes, whereby the ranking system was essential for the exposure of true positive results. Such advantage was also noted for the VCV·HCl co-crystal prediction, whereby the comparatively low ranking score of a specific hierarchal route exhibited its ability to rank coformers adequately, despite having negative MD-scores.

6.1.4 Conclusion and achievements

The outcomes described above entailed a detailed characterisation of solid-state behaviour of GCV, FCV and VCV·HCl, with particular attention with respect to their administered form, using both experimental and computational methods. This characterisation extended from a molecular to a supramolecular level, and enfolded further into the investigation of co-crystallisation, thereby fulfilling all the objectives of this project. Despite having similar backbones, all antivirals resulted to be characterised by unique features at all molecular levels, which ultimately explain their distinct physical properties.

The study of molecular flexibility led to the formulation of a new scoring system, which provided improved results relative to the NRot variable. This new flexibility score has the potential to be applicable for a wide range of molecules. The methodology concerning polymorphism, exposed the benefits of the integration of various tools, each probing at different angle to investigate thermodynamic stability. The accumulation of the results provided an exhaustive view of the factors which interplayed together and manifested the characteristics of each form. This analysis contributed towards having a better understanding of such interplay which dictates the stability of each crystalline material. The presented framework of procedures resulted to be a robust pathway to investigate any organic crystal structure. The developed methodologies for the prediction of co-crystallisation, together with the inferences made through experimental work, also show sufficient promise for a wider application.

6.2 Strengths

One of the strongest pillars of this work rests on the utilisation of both experimental data and computational work. This methodology enabled the assessment of particular *in silico* procedures, hence providing a clearer picture of their respective behaviour and applicability. This was especially useful during the analysis of new / modified co-crystallisation prediction methods. This work also illustrates how such combined approach can assist in obtaining a more holistic view. Furthermore, the utilisation of experimental information within theoretical methods proved to be advantageous, as demonstrated through the new flexibility scoring system, which utilised geometric data from experimentally solved structures. Despite further testing is required, this flexibility score performed very well on the set of molecules presented here, also in line with the observations made through the statistical analysis of Hirshfeld surface descriptors.

This study showcases a diverse range of computational tools, thus enabling the investigation of every objective more exhaustively. Such advantage was illustrated through the results obtained from each method, which facilitated the understanding of the underlying effects. This was particularly displayed through the analysis of the thermodynamic stability of the antivirals, which provided a detailed, reliable insight by combining the outcomes from *IsoStar* contour plots, full interaction maps, Hirshfeld surface analysis and energy frameworks. In addition, the use of multiple approaches also enabled the comparison of their results, thus providing more confidence with respect to the reliability of the derived conclusions when in agreement. Such multitude of routes together with other considerations involving carbon as a HB donor and molecular flexibility, made the methodology more exhaustive and robust.

6.3 Limitations

The most significant limitation within this study concerns the sample size and lack of diversity with respect to the chemical nature of the chosen antivirals. This restriction was applicable to all results which were obtained from newly developed or adjusted methodologies, primarily those concerning the quantification of molecular flexibility and co-crystallisation prediction. Despite still being a significant indication of performance, a larger sample would enable a proper validation process using statistical testing. Moreover, the extrapolation of outcomes is only applicable to molecules with similar chemical environments as the investigated antivirals. The absence of determined VCV·HCl crystal structures limited the extent of analysis in chapter four.

Another limitation concerns the experimental trials performed throughout this project, particularly concerning co-crystal screening. Due to time restrictions, only one route of such screening was employed, thereby potentially reducing the exhaustiveness of the trials. Moreover, a larger list of coformers would have supported the observations noted for the three antivirals. Another experimental limitation regards the lack of controlled environmental conditions during vt-PXRD runs which might have caused some discrepancies in resultant patterns, and thereby obstructing the comparison with literature or STA results. This was particularly evident in case of GCV form II and VCV·HCl polymorphs. Furthermore, the absence of purification steps, especially concerning various mixed samples of VCV·HCl hindered the process of characterisation.

6.4 Further work

Given the above limitations, it is imperative to invest time in applying the developed methodologies to a more diverse training set in order to test whether their performance is satisfactory and the observations are still valid. This is particularly relevant for the testing of the flexibility score and co-crystallisation prediction methods. Even though it was conclusive for GCV co-crystal trials, the exclusion of CH as HB donor still requires further testing to be confirmed.

Another crucial objective would be the crystal structure solution of VCV·HCl form I so as to validate the predictions inferred regarding its non-bonding network. Further investigations would entail a more detailed analysis towards the role of $\pi\cdots\pi$ interactions within all three antivirals.[326] In addition, it would be interesting to conduct CSP for GCV using a different computational algorithm, such as GRACE, to test whether the generation of form I would be attainable. Such procedures can also be applied to FCV and VCV·HCl, due to the outlined limitations of the HBP models approach. Despite hinted in literature, more exhaustive studies are needed to monitor the kinetic stability of the administered forms, particularly FCV form I, in high relative humidity environments. Solubility tests are required to be conducted for the newly formed co-crystals, so as to check whether they can classify as potential candidates to improve the pharmacokinetic properties of the present administered API form.

REFERENCES

1. Hilfiker, R.; Blatter, F.; Raumer, M. v., Relevance of Solid-state Properties for Pharmaceutical Products. In *Polymorphism*, Wiley-VCH Verlag GmbH & Co. KGaA: 2006; pp 1-19.

2. Berdigaliyev, N.; Aljofan, M., An overview of drug discovery and development. *Future Medicinal Chemistry* **2020,** *12* (10), 939-947.

3. Koenig, S. G.; Leahy, D. K.; Wells, A. S., Evaluating the Impact of a Decade of Funding from the Green Chemistry Institute Pharmaceutical Roundtable. *Organic Process Research & Development* **2018,** *22* (10), 1344-1359.

4. Abramov, Y. A., Computational Pharmaceutical Solid-State Chemistry. In *Computational Pharmaceutical Solid State Chemistry*, John Wiley & Sons, Inc: 2016; pp 1-13.

5. Wu, F.; Zhou, Y.; Li, L.; Shen, X.; Chen, G.; Wang, X.; Liang, X.; Tan, M.; Huang, Z., Computational Approaches in Preclinical Studies on Drug Discovery and Development. *Frontiers in chemistry* **2020,** *8*, 726.

6. Poole, C. L.; James, S. H., Antiviral Therapies for Herpesviruses: Current Agents and New Directions. *Clin Ther* **2018,** *40* (8), 1282-1298.

7. Chemburkar, S. R.; Bauer, J.; Deming, K.; Spiwek, H.; Patel, K.; Morris, J.; Henry, R.; Spanton, S.; Dziki, W.; Porter, W.; Quick, J.; Bauer, P.; Donaubauer, J.; Narayanan, B. A.; Soldani, M., et al., Dealing with the Impact of Ritonavir Polymorphs on the Late Stages of Bulk Drug Process Development. *Organic Process Research & Development* **2000,** *4* (5), 413-417.

8. Recalled Transdermal PD Patch Should be Down-Titrated *Neurology Today* [Online], 2008. (accessed on February 2021).

9. Gao, Z.; Rohani, S.; Gong, J.; Wang, J., Recent Developments in the Crystallization Process: Toward the Pharmaceutical Industry. *Engineering* **2017,** *3* (3), 343-353.

10. Couillaud, B. M.; Espeau, P.; Mignet, N.; Corvis, Y., State of the Art of Pharmaceutical Solid Forms: from Crystal Property Issues to Nanocrystals Formulation. *ChemMedChem* **2019,** *14* (1), 8-23.

11. Kramer, H. J. M.; Rosmalen, G. M., Crystallization. In *Encyclopedia of Separation Science*, Wilson, I. D., Ed. Academic Press: Oxford, 2000; pp 64-84.

12. Ramachandran, V.; Halfpenny, P. J.; Roberts, K. J., Crystal Science Fundamentals. In *From Molecules to Crystallizers- An Introduction to Crystallization*, Roberts, K. J.; Docherty, R.; Tamura, R., Eds. Oxford University Press: New York, 2000.

13. Wunderlich, B.; Grebowicz, J., Thermotropic mesophases and mesophase transitions of linear, flexible macromolecules. In *Liquid Crystal Polymers II/III*, Springer: Berlin, 1984; Vol. 60/61, pp 1-59.

14. Camacho Corzo, D. M.; Ma, C. Y.; Ramachandran, V.; Mahmud, T.; Roberts, K. J., Crystallisation Route Map. In *Engineering Crystallography: From Molecule to Crystal to Functional Form*, Roberts, K. J.; Docherty, R.; Tamura, R., Eds. Springer Netherlands: 2017.

15. Crystallisation. https://syrris.co.jp/applications/what-is-crystallization-and-what-are-the-methods-of-crystallization/ (accessed April 2020).

16. Volmer, M.; Weber, A. Z., Nucleus Formation in Supersaturated Systems. *Zeitschrift für Physikalische Chemie* **1926,** (119), 277-301.

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

18. Gibbs, W., On the equilibrium of heterogeneous substances. *Transactions of the Connecticut Academy of Arts and Sciences* **1875-1878,** *3*.

19. Rowlinson, J. S., Translation of J. D. van der Waals' “The thermodynamik theory of capillarity under the hypothesis of a continuous variation of density”. *Journal of Statistical Physics* **1979,** *20* (2), 197-200.

20. Ostwald, W., Studien über die Bildung und Umwandlung fester Körper. *Zeitschrift für Physikalische Chemie* **1897,** *22U* (1), 289-330.

21. Schmelzer, J.; Abyzov, A., How Do Crystals Nucleate and Grow: Ostwald’s Rule of Stages and Beyond. In *Thermal Physics and Thermal Analysis*, 2017; pp 195-211.

22. Schmelzer, J. W. P.; Boltachev, G. S.; Baidakov, V. G., Classical and generalized Gibbs’ approaches and the work of critical cluster formation in nucleation theory. *The Journal of Chemical Physics* **2006,** *124* (19), 194503.

23. Cahn, J. W.; Hilliard, J. E., Free Energy of a Nonuniform System. III. Nucleation in a Two-Component Incompressible Fluid. *The Journal of Chemical Physics* **1959,** *31* (3), 688-699.

24. Dalmolen, J., Crystal Growth from Solution. In *Synthesis and application of new chiral amines in Dutch resolution: Family Behaviour in Nucleation Inhibition.*, Groningen, 2005.

25. Pritula, I.; Sangwal, K., 29 - Fundamentals of Crystal Growth from Solutions. In *Handbook of Crystal Growth (Second Edition)*, Rudolph, P., Ed. Elsevier: Boston, 2015; pp 1185-1227.

26. Mann, A., Organic Structure and Bonding. In *Pharmaceutical Chemistry*, Rostron, C. B. J., Ed. Oxford University Press: 2021.

27. Jespersen, N. D.; Hyslop, A., Intermolecular Attractions and the Properties of Liquids and Solids. In *Chemistry: The Molecular Nature of Matter*, Wiley: 2021.

28. Davis, R. F., *Diamond Films and Coatings - Development, Properties and Applications*. Park Ridge, N.J: 1993.

29. Bernstein, J., Fundamentals. In *Polymorphism in Molecular Crystals*, OUP Oxford: 2007.

30. Aitipamula, S.; Chow, P. S.; Tan, R. B. H., Conformational and enantiotropic polymorphism of a 1 : 1 cocrystal involving ethenzamide and ethylmalonic acid. *CrystEngComm* **2010,** *12* (11), 3691-3697.

31. McCrone, W. C., Polymorphism. In *Physics and Chemistry of organic solid state*, Fox, D.; Labes, M. M.; Weissberger, A., Eds. USA, 1965; Vol. 2.

32. Haleblian, J.; McCrone, W., Pharmaceutical applications of polymorphism. *Journal of Pharmaceutical Sciences* **1969,** *58* (8), 911-929.

33. Bernstein, J., Introduction and historical background. In *Polymorphism in Molecular Crystals* Oxford University Press: 2007.

34. Sengupta, B.; Sengupta, P.; Grant, R.; Beasley, M.; Mason, B.; Love, T.; Barroso, L.; Alvarado, M.; Zaman, M., Conformational Polymorphism in Organic Crystals: Structural and Functional aspects -A Review Article Details. **2019**.

35. Beyer, T.; Day, G. M.; Price, S. L., The prediction, morphology, and mechanical properties of the polymorphs of paracetamol. *J Am Chem Soc* **2001,** *123* (21), 5086-94.

36. Nangia, A., Conformational Polymorphism in Organic Crystals. *Accounts of Chemical Research* **2008,** *41* (5), 595-604.

37. Yu, L.; Stephenson, G. A.; Mitchell, C. A.; Bunnell, C. A.; Snorek, S. V.; Bowyer, J. J.; Borchardt, T. B.; Stowell, J. G.; Byrn, S. R., Thermochemistry and Conformational Polymorphism of a Hexamorphic Crystal System. *Journal of the American Chemical Society* **2000,** *122* (4), 585-591.

38. Surov, A. O.; Manin, A. N.; Voronin, A. P.; Churakov, A. V.; Perlovich, G. L.; Vener, M. V., Weak Interactions Cause Packing Polymorphism in Pharmaceutical Two-

Component Crystals. The Case Study of the Salicylamide Cocrystal. *Crystal Growth & Design* **2017,** *17* (3), 1425-1437.

39. Lohani, S.; Grant, D. J., Thermodynamics of Polymorphs. In *Polymorphism: In the Pharmaceutical Industry*, Hilfiker, R., Ed. John Wiley & Sons: 2006.

40. Buerger, M. J.; Bloom, M. C., Crystal Polymorphism. In *Zeitschrift für Kristallographie - Crystalline Materials*, 1937; Vol. 96, p 182.

41. Kawakami, K., Reversibility of Enantiotropically Related Polymorphic Transformations from a Practical Viewpoint: Thermal Analysis of Kinetically Reversible/Irreversible Polymorphic Transformations. *Journal of Pharmaceutical Sciences* **2007,** *96* (5), 982-989.

42. Fedorov, A. Y.; Rychkov, D. A.; Losev, E. A.; Zakharov, B. A.; Stare, J.; Boldyreva, E. V., Effect of pressure on two polymorphs of tolazamide: why no interconversion? *CrystEngComm* **2017,** *19* (16), 2243-2252.

43. Rietveld, I. B.; Barrio, M.; Lloveras, P.; Céolin, R.; Tamarit, J.-L., Polymorphism of spironolactone: An unprecedented case of monotropy turning to enantiotropy with a huge difference in the melting temperatures. *International Journal of Pharmaceutics* **2018,** *552* (1), 193-205.

44. Gu, C. H.; Grant, D. J., Estimating the relative stability of polymorphs and hydrates from heats of solution and solubility data. *J Pharm Sci* **2001,** *90* (9), 1277-87.

45. Burger, A.; Ramberger, R., On the polymorphism of pharmaceuticals and other molecular crystals. I - Theory of thermodynamics rules. *Microchimica Acta* **1979,** *72* (3-4), 259-271.

46. Yu, L., Inferring thermodynamic stability relationship of polymorphs from melting data. *Journal of Pharmaceutical Sciences* **1995,** *84* (8), 966-974.

47. Qi, M.-H.; Hong, M.-H.; Liu, Y.; Wang, E.-F.; Ren, F.-Z.; Ren, G.-B., Estimating Thermodynamic Stability Relationship of Polymorphs of Sofosbuvir. *Crystal Growth & Design* **2015,** *15* (10), 5062-5067.

48. Yang, L.; Yin, Q.; Hou, B.; Wang, Y.; Bao, Y.; Wang, J.; Hao, H., Solubility and Thermodynamic Stability of the Enantiotropic Polymorphs of 2,3,5-Trimethyl-1,4-diacetoxybenzene. *Industrial & Engineering Chemistry Research* **2013,** *52* (7), 2477-2485.

49. Nicoud, L.; Licordari, F.; Myerson, A. S., Estimation of the Solubility of Metastable Polymorphs: A Critical Review. *Crystal Growth & Design* **2018,** *18* (11), 7228-7237.

50. Guillory, J. K.; Poust, R. I., Chemical kinetics and drug stability. In *Modern Pharmaceutics, Drugs and the Pharmaceutical Sciences*, Banker, G. S.; Siepmann, J.; Rhodes, C., Eds. CRC Press: New York, 2002; pp 111-138.

51. Joiris, E.; Martino, P. D.; Berneron, C.; Guyot-Hermann, A.-M.; Guyot, J.-C., Compression Behavior of Orthorhombic Paracetamol. *Pharm Res* **1998,** *15* (7), 1122-1130.

52. Capece, M.; Davé, R., Enhanced Physical Stability of Amorphous Drug Formulations via Dry Polymer Coating. *Journal of Pharmaceutical Sciences* **2015,** *104* (6), 2076-2084.

53. Campeta, A. M.; Chekal, B. P.; Abramov, Y. A.; Meenan, P. A.; Henson, M. J.; Shi, B.; Singer, R. A.; Horspool, K. R., Development of a targeted polymorph screening approach for a complex polymorphic and highly solvating API. *J Pharm Sci* **2010,** *99* (9), 3874-86.

54. Kras, W.; Carletta, A.; Montis, R.; Sullivan, R. A.; Cruz-Cabeza, A. J., Switching polymorph stabilities with impurities provides a thermodynamic route to benzamide form III. *Communications Chemistry* **2021,** *4* (1), 38.

55. Piqueras, S.; Duponchel, L.; Tauler, R.; de Juan, A., Monitoring polymorphic transformations by using in situ Raman hyperspectral imaging and image multiset analysis. *Analytica Chimica Acta* **2014,** *819*, 15-25.

56. Arlin, J. B.; Price, L. S.; Price, S. L.; Florence, A. J., A strategy for producing predicted polymorphs: catemeric carbamazepine form V. *Chem Commun (Camb)* **2011,** *47* (25), 7074-6.

57. Potticary, J.; Terry, L. R.; Bell, C.; Papanikolopoulos, A. N.; Christianen, P. C.; Engelkamp, H.; Collins, A. M.; Fontanesi, C.; Kociok-Köhn, G.; Crampin, S.; Da Como, E.; Hall, S. R., An unforeseen polymorph of coronene by the application of magnetic fields during crystal growth. *Nat Commun* **2016,** *7*, 11555.

58. Bučar, D.-K.; Lancaster, R. W.; Bernstein, J., Disappearing polymorphs revisited. *Angew Chem Int Ed Engl* **2015,** *54* (24), 6972-6993.

59. Hasa, D.; Marosa, M.; Bučar, D.-K.; Corpinot, M. K.; Amin, D.; Patel, B.; Jones, W., Mechanochemical Formation and “Disappearance” of Caffeine–Citric-Acid Cocrystal Polymorphs. *Crystal Growth & Design* **2020,** *20* (2), 1119-1129.

60. Gadade, D. D.; Pekamwar, S. S., Pharmaceutical Cocrystals: Regulatory and Strategic Aspects, Design and Development. *Adv Pharm Bull* **2016,** *6* (4), 479-494.

61. Tieknik, E. R. T.; Zukerman- Schpector, J., *Multi-component Crystals: Synthesis, Concepts, Function*. 2017.

62. Cinčić, D.; Friščić, T.; Jones, W., A Stepwise Mechanism for the Mechanochemical Synthesis of Halogen-Bonded Cocrystal Architectures. *Journal of the American Chemical Society* **2008,** *130* (24), 7524-7525.

63. Aakeröy, C. B.; Schultheiss, N. C.; Rajbanshi, A.; Desper, J.; Moore, C., Supramolecular Synthesis Based on a Combination of Hydrogen and Halogen Bonds. *Crystal Growth & Design* **2009,** *9* (1), 432-441.

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

65. Maeno, Y.; Fukami, T.; Kawahata, M.; Yamaguchi, K.; Tagami, T.; Ozeki, T.; Suzuki, T.; Tomono, K., Novel pharmaceutical cocrystal consisting of paracetamol and trimethylglycine, a new promising cocrystal former. *Int J Pharm* **2014,** *473* (1-2), 179-86.

66. Zhou, Z.; Li, W.; Sun, W. J.; Lu, T.; Tong, H. H. Y.; Sun, C. C.; Zheng, Y., Resveratrol cocrystals with enhanced solubility and tabletability. *Int J Pharm* **2016,** *509* (1-2), 391-399.

67. Basavoju, S.; Boström, D.; Velaga, S. P., Indomethacin-saccharin cocrystal: design, synthesis and preliminary pharmaceutical characterization. *Pharm Res* **2008,** *25* (3), 530-41.

68. Schartman, R. R., On the thermodynamics of cocrystal formation. *International Journal of Pharmaceutics* **2009,** *365* (1-2), 77-80.

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

70. Dubey, R.; Desiraju, G. R., Combinatorial selection of molecular conformations and supramolecular synthons in quercetin cocrystal landscapes: a route to ternary solids. *IUCrJ* **2015,** *2* (4), 402-408.

71. Bond, A. D., Fundamental Aspects of Salts and Co-crystals. In *Pharmaceutical Salts and Co-crystals*, Wouters, J.; Quéré, L., Eds. The Royal Society of Chemistry: 2011; p 0.

72. Aakeröy, C. B.; Fasulo, M. E.; Desper, J., Cocrystal or salt: does it really matter? *Mol Pharm* **2007,** *4* (3), 317-22.

73. Mohamed, S.; Tocher, D. A.; Price, S. L., Computational prediction of salt and cocrystal structures—Does a proton position matter? *International Journal of Pharmaceutics* **2011,** *418* (2), 187-198.

74. Childs, S. L.; Stahly, G. P.; Park, A., The salt-cocrystal continuum: the influence of crystal structure on ionization state. *Mol Pharm* **2007,** *4* (3), 323-38.

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

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

77. Price, S. L., Crystal Structure Prediction. In *Encyclopedia of Supramolecular Chemistry*, Taylor and Francis: New York: 2007; pp 371-379.

78. Price, S. L., Predicting crystal structures of organic compounds. *Chemical Society Reviews* **2014,** *43* (7), 2098-2111.

79. Marchese Robinson, R. L.; Geatches, D.; Morris, C.; Mackenzie, R.; Maloney, A. G. P.; Roberts, K. J.; Moldovan, A.; Chow, E.; Pencheva, K.; Vatvani, D. R. M., Evaluation of Force-Field Calculations of Lattice Energies on a Large Public Dataset, Assessment of Pharmaceutical Relevance, and Comparison to Density Functional Theory. *Journal of Chemical Information and Modeling* **2019,** *59* (11), 4778-4792.

80. Price, S. L.; Price, L. S., Modelling Intermolecular Forces for Organic Crystal Structure Prediction. In *Intermolecular Forces and Clusters I*, Wales, D. J., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 81-123.

81. 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., et al., Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B* **2016,** *72* (4), 439-459.

82. Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S. X. M.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; Della Valle, R. G.; Desiraju, G. R.; van Eijck, B. P.; Facelli, J. C.; Ferraro, M. B.; Grillo, D.; Habgood, M., et al., Towards crystal structure prediction of complex organic compounds - a report on the fifth blind test. *Acta Crystallographica Section B* **2011,** *67* (6), 535-551.

83. Mattei, A.; Li, T., Intermolecular Interactions and Computational Modeling. *Pharmaceutical Crystals* **2018**, 123-167.

84. P Thomas, S.; Spackman, M., The Polymorphs of ROY: A Computational Study of Lattice Energies and Conformational Energy Differences*. *Australian Journal of Chemistry* **2018,** *71*.

85. Greenwell, C.; Beran, G. J. O., Inaccurate Conformational Energies Still Hinder Crystal Structure Prediction in Flexible Organic Molecules. *Crystal Growth & Design* **2020,** *20* (8), 4875-4881.

86. Karamertzanis, P. G.; Pantelides, C. C., Ab initio crystal structure prediction-I. Rigid molecules. *J Comput Chem* **2005,** *26* (3), 304-24.

87. Karfunkel, H. R.; Gdanitz, R. J., Ab Initio prediction of possible crystal structures for general organic molecules. *Journal of Computational Chemistry* **1992,** *13* (10), 1171-1183.

88. Neumann, M. A.; Leusen, F. J.; Kendrick, J., A major advance in crystal structure prediction. *Angew Chem Int Ed Engl* **2008,** *47* (13), 2427-30.

89. Iuzzolino, L.; Reilly, A. M.; McCabe, P.; Price, S. L., Use of Crystal Structure Informatics for Defining the Conformational Space Needed for Predicting Crystal Structures of Pharmaceutical Molecules. *Journal of Chemical Theory and Computation* **2017,** *13* (10), 5163-5171.

90. Thompson, H. P. G.; Day, G. M., Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape. *Chemical Science* **2014,** *5* (8), 3173-3182.

91. Dubbeldam, D.; Walton, K. S.; Vlugt, T. J. H.; Calero, S., Design, Parameterization, and Implementation of Atomic Force Fields for Adsorption in Nanoporous Materials. *Advanced Theory and Simulations* **2019,** *2* (11), 1900135.

92. Taylor, C. R.; Mulvee, M. T.; Perenyi, D. S.; Probert, M. R.; Day, G. M.; Steed, J. W., Minimizing Polymorphic Risk through Cooperative Computational and Experimental Exploration. *Journal of the American Chemical Society* **2020,** *142* (39), 16668-16680.

93. Nyman, J.; Day, G. M., Static and lattice vibrational energy differences between polymorphs. *CrystEngComm* **2015,** *17* (28), 5154-5165.

94. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010,** *132* (15), 154104.

95. Tkatchenko, A.; Scheffler, M., Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Physical Review Letters* **2009,** *102* (7), 073005.

96. Akkermans, R. L. C.; Spenley, N. A.; Robertson, S. H., Monte Carlo methods in Materials Studio. *Molecular Simulation* **2013,** *39* (14-15), 1153-1164.

97. Isele-Holder, R. E.; Mitchell, W.; Ismail, A. E., Development and application of a particle-particle particle-mesh Ewald method for dispersion interactions. *The Journal of Chemical Physics* **2012,** *137* (17), 174107.

98. Sun, H.; Ren, P.; Fried, J. R., The COMPASS force field: parameterization and validation for phosphazenes. *Computational and Theoretical Polymer Science* **1998,** *8* (1), 229-246.

99. Akkermans, R. L. C.; Spenley, N. A.; Robertson, S. H., COMPASS III: automated fitting workflows and extension to ionic liquids. *Molecular Simulation* **2020**, 1-12.

100. Neumann, M. A., Tailor-made force fields for crystal-structure prediction. *J Phys Chem B* **2008,** *112* (32), 9810-29.

101. Mohamed, S., Integrating Computed Crystal Energy Landscapes in Crystal Form Discovery and Characterisation. In *Understanding Intermolecular Interactions in the Solid State: Approaches and Techniques*, The Royal Society of Chemistry: 2019; pp 1-31.

102. Braun, D. E.; McMahon, J. A.; Koztecki, L. H.; Price, S. L.; Reutzel-Edens, S. M., Contrasting Polymorphism of Related Small Molecule Drugs Correlated and Guided by the Computed Crystal Energy Landscape. *Crystal Growth & Design* **2014,** *14* (4), 2056-2072.

103. Leach, A. R., Empirical Force-Field Models: Molecular Mechanics. In *Molecular Modelling: Principles and Applications*, 2 ed.; Pearson Education Limited: England, 2001.

104. Brodersen, S.; Wilke, S.; Leusen, F. J. J.; Engel, G., A study of different approaches to the electrostatic interaction in force field methods for organic crystals. *Physical Chemistry Chemical Physics* **2003,** *5* (21), 4923-4931.

105. Williams, D. E., Improved intermolecular force field for molecules containing H, C, N, and O atoms, with application to nucleoside and peptide crystals. *Journal of Computational Chemistry* **2001,** *22* (11), 1154-1166.

106. Gavezzotti, A.; Filippini, G., Non-Covalent Interactions in Organic Crystals, and the Calibration of Empirical Force Fields. In *Computational Approaches in Supramolecular Chemistry*, Wipff, G., Ed. Springer Netherlands: Dordrecht, 1994; pp 51-62.

107. Nyman, J.; Pundyke, O. S.; Day, G. M., Accurate force fields and methods for modelling organic molecular crystals at finite temperatures. *Physical Chemistry Chemical Physics* **2016,** *18* (23), 15828-15837.

108. Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M., Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *The Journal of Physical Chemistry* **1996,** *100* (18), 7352-7360.

109. Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M., Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Physical Chemistry Chemical Physics* **2010,** *12* (30), 8478-8490.

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

111. Gavezzotti, A., Efficient computer modeling of organic materials. The atom–atom, Coulomb–London–Pauli (AA-CLP) model for intermolecular electrostatic-polarization, dispersion and repulsion energies. *New Journal of Chemistry* **2011,** *35* (7), 1360-1368.

112. Mackenzie, C. F.; Spackman, P. R.; Jayatilaka, D.; Spackman, M. A., CrystalExplorer model energies and energy frameworks: extension to metal coordination compounds, organic salts, solvates and open-shell systems. *IUCrJ* **2017,** *4* (5), 575-587.

113. Wolff, S. K.; Grimwood, D. J.; McKinnon, J. J.; Turner, M. J.; Jayatilaka, D.; Spackman, M. A. *CrystalExplorer*, 17; University of Western Australia, 2017.

114. Li, T.; Ayers, P. W.; Liu, S.; Swadley, M. J.; Aubrey-Medendorp, C., Crystallization Force—A Density Functional Theory Concept for Revealing Intermolecular Interactions and Molecular Packing in Organic Crystals. *Chemistry – A European Journal* **2009,** *15* (2), 361-371.

115. Vasileiadis, M.; Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C., The polymorphs of ROY: application of a systematic crystal structure prediction technique. *Acta Crystallographica Section B* **2012,** *68* (6), 677-685.

116. Habgood, M.; Sugden, I. J.; Kazantsev, A. V.; Adjiman, C. S.; Pantelides, C. C., Efficient Handling of Molecular Flexibility in Ab Initio Generation of Crystal Structures. *Journal of Chemical Theory and Computation* **2015,** *11* (4), 1957-1969.

117. Chen, S.; Guzei, I. A.; Yu, L., New Polymorphs of ROY and New Record for Coexisting Polymorphs of Solved Structures. *Journal of the American Chemical Society* **2005,** *127* (27), 9881-9885.

118. Stein, M.; Heimsaat, M., Intermolecular Interactions in Molecular Organic Crystals upon Relaxation of Lattice Parameters. *Crystals* **2019,** *9* (12).

119. Beran, G., Modeling Polymorphic Molecular Crystals with Electronic Structure Theory. *Chemical reviews* **2016,** *116 9*, 5567-613.

120. Červinka, C.; Fulem, M., State-of-the-Art Calculations of Sublimation Enthalpies for Selected Molecular Crystals and Their Computational Uncertainty. *Journal of Chemical Theory and Computation* **2017,** *13* (6), 2840-2850.

121. Mortazavi, M.; Hoja, J.; Aerts, L.; Quéré, L.; van de Streek, J.; Neumann, M. A.; Tkatchenko, A., Computational polymorph screening reveals late-appearing and poorly-soluble form of rotigotine. *Communications Chemistry* **2019,** *2* (1), 70.

122. Hoja, J.; Ko, H.-Y.; Neumann, M. A.; Car, R.; DiStasio, R. A.; Tkatchenko, A., Reliable and practical computational description of molecular crystal polymorphs. *Science Advances* **2019,** *5* (1), eaau3338.

123. Buchholz, H. K.; Stein, M., Accurate lattice energies of organic molecular crystals from periodic turbomole calculations. *Journal of Computational Chemistry* **2018,** *39* (19), 1335-1343.

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

125. Chan, H. C. S.; Kendrick, J.; Neumann, M. A.; Leusen, F. J. J., Towards ab initio screening of co-crystal formation through lattice energy calculations and crystal structure prediction of nicotinamide, isonicotinamide, picolinamide and paracetamol multi-component crystals. *CrystEngComm* **2013,** *15* (19), 3799-3807.

126. Taylor, C. R.; Day, G. M., Evaluating the Energetic Driving Force for Cocrystal Formation. *Cryst Growth Des* **2018,** *18* (2), 892-904.

127. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chemical Reviews* **2005,** *105* (8), 2999-3094.

128. Klamt, A., Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *The Journal of Physical Chemistry* **1995,** *99* (7), 2224-2235.

129. Onsager, L., Electric Moments of Molecules in Liquids. *Journal of the American Chemical Society* **1936,** *58* (8), 1486-1493.

130. Quan, C.; Stamm, B., Meshing Molecular Surfaces Based on Analytical Implicit Representation. *Journal of Molecular Graphics and Modelling* **2016,** *71*.

131. Klamt, A.; Schuurmann, G., COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2* **1993,** (5), 799-805.

132. Wichmann, K.; Klamt, A., Drug Solubility and Reaction Thermodynamics. In *Chemical Engineering in the Pharmaceutical Industry*, John Wiley & Sons, Inc.: 2010; pp 457-476.

133. Spiteri, L. A knowledged-based approach towards the evaluation, analysis and prediction of the properties of crystal forms. University of Malta, 2019.

134. Klamt, A., The COSMO and COSMO-RS solvation models. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011,** *1* (5), 699-709.

135. Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W., Refinement and Parametrization of COSMO-RS. *The Journal of Physical Chemistry A* **1998,** *102* (26), 5074-5085.

136. Loschen, C.; Klamt, A., New Developments in Prediction of Solid-State Solubility and Cocrystallization Using COSMO-RS Theory. In *Computational Pharmaceutical Solid State Chemistry*, John Wiley & Sons, Inc: 2016; pp 211-233.

137. Mullins, E.; Oldland, R.; Liu, Y. A.; Wang, S.; Sandler, S. I.; Chen, C.-C.; Zwolak, M.; Seavey, K. C., Sigma-Profile Database for Using COSMO-Based Thermodynamic Methods. *Industrial & Engineering Chemistry Research* **2006,** *45* (12), 4389-4415.

138. Eckert, F.; Klamt, A., Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE Journal* **2002,** *48* (2), 369-385.

139. Reinisch, J.; Diedenhofen, M.; Wilcken, R.; Udvarhelyi, A.; Glöß, A., Benchmarking Different QM Levels for Usage with COSMO-RS. *Journal of Chemical Information and Modeling* **2019,** *59* (11), 4806-4813.

140. Grensemann, H.; Gmehling, J., Performance of a Conductor-Like Screening Model for Real Solvents Model in Comparison to Classical Group Contribution Methods. *Industrial & Engineering Chemistry Research* **2005,** *44* (5), 1610-1624.

141. Han, J.; Dai, C.; Yu, G.; Lei, Z., Parameterization of COSMO-RS model for ionic liquids. *Green Energy & Environment* **2018,** *3* (3), 247-265.

142. Paduszyński, K., An overview of the performance of the COSMO-RS approach in predicting the activity coefficients of molecular solutes in ionic liquids and derived properties at infinite dilution. *Physical Chemistry Chemical Physics* **2017,** *19* (19), 11835-11850.

143. *Biovia COSMOtherm*, 18.0.2; Dassault Systemes: Germany.

144. Hyttinen, N.; Prisle, N. L., Improving Solubility and Activity Estimates of Multifunctional Atmospheric Organics by Selecting Conformers in COSMOtherm. *The Journal of Physical Chemistry A* **2020,** *124* (23), 4801-4812.

145. *COSMOconf* 4.2; COSMOlogic GmbH & Co. KG: Leverkusen, Germany.

146. Loschen, C.; Klamt, A., COSMOquick: A Novel Interface for Fast σ-Profile Composition and Its Application to COSMO-RS Solvent Screening Using Multiple Reference Solvents. *Industrial & Engineering Chemistry Research* **2012,** *51* (43), 14303-14308.

147. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C., Electronic structure calculations on workstation computers: The program system turbomole. *Chemical Physics Letters* **1989,** *162* (3), 165-169.

148. Steffen, C.; Thomas, K.; Huniar, U.; Hellweg, A.; Rubner, O.; Schroer, A., TmoleX—A graphical user interface for TURBOMOLE. *Journal of Computational Chemistry* **2010,** *31* (16), 2967-2970.

149. Gusarov, S.; Stoyanov, S. R., COSMO-RS-Based Descriptors for the Machine Learning-Enabled Screening of Nucleotide Analogue Drugs against SARS-CoV-2. *The Journal of Physical Chemistry Letters* **2020,** *11* (21), 9408-9414.

150. Loschen, C.; Klamt, A., Solubility prediction, solvate and cocrystal screening as tools for rational crystal engineering. *The Journal of pharmacy and pharmacology* **2015,** *67* (6), 803-11.

151. Palomar, J.; Torrecilla, J. S.; Lemus, J.; Ferro, V. R.; Rodríguez, F., Prediction of non-ideal behavior of polarity/polarizability scales of solvent mixtures by integration of a novel COSMO-RS molecular descriptor and neural networks. *Physical Chemistry Chemical Physics* **2008,** *10* (39), 5967-5975.

152. Khan, A. S.; Ibrahim, T. H.; Rashid, Z.; Khamis, M. I.; Nancarrow, P.; Jabbar, N. A., COSMO-RS based screening of ionic liquids for extraction of phenolic compounds from aqueous media. *Journal of Molecular Liquids* **2021,** *328*, 115387.

153. Moya, C.; Hospital-Benito, D.; Santiago, R.; Lemus, J.; Palomar, J., Prediction of CO2 chemical absorption isotherms for ionic liquid design by DFT/COSMO-RS calculations. *Chemical Engineering Journal Advances* **2020,** *4*, 100038.

154. Chapman, C. J.; Groven, L. J., Evaluation of solvate and co-crystal screening methods for CL-20 containing energetic materials. *Journal of Energetic Materials* **2022,** *40* (3), 258-272.

155. Nunes Costa, R.; Choquesillo-Lazarte, D.; Cuffini, S. L.; Pidcock, E.; Infantes, L., Optimization and comparison of statistical tools for the prediction of multicomponent forms of a molecule: the antiretroviral nevirapine as a case study. *CrystEngComm* **2020,** *22* (43), 7460-7474.

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

157. Allen Frank, H., The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallographica Section B* **2008,** *58* (3-1), 380-388.

158. CSD Statistics. https://www.ccdc.cam.ac.uk/CCDCStats/Stats (accessed on February, 2023).

159. Bryant, M.; Black, S.; Blade, H.; Docherty, R.; Maloney, A.; Taylor, S., The CSD Drug Subset: The Changing Chemistry and Crystallography of Small Molecule Pharmaceuticals. *Journal of Pharmaceutical Sciences* **2019,** *108*.

160. Galek, P. T. A.; Pidcock, E.; Wood, P. A.; Bruno, I. J.; Groom, C. R., One in half a million: a solid form informatics study of a pharmaceutical crystal structure. *CrystEngComm* **2012,** *14* (7), 2391-2403.

161. Galek, P. T. A.; Pidcock, E.; Wood, P. A.; Feeder, N.; Allen, F. H., Navigating the Solid Form Landscape with Structural Informatics. *Computational Pharmaceutical Solid State Chemistry* **2016**, 15-35.

162. Spiteri, L.; Baisch, U.; Vella-Zarb, L., Understanding Conformational Polymorphism in Ganciclovir: A Holistic Approach. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2021,** *3* (1).

163. Taylor, R.; Wood, P. A., A Million Crystal Structures: The Whole Is Greater than the Sum of Its Parts. *Chemical Reviews* **2019**.

164. Takahashi, O.; Kohno, Y.; Nishio, M., Relevance of Weak Hydrogen Bonds in the Conformation of Organic Compounds and Bioconjugates: Evidence from Recent Experimental Data and High-Level ab Initio MO Calculations. *Chemical Reviews* **2010,** *110* (10), 6049-6076.

165. Galek, P. T. A.; Fabian, L.; Allen, F. H., Universal prediction of intramolecular hydrogen bonds in organic crystals. *Acta Crystallographica Section B* **2010,** *66* (2), 237-252.

166. *Mogul User Guide and Tutorials*; Cambridge Crystallographic Data Centre: 2018.

167. Wright, S. E.; Bryant, M. J.; Cruz-Cabeza, A. J., Is it usual to be unusual? An investigation into molecular conformations in organic crystals. *CrystEngComm* **2020**.

168. Bruno, I.; Cole, J.; Lommerse, J.; Rowland, R.; Taylor, R.; Verdonk, M., IsoStar: A library of information about nonbonded interactions. *Journal of Computer-Aided Molecular Design* **1997,** *11* (6), 525-537.

169. 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., Mercury 4.0: from visualization to analysis, design and prediction. *Journal of Applied Crystallography* **2020,** *53* (1), 226-235.

170. Chisholm, J.; Pidcock, E.; van de Streek, J.; Infantes, L.; Motherwell, S.; Allen, F. H., Knowledge-based approaches to crystal design. *CrystEngComm* **2006,** *8* (1), 11-28.

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

172. Sarma, J. A. R. P.; Desiraju, G. R., The Supramolecular Synthon Approach to Crystal Structure Prediction. *Crystal Growth & Design* **2002,** *2* (2), 93-100.

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

174. Nauha, E.; Bernstein, J., “Predicting” Crystal Forms of Pharmaceuticals Using Hydrogen Bond Propensities: Two Test Cases. *Crystal Growth & Design* **2014,** *14* (9), 4364-4370.

175. Wood, P. A.; Feeder, N.; Furlow, M.; Galek, P. T. A.; Groom, C. R.; Pidcock, E., Knowledge-based approaches to co-crystal design. *CrystEngComm* **2014,** *16* (26), 5839-5848.

176. Galek, P. T. A.; Chisholm, J. A.; Pidcock, E.; Wood, P. A., Hydrogen-bond coordination in organic crystal structures: statistics, predictions and applications. *Acta Crystallographica Section B* **2014,** *70* (1), 91-105.

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

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

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

180. Khalaji, M.; Potrzebowski, M. J.; Dudek, M. K., Virtual Cocrystal Screening Methods as Tools to Understand the Formation of Pharmaceutical Cocrystals—A Case Study of Linezolid, a Wide-Range Antibacterial Drug. *Crystal Growth & Design* **2021,** *21* (4), 2301-2314.

181. Tilbury, C. J.; Chen, J.; Mattei, A.; Chen, S.; Sheikh, A. Y., Combining Theoretical and Data-Driven Approaches To Predict Drug Substance Hydrate Formation. *Crystal Growth & Design* **2018,** *18* (1), 57-67.

182. Zarrouk, K.; Piret, J.; Boivin, G., Herpesvirus DNA polymerases: Structures, functions and inhibitors. *Virus Research* **2017,** *234* (Supplement C), 177-192.

183. Vere Hodge, R. A.; Field, H. J., Antiviral agents for herpes simplex virus. *Advances in pharmacology (San Diego, Calif.)* **2013,** *67*, 1-38.

184. *Herpes Simplex Virus*; World Health Organisation: January, 2017.

185. Grinde, B., Herpesviruses: latency and reactivation – viral strategies and host response. *Journal of Oral Microbiology* **2013,** *5*, 10.3402/jom.v5i0.22766.

186. Bloom, D. C.; Dhummakupt, A., The Herpes Simplex Viruses. In *Neurotropic Viral Infections*, Shoshkes Reiss, C., Ed. 2008.

187. Rangel-Sosa, M. M.; Aguilar-Córdova, E.; Rojas-Martínez, A., Immunotherapy and gene therapy as novel treatments for cancer. *Colomb Med (Cali)* **2017,** *48* (3), 138-147.

188. Rangel-Sosa, M.; Aguilar-Córdova, E.; Rojas-Martinez, A., Immunotherapy and gene therapy as novel treatments for cancer. *Colombia Medica* **2017,** *48*, 137-146.

189. Zhu, R.; Chen, D.; Lin, D.; Lu, F.; Yin, J.; Li, N., Adenovirus vector-mediated herpes simplex virus-thymidine kinase gene/ganciclovir system exhibits anti-tumor effects in an orthotopic hepatocellular carcinoma model. *Die Pharmazie* **2014,** *69* (7), 547-52.

190. de Melo, S. M.; Bittencourt, S.; Ferrazoli, E. G.; da Silva, C. S.; da Cunha, F. F.; da Silva, F. H.; Stilhano, R. S.; Denapoli, P. M.; Zanetti, B. F.; Martin, P. K.; Silva, L. M.; dos Santos, A. A.; Baptista, L. S.; Longo, B. M.; Han, S. W., The Anti-Tumor Effects of Adipose Tissue Mesenchymal Stem Cell Transduced with HSV-Tk Gene on U-87-Driven Brain Tumor. *PloS one* **2015,** *10* (6), e0128922.

191. Pan, J. G.; Luo, R. Q.; Zhou, X.; Han, R. F.; Zeng, G. W., Potent antitumor activity of the combination of HSV-TK and endostatin by adeno-associated virus vector for bladder cancer in vivo. *Clinical laboratory* **2013,** *59* (9-10), 1147-58.

192. Parr, A.; Hidalgo, I. J.; Bode, C.; Brown, W.; Yazdanian, M.; Gonzalez, M. A.; Sagawa, K.; Miller, K.; Jiang, W.; Stippler, E. S., The Effect of Excipients on the Permeability of BCS Class III Compounds and Implications for Biowaivers. *Pharm Res* **2016,** *33* (1), 167-176.

193. WHO Drug Information. **2017,** *31* (3).

194. Paintsil, E.; Cheng, Y.-C., Antiviral Agents A2 - Schaechter, Moselio. In *Encyclopedia of Microbiology* 3ed.; Academic Press: Oxford, 2009; pp 223-257.

195. Verheyden, J. P.; Martin, J. C. 9-(1,3-dihydroxy-2-propoxymethyl)-

guanine as an antiviral US 4355032, Oct. 19, 1982.

196. Mallipeddi, S. M.; Sterling, J. E.; Koneru, P. B. Ganciclovir Compositions and Related Methods. US 9486530, Nov. 8, 2016.

197. Sarbajna, R.; Anil, P.; Sivalakshmi Devi, A.; V. Suryanarayana, M.; Sethi, M.; Dutta, D., *Studies on Crystal Modifications of Ganciclovir*. 2011; Vol. 537, p 141-154.

198. Kawamura, T.; Hirayama, N., Crystal Structure of Ganciclovir. *X-ray Structure Analysis Online* **2009,** *25*, 51-52.

199. Fernandes, J. A.; Galli, S.; Palmisano, G.; Volante, P.; Mendes, R. F.; Paz, F. A. A.; Masciocchi, N., Reviewing the Manifold Aspects of Ganciclovir Crystal Forms. *Crystal Growth & Design* **2016,** *16* (7), 4108-4118.

200. Roque-Flores, R. L.; Guzei, I. A.; Matos, J. d. R.; Yu, L., Polymorphs of the antiviral drug ganciclovir. *Acta Crystallographica Section C* **2017,** *73* (12), 1116-1120.

201. Roque-Flores, R. L.; Matos, J. d. R., Simultaneous measurements of X-ray diffraction–differential scanning calorimetry. *Journal of Thermal Analysis and Calorimetry* **2019,** *137* (4), 1347-1358.

202. Gill, K. S.; Wood, M. J., The clinical pharmacokinetics of famciclovir. *Clinical pharmacokinetics* **1996,** *31* (1), 1-8.

203. Harnden, M. R.; Jarvest, R. L.; Slawin, A. M. Z.; Williams, D. J., Crystal and Molecular Structures of the Antiviral Acyclonucleoside 9-[4-Hydroxy-3-(hydroxymethyl)butyl]guanine (BRL 39123, Penciclovir) and its Prodrug 9-[4-Acetoxy-3-(acetoxymethyl)butyl]-2-aminopurine (BRL 42810, Famciclovir). *Nucleosides and Nucleotides* **1990,** *9* (4), 499-513.

204. Harnden, M. R.; Jarvest, R. L. Purine derivatives. US5246937A, 1993.

205. Dolitzky, B.-Z.; Wizel, S.; Reany, O.; Shammai, J. Crystalline solid famciclovir forms i, ii, iii and preparation thereof. WO2004018470A3, April, 2004.

206. Vella-Zarb, L.; Baisch, U., Born out of Fire and Ice: Polymorph Studies of the Antiviral Famciclovir. *Crystals* **2021,** *11* (2).

207. Campbell, K. C.; Greenway, M. J.; Hancock, S. A. Famciclovir monohydrate and its pharmaceutical use. WO1997029108A1, 1997.

208. MacDougall, C.; Guglielmo, B. J., Pharmacokinetics of valaciclovir. *Journal of Antimicrobial Chemotherapy* **2004,** *53* (6), 899-901.

209. Zhang, S.; Zheng, M.; Zhou, M.; Chen, T.; Wang, N.; Zhang, Z.; Zhang, G. A New Hemihydrate of Valacyclovir Hydrochloride *Crystals* [Online], 2017.

210. Fain, H.; Jones, D.; Lake, P. G. Anhydrous crystal form of valaciclovir hydrochloride. WO 03/040145 A1, 2003.

211. Aslan, T.; Ozarslan, A. E.; Sahbaz, F.; Uluozyurt, N. S.; Arslan, N. Crystalline forms of valacyclovir hydrochloride. USOO77863O2B2, 2012.

212. Pizzocaro, F. Valacyclovir polymorphs and process for the preparation thereof. US 20070021444A1, 2007.

213. Carter, B. H.; Partin, J. M.; Varlashkin, P. G.; Winnike, R. A.; Grubb, W. B.; Conway, G. A.; Lake, P. G.; Skinner, D. M.; Whatrup, D. J. Guanine Derivative. US006107302A, 2000.

214. Wizel, S.; Aronhime, J.; Niddam-Hildesheim, V.; Dolitzky, B.-Z.; Etinger, M.; Yuzefovich, M.; Nisnevich, G.; Pertsikov, B.; Tishin, B.; Blasberger, D. Crystalline forms of valacyclovir hydrochloride. US 2003/0114470 A1, 2003.

215. Rani, D.; Goyal, P.; Chadha, R., Conformational flexibility and packing plausibility of repaglinide polymorphs. *Journal of Molecular Structure* **2018,** *1157*, 263-275.

216. Bajorath, J., Selected Concepts and Investigations in Compound Classification, Molecular Descriptor Analysis, and Virtual Screening. *Journal of Chemical Information and Computer Sciences* **2001,** *41* (2), 233-245.

217. Yu, L.; Reutzel-Edens, S. M.; Mitchell, C. A., Crystallization and Polymorphism of Conformationally Flexible Molecules: Problems, Patterns, and Strategies. *Organic Process Research & Development* **2000,** *4* (5), 396-402.

218. Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D., Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *Journal of Medicinal Chemistry* **2002,** *45* (12), 2615-2623.

219. COSMOlogic *COSMOquick User guide*; Germany.

220. Caron, G.; Digiesi, V.; Solaro, S.; Ermondi, G., Flexibility in early drug discovery: focus on the beyond-Rule-of-5 chemical space. *Drug Discovery Today* **2020,** *25* (4), 621-627.

221. Anderson, K. M.; Probert, M. R.; Goeta, A. E.; Steed, J. W., Size does matter—the contribution of molecular volume, shape and flexibility to the formation of co-crystals and structures with Z′ > 1. *CrystEngComm* **2011,** *13* (1), 83-87.

222. Todeschini, R.; Consonni, V., Flexibility Indices. In *Handbook of Molecular Descriptors*, Mannhold, R.; Kubinyi, H.; Timmerman, H., Eds. Wiley-VCH: 2000; Vol. 11, pp 178-180.

223. McCabe, P.; Korb, O.; Cole, J., Kernel Density Estimation Applied to Bond Length, Bond Angle, and Torsion Angle Distributions. *Journal of Chemical Information and Modeling* **2014,** *54* (5), 1284-1288.

224. *Origin(Pro)*, 2019b; OriginLab Corporation: Northampton, MA, USA.

225. Fisher, N. I., Smoothing a sample of circular data. *Journal of Structural Geology* **1989,** *11* (6), 775-778.

226. Taylor, C. C., Automatic bandwidth selection for circular density estimation. *Computational Statistics & Data Analysis* **2008,** *52* (7), 3493-3500.

227. Oliveira, M.; Crujeiras, R. M.; Rodríguez-Casal, A., NPCirc: An R Package for Nonparametric Circular Methods. *Journal of Statistical Software; Vol 1, Issue 9 (2014)* **2014**.

228. *R: A Language and Environment for statistical computing*, R Foundation for Statistical Computing: Vienna, Austria, 2020. https://www.R-project.org.

229. Oliveira, M.; Crujeiras, R. M.; Rodríguez-Casal, A., NPCirc: An R Package for Nonparametric Circular Methods. *2014* **2014,** *61* (9), 26.

230. Perdew, J. P., Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B* **1986,** *33* (12), 8822-8824.

231. Delley, B., From molecules to solids with the DMol3 approach. *The Journal of Chemical Physics* **2000,** *113* (18), 7756-7764.

232. *BIOVIA Material Studio*, 2021. http://accelrys.com/products/collaborative-science/biovia-materials-studio/.

233. Kirschner, K. N.; Reith, D.; Heiden, W., The performance of Dunning, Jensen, and Karlsruhe basis sets on computing relative energies and geometries. *Soft Materials* **2020**, 1-15.

234. Grimme, S.; Steinmetz, M., Effects of London dispersion correction in density functional theory on the structures of organic molecules in the gas phase. *Physical Chemistry Chemical Physics* **2013,** *15* (38), 16031-16042.

235. Weigend, F., Accurate Coulomb-fitting basis sets for H to Rn. **2005,** (1463-9076 (Print)).

236. Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S., A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Physical Chemistry Chemical Physics* **2017,** *19* (48), 32184-32215.

237. Vainio, M. J.; Johnson, M. S., Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *Journal of Chemical Information and Modeling* **2007,** *47* (6), 2462-2474.

238. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **1988,** *38* (6), 3098-3100.

239. *COSMObuild* 17.0; COSMOlogic GmbH & Co. KG: Leverkusen, Germany.

240. Cole, J. C.; Korb, O.; McCabe, P.; Read, M. G.; Taylor, R., Knowledge-Based Conformer Generation Using the Cambridge Structural Database. *J.Chem.Inf.Model* **2018,** (1549-960X (Electronic)).

241. Sargsyan, K.; Hua, Y. H.; Lim, C., Clustangles: An Open Library for Clustering Angular Data. *Journal of Chemical Information and Modeling* **2015,** *55* (8), 1517-1520.

242. Sargsyan, K.; Wright, J.; Lim, C., GeoPCA: a new tool for multivariate analysis of dihedral angles based on principal component geodesics. *Nucleic Acids Res* **2012,** *40* (3), e25-e25.

243. Van Rossum, G.; Drake Jr, F. L., *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam: 1995.

244. Matta, C. F.; Hernández-Trujillo, J.; Tang, T.-H.; Bader, R. F. W., Hydrogen–Hydrogen Bonding: A Stabilizing Interaction in Molecules and Crystals. *Chemistry – A European Journal* **2003,** *9* (9), 1940-1951.

245. Hirshfeld, F. L., Bonded-atom fragments for describing molecular charge densities. *Theoretica chimica acta* **1977,** *44* (2), 129-138.

246. Spackman, M. A.; Jayatilaka, D., Hirshfeld surface analysis. *CrystEngComm* **2009,** *11* (1), 19-32.

247. Lu, T.; Chen, F., Multiwfn: A multifunctional wavefunction analyzer. *Journal of Computational Chemistry* **2012,** *33* (5), 580-592.

248. Lu, T., *Multiwfn - A Multifunctional Wavefunction Analyzer - Software Manual*. 2020.

249. Humphrey, W.; Dalke, A.; Schulten, K., VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **1996,** *14.1*, 33-38.

250. Mayr, A.; Hofner, B.; Weinhold, L.; Schmid, M., Package 'betaboost' - Boosting Beta Regreesion. Mayr, A., Ed. 2018. https://github.com/boost-R/betaboost. (accessed on June 2020)

251. Zeileis, A.; Cribari-Neto, F.; Gruen, B.; Kosmidis, I.; Simas, A. B.; Rocha, A. V., Package: 'betareg' - Beta regression. In [Online] 2020. https://cran.r-project.org/web/packages/betareg/betareg.pdf. (accessed on June 2020)

252. Espinheira, P. L.; da Silva, L. C. M.; Silva, A. D.; Ospina, R., Model Selection Criteria on Beta Regression for Machine Learning. *Machine Learning and Knowledge Extraction* **2019,** *1* (1).

253. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **2001,** *46* (1-3), 3-26.

254. Brameld, K. A.; Kuhn, B.; Reuter, D. C.; Stahl, M., Small Molecule Conformational Preferences Derived from Crystal Structure Data. A Medicinal Chemistry Focused Analysis. *Journal of Chemical Information and Modeling* **2008,** *48* (1), 1-24.

255. Katajisto, J.; Linnolahti, M.; Haukka, M.; Pakkanen, T. A., Development of a New Force Field for Property Prediction of Cyclo-Olefin Copolymers. *The Journal of Physical Chemistry B* **2004,** *108* (7), 2168-2172.

256. Schärfer, C.; Schulz-Gasch, T.; Ehrlich, H.-C.; Guba, W.; Rarey, M.; Stahl, M., Torsion Angle Preferences in Druglike Chemical Space: A Comprehensive Guide. *Journal of Medicinal Chemistry* **2013,** *56* (5), 2016-2028.

257. Puth, M.-T.; Neuhäuser, M.; Ruxton, G. D., Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour* **2015,** *102*, 77-84.

258. Anderson, M. J., Distance-Based Tests for Homogeneity of Multivariate Dispersions. *Biometrics* **2006,** *62* (1), 245-253.

259. Anderson, M. J.; Walsh, D. C. I., PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecological Monographs* **2013,** *83* (4), 557-574.

260. Stier, A. C.; Geange, S. W.; Hanson, K. M.; Bolker, B. M., Predator density and timing of arrival affect reef fish community assembly. *Ecology* **2013,** *94* (5), 1057-1068.

261. Kier, L. B.; Hall, L. H., A Differential Molecular Connectivity Index. *Quantitative Structure-Activity Relationships* **1991,** *10* (2), 134-140.

262. Wicker, J. G. P.; Cooper, R. I., Beyond Rotatable Bond Counts: Capturing 3D Conformational Flexibility in a Single Descriptor. *Journal of Chemical Information and Modeling* **2016,** *56* (12), 2347-2352.

263. Ferrari, S.; Cribari-Neto, F., Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics* **2004,** *31* (7), 799-815.

264. Cribari-Neto, F.; Zeileis, A., Beta Regression in R. *Journal of Statistical Software; Vol 1, Issue 2 (2010)* **2010**.

265. Grün, B.; Kosmidis, I.; Zeileis, A., Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned. *Journal of Statistical Software; Vol 1, Issue 11 (2012)* **2012**.

266. *Crystalline PV - User Manual*. 1.5.1 ed.; Technobis Crystallisation systems: Netherlands, 2015.

267. Wood, P. A.; Olsson, T. S. G.; Cole, J. C.; Cottrell, S. J.; Feeder, N.; Galek, P. T. A.; Groom, C. R.; Pidcock, E., Evaluation of molecular crystal structures using Full Interaction Maps. *CrystEngComm* **2013,** *15* (1), 65-72.

268. Bruno Ian, J.; Cole Jason, C.; Edgington Paul, R.; Kessler, M.; Macrae Clare, F.; McCabe, P.; Pearson, J.; Taylor, R., New software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallographica Section B* **2008,** *58* (3-1), 389-397.

269. Wood, P. A.; Allen, F. H.; Pidcock, E., Hydrogen-bond directionality at the donor H atom—analysis of interaction energies and database statistics. *CrystEngComm* **2009,** *11* (8), 1563-1571.

270. Gavezzotti, A.; Filippini, G., Geometry of the Intermolecular X-H.cntdot..cntdot..cntdot.Y (X, Y = N, O) Hydrogen Bond and the Calibration of Empirical Hydrogen-Bond Potentials. *The Journal of Physical Chemistry* **1994,** *98* (18), 4831-4837.

271. Gavezzotti, A., Are Crystal Structures Predictable? *Accounts of Chemical Research* **1994,** *27* (10), 309-314.

272. Spek, A., Single-crystal structure validation with the program PLATON. *Journal of Applied Crystallography* **2003,** *36* (1), 7-13.

273. *Mercury User Guide and Tutorials*; Cambridge Crystallographic Data Centre: 2018.

274. Thomas, S. P.; Spackman, P. R.; Jayatilaka, D.; Spackman, M. A., Accurate Lattice Energies for Molecular Crystals from Experimental Crystal Structures. *Journal of Chemical Theory and Computation* **2018,** *14* (3), 1614-1623.

275. Folmsbee, D.; Hutchison, G., Assessing conformer energies using electronic structure and machine learning methods. *International Journal of Quantum Chemistry* **2021,** *121* (1), e26381.

276. Rappoport, D.; Furche, F., Property-optimized Gaussian basis sets for molecular response calculations. *The Journal of Chemical Physics* **2010,** *133* (13), 134105.

277. Natarajan, K.; Duraisamy, S.; Rubalingam, P.; Kandasamy, S., Molecular modelling of an organic salt 4-chloroanilinium hydrogen malonate (4CAHM): crystal structure, theoretical and Hirshfeld surface analysis. *Materials Research Innovations* **2017,** *21* (5), 304-312.

278. *CSD Space Group Statistics - Space Group Number Ordering*; Cambridge Structural Database: 2022.

279. Del Bene, J. E.; Jordan, M. J. T., Vibrational spectroscopy of the hydrogen bond: An ab initio quantum-chemical perspective. *International Reviews in Physical Chemistry* **1999,** *18* (1), 119-162.

280. Tan, S. L.; Jotani, M. M.; Tiekink, E. R. T., Utilizing Hirshfeld surface calculations, non-covalent inter-action (NCI) plots and the calculation of inter-action energies in the analysis of mol-ecular packing. *Acta Crystallogr E Crystallogr Commun* **2019,** *75* (Pt 3), 308-318.

281. Vella-Zarb, L.; Dinnebier, R. E.; Baisch, U., The Devil is in the Detail: A Rare H-Bonding Motif in New Forms of Docetaxel. *Crystal Growth & Design* **2013,** *13* (10), 4402-4410.

282. Nishio, M., The CH/π hydrogen bond in chemistry. Conformation, supramolecules, optical resolution and interactions involving carbohydrates. *Physical Chemistry Chemical Physics* **2011,** *13* (31), 13873-13900.

283. McKinnon, J. J.; Fabbiani, F. P. A.; Spackman, M. A., Comparison of Polymorphic Molecular Crystal Structures through Hirshfeld Surface Analysis. *Crystal Growth & Design* **2007,** *7* (4), 755-769.

284. Carugo, O.; Blatova, O. A.; Medrish, E. O.; Blatov, V. A.; Proserpio, D. M., Packing topology in crystals of proteins and small molecules: a comparison. *Scientific Reports* **2017,** *7* (1), 13209.

285. Alvarez, S., A cartography of the van der Waals territories. *Dalton Transactions* **2013,** *42* (24), 8617-8636.

286. Yao, Z.-F.; Wang, J.-Y.; Pei, J., Control of π–π Stacking via Crystal Engineering in Organic Conjugated Small Molecule Crystals. *Crystal Growth & Design* **2018,** *18* (1), 7-15.

287. Shamai, G.; Antebi, S.; Ioffe, D.; Dolitzky, B.-Z.; Kauffmann, B. Process for preparing famciclovir. US 2005/0143400 A1, 2005.

288. Wu, C.-j.; You, J.-z.; Wang, X.-j., Study on the thermal decomposition of famciclovir. *Journal of Thermal Analysis and Calorimetry* **2017,** *131*, 1-11.

289. *IsoStar User guide and Tutorials*; 2018.

290. Desiraju, G. R.; Steiner, T., *The Weak Hydrogen Bond In Structural Chemistry and Biology* Oxford University Press: Oxford and New York, 1999.

291. Tedesco, E.; Giron, D.; Pfeffer, S., Crystal structure elucidation and morphology study of pharmaceuticals in development. *CrystEngComm* **2002,** *4* (67), 393-400.

292. Moreno-Calvo, E.; Calvet, T.; Cuevas-Diarte, M. A.; Aquilano, D., Relationship between the Crystal Structure and Morphology of Carboxylic Acid Polymorphs. Predicted and Experimental Morphologies. *Crystal Growth & Design* **2010,** *10* (10), 4262-4271.

293. Gavezzotti, A.; Presti, L. L., Building Blocks of Crystal Engineering: A Large-Database Study of the Intermolecular Approach between C–H Donor Groups and O, N, Cl, or F Acceptors in Organic Crystals. *Crystal Growth & Design* **2016,** *16* (5), 2952-2962.

294. Müller, P., Disorder. In *Crystal Structure Refinement: A Crystallographer's Guide to SHELXL*, Müller, P.; Herbst-Irmer, R.; Spek, A. L.; Schneider, T. R.; Sawaya, M. R., Eds. Oxford University Press: 2006; p 0.

295. van de Streek, J.; Neumann, M. A., Validation of experimental molecular crystal structures with dispersion-corrected density functional theory calculations. *Acta Crystallographica Section B* **2010,** *66* (5), 544-558.

296. Le, C. T., *Applied Categorical Data Analysis and Translational Research*. Wiley: 2010.

297. Forte, R. M., *Mastering Predictive Analytics with R*. Packt Publishing: 2015.

298. Wu, Y.-F.; Zhang, H.; Wang, N.-X.; Chen, T.; Liu, Y., A Study on the Crystal Transformation Relationships of Valacyclovir Hydrochloride Polymorphs: Sesquihydrate, Form I, and Form II. *Crystal Research and Technology* **2021,** *56* (12), 2100084.

299. *ChemAxon Marvin*, 2022. https://www.chemaxon.com.

300. Abramov, Y. A., Current Computational Approaches to Support Pharmaceutical Solid Form Selection. *Organic Process Research & Development* **2013,** *17* (3), 472-485.

301. Fukte, S. R.; Wagh, M. P.; Rawat, S., Coformer selection: An important tool in cocrystal formation. *International Journal of Pharmacy and Pharmaceutical Sciences* **2014,** *6* (7), 9-14.

302. Wood, P. A. CSD Molecular Complementarity Tool Domain of Applicability 2022. https://www.ccdc.cam.ac.uk/discover/blog/csd-molecular-complementarity-domain-applicability/ (accessed on December 2022).

303. Udvarhelyi, A.; Rodde, S.; Wilcken, R., ReSCoSS: a flexible quantum chemistry workflow identifying relevant solution conformers of drug-like molecules. *Journal of Computer-Aided Molecular Design* **2021,** *35* (4), 399-415.

304. Lewis-Atwell, T.; Townsend, P. A.; Grayson, M. N., Comparisons of different force fields in conformational analysis and searching of organic molecules: A review. *Tetrahedron* **2021,** *79*, 131865.

305. Kanal, I. Y.; Keith, J. A.; Hutchison, G. R., A sobering assessment of small-molecule force field methods for low energy conformer predictions. *International Journal of Quantum Chemistry* **2018,** *118* (5), e25512.

306. Landrum, G. A., RDKit Documentation. 2019.

307. Wang, S.; Witek, J.; Landrum, G. A.; Riniker, S., Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences. *Journal of Chemical Information and Modeling* **2020,** *60* (4), 2044-2058.

308. Riniker, S.; Landrum, G. A., Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *Journal of Chemical Information and Modeling* **2015,** *55* (12), 2562-2574.

309. Eckert, F., COSMOtherm Reference Manual. Germany, 2017.

310. Klamt, A.; Diedenhofen, M., A refined cavity construction algorithm for the conductor-like screening model. *Journal of Computational Chemistry* **2018,** *39* (21), 1648-1655.

311. Brandenburg, J. G.; Bannwarth, C.; Hansen, A.; Grimme, S., B97-3c: A revised low-cost variant of the B97-D density functional method. *J Chem Phys* **2018,** *148* (6), 064104.

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

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

314. Keith, T. A. AIMAII software. http://aim.tkgristmill.com/index.html.

315. Neese, F., The ORCA program system. *WIREs Computational Molecular Science* **2012,** *2* (1), 73-78.

316. Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C., The ORCA quantum chemistry program package. *The Journal of Chemical Physics* **2020,** *152* (22), 224108.

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

318. Ellul, J. Polymorph and cocrystal screening of Ganciclovir for an alteration in its physical properties. University of Malta, 2021.

319. Vella-Zarb, L.; Baisch, U. Co-crystals comprising valacyclovir. WO 2019/038396 Al, 2019.

320. Solares-Briones, M.; Coyote-Dotor, G.; Páez-Franco, J. C.; Zermeño-Ortega, M. R.; de la O Contreras, C. M.; Canseco-González, D.; Avila-Sorrosa, A.; Morales-Morales, D.; Germán-Acacio, J. M., Mechanochemistry: A Green Approach in the Preparation of Pharmaceutical Cocrystals. *Pharmaceutics* **2021,** *13* (6), 790.

321. Friščić, T.; Jones, W., Recent Advances in Understanding the Mechanism of Cocrystal Formation via Grinding. *Crystal Growth & Design* **2009,** *9* (3), 1621-1637.

322. Loschen, C.; Klamt, A., COSMOtherm as a valuable tool for cocrystal screening and development. **2011**.

323. Al-Badr, A. A.; Ajarim, T. D. S., Chapter One - Ganciclovir. In *Profiles of Drug Substances, Excipients and Related Methodology*, Brittain, H. G., Ed. Academic Press: 2018; Vol. 43, pp 1-208.

324. Duggirala, N. K.; Frericks Schmidt, H. L.; Lei, Z.; Zaworotko, M. J.; Krzyzaniak, J. F.; Arora, K. K., Solid-State Characterization and Relative Formation Enthalpies To Evaluate Stability of Cocrystals of an Antidiabetic Drug. *Molecular Pharmaceutics* **2018,** *15* (5), 1901-1908.

325. Sandhu, B.; McLean, A.; Sinha, A. S.; Desper, J.; Aakeröy, C. B., Assessment of Computational Tools for Predicting Supramolecular Synthons. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2021,** *3* (2).

326. Hunter, C. A.; Sanders, J. K. M., The nature of .pi.-.pi. interactions. *Journal of the American Chemical Society* **1990,** *112* (14), 5525-5534.