Free Energy calculations for Crystal Structure Prediction studies

Stefanos Konstantinopoulos

A thesis submitted for the Doctor of Philosophy of Imperial College London and for the Diploma of the Imperial College

Supervisors:
Claire S. Adjiman
Constantinos C. Pantelides

Centre for Process Systems Engineering
Department of Chemical Engineering
Imperial College London
London SW2 2AZ
United Kingdom

December, 2023

Declaration of Authorship

I, Stefanos Konstantinopoulos, declare that this thesis titled, 'Free Energy calculations for Crystal Structure Prediction studies' and the work presented in it are my own. All information derived from other sources is appropriately referenced.

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND). Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

The accurate prediction of the crystal structure(s) that an organic molecule can be arranged into remains an unsolved theoretical problem of significant importance to industries delivering products in crystalline forms. From a thermodynamic standpoint, polymorphs can be identified as minima in the free energy landscape, motivating the development of crystal structure prediction (CSP) methods. CSP computational methods aim to determine the most stable crystal structure and other metastable forms, with the correct order of stability. Despite recent advances, most state-of-the-art CSP methods rely on lattice energy instead of Gibbs free energy to rank plausible crystal structures. Methods for evaluating the vibrational free energy of predicted crystals either employ computationally expensive atomistic DFT-d calculations or utilise force fields that adopt a rigid body treatment, neglecting intramolecular contributions.

In this thesis, we expand the applicability of the free energy calculation method proposed by Vasileiadis [1] to organic compounds that exhibit molecular flexibility. The proposed workflow for computing the Helmholtz free energy of molecular crystals, involves lattice energy minimisation using an atomistic force field and free energy calculation employing our in-house CrystalDynamics algorithm. The free energy calculation method is based on the harmonic approximation of lattice dynamics theory. The analytical derivations for the electrostatic expressions contribute to the development of an efficient yet accurate energy model applicable to a large number of structures generated in CSP studies. We apply the proposed free energy workflow to refine the CSP landscapes of three molecules of progressively increasing size and flexibility, as well as a set of experimental crystal structures of polymorphic flexible molecules with well-characterised enantiotropic transition temperatures. The workflow presented in this thesis can be utilised as a final refinement stage for CSP-generated landscapes of flexible molecules, bridging the gap between static lattice energies computed at 0 K and temperature-dependent free energies.

Acknowledgements

I would like to thank a number of individuals, whose support have been invaluable throughout the journey of my PhD. First and foremost, I would like to express my sincere gratitude to my supervisor, Claire Adjiman, for her unwavering guidance, support and insightful contributions, which have significantly shaped the outcome of this research. I am deeply thankful to my other supervisor, Costas Pentelides, for teaching me how to deal with challenging problems that initially seem daunting. I would like to express my gratitude to Isaac for his dedicated guidance and the rest of the members of the CSP group including David, Yizu, Ben, and Leticia for their creative input and collaborative spirit.

I am grateful to Spyros, with whom I have engaged in countless hours of fruitful discussions. My sincere thanks extend to my colleagues and ex-flatmates, Fabian and Matthias, for both the nice time both inside and outside the academic realm. I thank Srikanth and Maziar for our collaboration during the early stages of my PhD. I would also like to acknowledge Felippe, Suela, Karl, Harry, Kezheng, Lauren, Tom, Malak, Griffin, Ed and Sacr as well as the rest of the members of the Molecular Systems Engineering group for the warm and stimulating atmosphere.

I would like to also thank my good friends Tadej, Liakos and Foivos for their support. Finally, I would like to express my gratitude to my brother and my mother for their unconditional support.

Contents

1 Introduction **1**

1.1 Background and motivation 1

1.2 Thesis Scope and Outline 3

2 Literature Review: CSP and free energy **5**

2.1 Thermodynamic definition of free energy 5

2.2 Crystal Structure Prediction framework 6

2.3 Free energy calculations with lattice dynamics 10

2.3.1 Approximations in lattice dynamics theory 10

2.3.2 The importance of vibrational free energy in CSP 11

2.3.3 Application of lattice dynamics in CSP 12

2.4 Free energy calculations with molecular dynamics 13

2.5 Disorder in molecular crystals 14

2.6 Overview & Summary 15

2.7 Conclusions 17

3 Lattice dynamics for a tailor-made CSP force field **19**

3.1 Mathematical description of molecular crystals 19

3.2 Harmonic approximation 20

3.3 Thermodynamic functions 23

3.3.1 Vibrational free energy 23

3.3.2 Entropy, heat capacity and thermal energy 25

3.4 Energy model 26

3.4.1 Intramolecular interactions 27

3.4.2 Pairwise additive potentials 27

3.4.3 Repulsion-dispersion interactions 28

3.4.4 Electrostatic interactions 28

3.5 Dynamical matrix 30

3.5.1 Derivatives of repulsion-dispersion interactions 30

3.5.2 Generalised Ewald summation method 31

3.5.3 Dynamical matrix expression 32

3.6 CSP approach adopted in this thesis 34

3.7 Summary 36

4 Assessment of CSP workflows & local energy minimisation algorithms 37

4.1 The CrystalDynamics algorithm 37

4.1.1 Input file & Supported types of calculation 39

4.1.2 Implementation 41

4.2 Lattice energy minimisation 44

4.2.1 Bilevel minimisation of lattice energy 44

4.2.2 Consistency in computed lattice energies 46

4.3 Application of free energy workflow 50

4.3.1 Lattice energy gradients 50

4.3.2 Dispersion curves & pDOS 52

4.3.3 Free energy & Thermodynamic properties 54

4.4 Sensitivity of the CSP workflow to algorithmic parameters 57

4.4.1 Number of flexible DoFs during minimisation 58

4.4.2 Finite difference step size & LAMs 59

4.5 Conclusions 61

5 Free energy for refining CSP landscapes 63

5.1 Tetracyanoethylene CSP study 64

5.1.1 Global Search 66

5.1.2 Refinement stage I - Lattice energy 68

5.1.3 Refinement stage II - Free energy 70

5.2 Benzoic Acid CSP study 75

5.2.1 Global Search 75

5.2.2 Refinement stage I - Lattice energy 76

5.2.3 Refinement stage II - Free energy 78

5.3 Target XXIX from 7th CSP blind test 82

5.3.1 Global Search 82

5.3.2 Refinement stage I - Lattice energy 86

5.3.3 Refinement stage II - Free energy 88

5.4 Conclusions 91

6 On the use of lattice dynamics to investigate polymorphic stability 93

6.1 Benchmark set of enantiotropically related polymorphs 93

6.2 Selection of an energy model 96

6.2.1 Algorithmic parameter values 96

6.2.2 Choosing an energy model: Paracetamol 97

6.3 Lattice energy minimisation of 13 polymoprhs 98

6.4 Free energy evaluation for the 13 polymorphs 100

6.4.1 Methodology 100

6.4.2 Results 101
6.5 Conclusions 105

7 Conclusions **107**
7.1 Key contributions 107
7.2 Direction for future studies 108

Bibliography **110**

List of Figures

1.1 Crystal structures of the (a) YN (QAXMEH04) and (b) OP (QAXMEH03) polymorphs of ROY. The unit cell is highlighted by a parallelepiped. 2

2.1 A flowchart describing the multistage CSP methodology [1] 7

3.1 A typical diagram of a unit cell 20

4.1 Diagrammatic flowchart of CrystalDynamics algorithm. Names inside parentheses denote calls to external algorithms. Each box is discussed in the text and Figure 4.2. 38
4.2 CrystalDynamics algorithm in a step-by-step procedure. 39
4.3 Breakdown of CrystalDynamics input file named "vibrations.in" 41
4.4 CrystalDynamics Initial subroutines called for each type of calculation 42
4.5 CrystalDynamics subroutines for free energy calculation 43
4.6 Histogram of lattice energy gradients with respect to atomic positions per each direction for the monoclinic form of TCNE (Z=2, N=10) minimised with CrystalOptimizer. 52
4.7 Dispersion curves of the monoclinic form describing the relation of phonon frequencies along each component of the reciprocal space vector (a) $\boldsymbol{k}_x$, (b) $\boldsymbol{k}_y$ and (c) $\boldsymbol{k}_z$ for frequencies less than 10 THz. 53
4.8 (a) Dispersion curves along the magnitude of reciprocal space vector (b) pDOS of the monoclinic form of TCNE calculated as a histogram of (a). 54
4.9 Calculated relative Helmholtz free energy between the cubic and monoclinic forms of TCNE, relative to the cubic form, as a function of temperature. 55
4.10 Entropic contributions (Entropy multiplied with temperature) of the two polymorphs of TCNE as a function of temperature. Open circles indicate the entropic contributions evaluated using DFT-d at the PBE-TS level of theory [2]. 56
4.11 Isochoric heat capacity of the two TCNE polymorphs. Blue line represents the heat capacity value calculated using the equipartition theorem. 57
4.12 Calculated thermal energy for the two TCNE polymorphs. Open circles indicate the thermal energy evaluated using DFT-d at the PBE-TS level of theory [2]. . . 57

5.1 Molecular connectivity diagram of TCNE. 65

5.2 Unit cells of (a) cubic form and (b) monoclinic TCNE polymorphs. 66

5.3 CrystalPredictor landscape for TCNE molecule. Open circles indicate the structures obtained at the global search stage after performing clustering. The predicted forms that match the experimental ones are highlighted with a cube and rhombus. The blue line indicates the 20kJ/mol threshold from the global minimum. 67

5.4 CrystalOptimizer landscape. Open circles indicate minimised crystal structures. Structures matching the experimental cubic and monoclinic form are highlighted with a rhombus and a cube symbol. The blue line indicates the 20kJ/mol threshold from the global minimum. 69

5.5 (a) ZPE energy landscape of TCNE. Free energy landscape at (b) 0 K, (c) 100 K, (d) 200 K, (e) 300 K and (d) 400 K. Structures shown as red and blue squares represent the cubic and monoclinic forms, respectively. 71

5.6 Calculated relative entropic contributions to free energy for the TCNE landscape (a) 100 K, (b) 200 K, (c) 300 K and (d) 400 K. Calculated thermal energy for the TCNE landscape at (e) 100 K and (f) 400 K. Structures shown as red and blue squares represent the cubic and monoclinic forms, respectively. 73

5.7 Rank of the 76 structures of the TCNE free energy landscape at various temperatures relative to their rank at 0 K. 74

5.8 Molecular connectivity diagram of Benzoic Acid. 75

5.9 CrystalPredictor Z'=1 landscape of Benzoic Acid after clustering. The structure predicted 52nd in the lattice energy landscape (red square) matches the experimental form with RefCode BENZAC02. 76

5.10 Landscape of the 100 lowest energy structures for Benzoic Acid minimised with CSO-FM. The structure matching the experimental form (BENZAC02) is highlighted with a red square and. 77

5.11 (a) ZPE landscape of Benzoic Acid and free energy landscapes at (b) 0 K, (c) 100 K, (d) 200 K, (e) 300 K and (d) 400 K. The predicted structure matching the experimental form (BENZAC02) is highlighted with a red square. 79

5.12 Entropic contributions of the Benzoic acid landscape at (a) 100 K, (b) 200 K, (c) 300 K, (d) 400 K. Thermal energy for the Benzoic acid at (e) 100 K and (f) 400 K. The predicted structure matching the experimental form (BENZAC02) is highlighted with a red square. 80

5.13 Rank of the 83 structures of the Benzoic acid free energy landscape at various temperatures relative to their rank at 0 K. 81

5.14 The molecular connectivity diagram of Target XXIX. The blue arrows represent the dihedrals: O1-C2-C3-C4 (dih4), H5-C1-O1-C2 (dih5), and H1-N1-C8-C7 (dih16), which have been treated as flexible during the global search stage. . . . 83

5.15 2D heat maps for various conformations of molecule XXIX. (a) uniform & (b) adaptive LAMs for dih4 and dih5, (c) uniform & (d) adaptive LAMs for dih4 and dih16, and (e) uniform & (f) adaptive LAMs for dih5 and dih16. Calculations were performed at the PBE0/6-311+G(d,p) level of theory. 84
5.16 CrystalPredictor Z'=1 landscape for molecule XXIX after clustering. The yellow triangle indicates the locally minimised experimental structure (FASMEV), which has Z'=3 and was not found in the search. The red and green dashed lines indicate the 10 kJ/mol and 20 kJ/mol thresholds from the global minimum. 85
5.17 Molecular connectivity diagram of Anthranilic Acid taken from CSD RefCode: AMBACO08. 86
5.18 CSO-FM landscape for molecule XXIX containing 100 CSO-FM minimised structures. The req square indicates the locally minimised experimental structure (FASMEV) 87
5.19 (a) ZPE energy landscape of XXIX. Free energy landscape at (b) 0 K, (c) 100 K, (d) 200 K, (e) 300 K and (d) 400 K. The red square represents the ZPE and free energy at various temperatures of the experimental form (FASMEV), which was not found in the CSP study. 89
5.20 Calculated relative entropic contributions to free energy for the landscape of Target XXIX at (a) 100 K, (b) 200 K, (c) 300 K and (d) 400 K. The red square represents the relative entropic contributions of the experimental form (FASMEV), which was not found in the CSP study 90
5.21 Rank of the 79 structures of Target XXIX free energy landscape at various temperatures relative to their rank at 0 K. 91

6.1 Molecular connectivity diagrams of all compounds included in the benchmark set. 94
6.2 Free energy difference relative to the low-temperature form ($\Delta\Delta A$) vs temperature for (a) Pyrazinamide, (b) Adenine, (c) Paracetamol, (d) Piracetam, (e) Theophylline and (f) Carbamazepine. 102

List of Tables

2.1 Summary of studies that include an evaluation of the free energies of CSP-predicted structures as part of the ranking procedure. Abbreviations not defined in the text: 6BT: Target molecules XXII - XXXVI of the 6$^{\text{th}}$ blind test, DMA: distributed multipoles analysis, PC: point charges, MetadD: Metadynamics. . . . 16

4.1 Default parameters values used in the CrystalDynamics algorithm. 40

4.2 Compounds studied in this Chapter. More details can be found in the relevant Sections in Chapter 5. 48

4.3 Comparison of intermolecular energy contributions calculated with CSO-FM and CrystalDynamics for various real space cutoffs (r_c^{real}) and splines fractions (ζ_s). Calculations are performed for the lowest energy structure of the CrystalPredictor landscape of Target XXIX. Energies are reported in kJ/mol. 49

4.4 Parameters values used in CrystalDynamics algorithm for calculations shown in Section 4.3. 50

4.5 Impact of the number of flexible DoFs on the stability of the 100 lowest energy structures of Benzoic Acid on the CSO-FM landscape 59

4.6 Effect of altering the FD step size for bonds in the landscape of Benzoic Acid. . . 60

4.7 Effect of using LAMs on the number of stable structures in the landscape of Target XXIX. 61

5.1 Geometry comparison of the cubic and monoclinic form generated at the CrystalPredictor stage with the experimental TCNE crystal structures in CSD. 68

5.2 Geometry comparison between the cubic and monoclinic form at refinement stage I performed with CrystalOptimizer. 69

5.3 Spearman rank correlation coefficient at various temperatures relative to 0 K. . . 74

5.4 Spearman rank correlation coefficient at various temperatures relative to 0 K for Benzoic acid. 81

5.5 Values of flexible dihedrals obtained from atomistic geometry optimisations of TCNE with the PBE0 functional and three different basis sets 83

5.6 Geometry and energy of the three Anthranilic Acid polymorphs locally minimised with CrystalOptimizer using two different parameter sets for the repulsion-dispersion interactions. 87

5.7 Spearman rank correlation coefficient at various temperatures relative to 0 K for Target XXIX. 91

6.1 Compound names, CSD RefCodes, names of polymorphs, Z/Z' numbers and enantiotropic transition temperatures of all compounds used in the benchmark set. . 95
6.2 Parameters values used in CSO-FM and CrystalDynamics for the lattice energy minimisation and free energy evaluation of all experimental structures of benchmark set. 96
6.3 Lattice energies and Helmholtz free energies at 0 K of the two paracetamol polymorphs using different levels of theory. 98
6.4 Geometry comparison between the experimental structure (see CSD RefCode) and locally minimised experimental structure with CSO-FM for all 13 structures in the benchmark set. Lattice energies are calculated at the PBE0/6-311+ level of theory using FIT repulsion-dispersion parameters. 99
6.5 Calculated $\Delta\Delta A$ at 0 K and 400 K and $T\Delta S$ at 400 K of all compounds in the benchmark set. The differences are calculated relative to the form that is known experimentally to be more stable at low temperatures. All reported energies are in kJ/mol. 104

Chapter 1

Introduction

1.1 Background and motivation

Organic crystals find numerous applications in pharmaceutical, agricultural, food, explosives and semiconductors industries [1]. The study of organic crystalline materials also poses an intriguing challenge for academia as the relation between the structure of a molecule and its most stable crystal structure is not generally well understood [3], especially as concerns an interesting phenomenon occurring in crystals, namely polymorphism [4]. Polymorphism refers to the ability of chemical compounds to exist in different crystalline forms, which are known as polymorphs. Various physical properties such as solubility, dissolution rate, morphology and mechanical strength depend on the crystal structure and as a result may vary significantly between different polymorphs. To date the compound with the largest number of characterised polymorphs at room temperature is 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecabonitrile. This molecule is commonly referred to as ROY, taking its name from the red (R), orange (O) and yellow (Y) colours of the experimental polymorphs [5, 6]. The crystal structures of the YN and OP polymorphs of ROY are shown in Figure 1.1. The crystal structures are taken from the Crystal Structure Database (CSD) with RefCodes QAXMEH04 and QAXMEH03 for the YN and OP polymorphs, respectively.

Crystal structure prediction (CSP) is the computational field of research that seeks to examine the most stable crystal structure of a target molecule(s) using very limited amount of information. The objective of CSP has been described by Pantelides *et al.* [7]: "*Given the molecular diagrams for all chemical species (neutral molecule(s) or ions) in the crystal, identify the thermodynamically most stable crystal structure at a given temperature and pressure, and, in correct order of decreasing stability, other (metastable) crystal structures that are also likely to occur in nature.*"

CSP can offer valuable insights for structural characterisation by providing putative structures to interpret experimental results, such as powder X-ray diffraction (PXRD) patterns, and by aiding in development of experimental protocols for the crystallisation of previously unobserved polymorphs [8]. Another important potential use of CSP is to examine the relation between crystal structures and properties, such as solubility, which can affect the bioavailability of pharmaceuticals delivered in crystalline form. The development of CSP methods thus aims

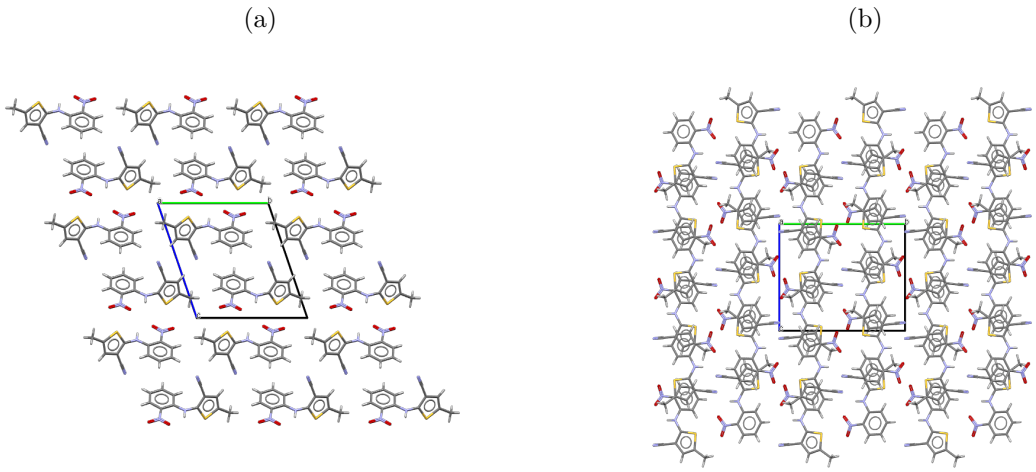


Figure 1.1: Crystal structures of the (a) YN (QAXMEH04) and (b) OP (QAXMEH03) polymorphs of ROY. The unit cell is highlighted by a parallelepiped.

to complement experimental findings by enhancing the understanding of the crystal landscape of the target compound(s). This understanding can contribute to risk assessments conducted during drug development or assist in identifying optimal drug formulations and manufacturing conditions, thereby preventing costly interruptions of supply. A well-known example that has motivated the development of CSP is the case of ritonavir [9]. Ritonavir, marketed under the brand name Norvir, has been developed to treat AIDS. The unexpected emergence of a previously unidentified and notably more stable polymorph of ritonavir led in the withdrawal of Norvir capsules from the market. The case of ritonavir threatened the supply of the Norvir life-saving AIDS treatment and resulted in a estimated 100-1000 million USD loss in revenue. The progress of the CSP methodologies is captured by a series of blinds test [10, 11] organised by Cambridge Crystallographic Data Centre (CCDC). Current state-of-the art CSP techniques can effectively model large flexible molecules, salts, hydrates and co-crystals of industrial relevance.

From a thermodynamic standpoint, Gibbs free energy is the appropriate metric governing polymorphic stability. The determination of the energetically more favourable crystal structure(s) requires a detailed investigation of the free energy landscape. At a given temperature and pressure, the most stable crystal structure corresponds to the global minimum in the Gibbs free energy landscape. Local minima in the Gibbs free energy surface are commonly referred to as metastable structures. The sets of candidate crystal structures produced in CSP studies are ranked by increasing energy. However, much of the focus of CSP efforts to date has been on identifying structures using lattice energy as the stability metric, without considering thermal and vibrational contributions to free energy. These neglected contributions prohibit the study of crystalline stability as a function of temperature. The increased computational cost, associated with the large number of structures generated in CSP studies, requires a method that provides rapid estimates of the free energy. The work presented in this thesis aims to bridge the gap between the lattice energies and free energies by employing a lattice dynamics method that takes

into account molecular flexibility.

1.2 Thesis Scope and Outline

A fundamental hypothesis in CSP is that the outcome of crystallisation is governed by thermodynamics [12]. Despite advances in accurately modelling interatomic forces, current state-of-the-art CSP techniques often rely on identifying experimental forms as low-energy minima at 0 K [13, 1]. However, at finite temperatures and pressure, thermodynamics enable the determination of relative stability and provide a link to experimentally measured properties, such as solubility. This necessitates moving beyond static energies calculated at 0 K and utilising accurate and efficient methods required for refining a variety of predicted structures generated in CSP studies. Our group at Imperial College London has developed a multistage methodology for performing CSP studies over the past two decades. The initial stage is focused on determining the conformational degrees of freedom that are treated as flexible and selecting an appropriate energy model. In the second stage, a global search is conducted to identify candidate crystal structures as low-energy minima on the lattice energy landscape. This is achieved by employing simplistic energy models that account for limited flexibility and utilise point charges to model electrostatic interactions. The third stage involves lattice energy minimisation for the most promising low-energy structures identified in the previous stage. In this stage, more detailed energy models are used, incorporating distributed multipole analysis for evaluating electrostatics and allowing for the treatment of extended molecular flexibility. Finally, in the fourth stage, the free energy is calculated using the harmonic approximation of lattice dynamics theory, with an energy model consistent with the one used in the previous stage. Further details of our CSP framework can be found in Section 3.6.

A methodology for evaluating the free energy of molecular crystals has been developed by Vasileiadis [14]. This methodology is based on the harmonic approximation of lattice dynamics theory and has been successfully used in ranking predicted structures generated in CSP studies of rigid molecules. In this thesis, the existing CSP workflow is extended to incorporate the evaluation of the free energy of flexible molecules. Our objectives involve examining the impact of molecular flexibility on the validity of the harmonic approximation and assessing the suitability of the proposed workflow in determining the stability of CSP-generated structures as a function of temperature. In addition, we investigate the suitability of the extended CSP workflow to reproduce the relative stability of experimental polymorphs across a range of temperatures.

The outline of this thesis is as follows. In Chapter 2, we define the free energy of molecular crystals in the context of chemical thermodynamics and review existing CSP methods used for generating and ranking predicted structures. Emphasis is placed on the few studies that have employed free energies as part of their ranking procedure. In Chapter 3, we introduce the formalism for calculating the free energy of molecular crystals. The utilised energy model is described in detail and we derive analytical expressions for performing the harmonic approximation within lattice dynamics theory. In Chapter 4, we describe the CrystalDynamics algorithm

and implementation. Additionally, we present our CSP workflow for evaluating the vibrational free energy contributions of flexible molecules. In Chapter 5, we perform full CSP studies for three compounds; a small rigid molecule, Benzoic acid, which exhibits one flexible degree of freedom, and Target XXIX from the 7^{th} blind test. These CSP studies includes global search and two refinement stages using lattice energy and free energy, respectively. In Chapter 6, we apply our free energy workflow to a set of polymorphic systems with well-characterised enantiotropic transition temperatures. The objective is to assess the suitability of our workflow in reproducing the experimentally determined relative stability. In Chapter 7, we present the main conclusions of this thesis and provide directions for future studies.

Chapter 2

Literature Review: CSP and free energy

In this chapter, we initially introduce key concepts necessary to understand this thesis. We begin by defining the Gibbs free energy for molecular crystals under the chemical thermodynamics framework. In turn, we describe the multistage approach adopted in CSP and provide characteristic examples of methodologies that have been successfully employed in CSP studies. We review alternative existing free energy evaluation methods used either in ranking structures generated in CSP studies or for determining polymorphic stability. The gaps in the literature are revealed to provide motivation for the work presented in this thesis.

2.1 Thermodynamic definition of free energy

The aim of this section is to provide a brief description of thermodynamic quantities as well as other metrics that have been used for ranking structures generated in CSP studies. The chemical thermodynamics framework provides the formalism for the description of a system by means of the so-called characteristic function [15, 16]

$$U = U(S, V), \tag{2.1.1}$$

where the internal energy, U, is expressed as a function of its natural variables entropy, S and volume, V. Equation (2.1.1) can be used as a basis for the evaluation of any other thermodynamic quantity of interest. In practice, this form of the characteristic function is not widely used as the natural variables can often be difficult to measure experimentally. For example, the lack of a method for directly measuring entropy hinders its use. The characteristic function can be transformed into equivalent forms, which are known as thermodynamic potentials. Thermodynamic potentials fully characterise the state of a system and provide information about its equilibrium condition. The Legendre transformation is commonly used to derive one thermodynamic potential from another by expressing the energy in terms of different pairs of conjugate variables (e.g., entropy and temperature or volume and pressure).

For a pure substance in any state, a suitable thermodynamic potential used as the metric to determine polymorphic stability at a given temperature, T, and pressure, P, is the Gibbs free energy, $G(T, P)$. For a perfect crystal, a Gibbs free energy difference, $\Delta G(T, P)$, is often used and defined relative to the ground-state energy of an isolated molecule at 0 K, with no thermal

and zero point motion, as

$$\Delta G(T, P) = \Delta U + PV - TS, \tag{2.1.2}$$

where ΔU is the difference in internal energy with respect to the isolated molecule reference state, V is the molar volume of the unit cell, and S is the entropy of the bulk crystal. Note that the entropy of the isolated molecule at 0 K reference state is zero. At a given temperature and pressure, the most stable crystal structure corresponds to the global minimum in the Gibbs free energy landscape, while local minima are referred to as metastable structures. The Gibbs free energy provides a more direct link to practically important and experimentally measurable properties, such as solubility and specific heat capacity, than the internal energy.

The PV term in equation (2.1.2) can usually be neglected under ambient conditions of temperature and pressure due to its small contribution to the overall free energy. At atmospheric pressure, the contribution of the PV term is typically on the order of J/mol, while the Gibbs free energies are on the order of kJ/mol. This significant difference in magnitude justifies the common practice of using Helmholtz free energy instead of Gibbs free energy as the stability criterion. The accurate evaluation of the entropic and thermal contributions for numerous structures, as generated in CSP studies, can be challenging due to the computational cost and complexity associated with the calculation of those contributions. To overcome this challenge, approximations of the free energy are often employed to determine the relative stability of structures generated in CSP studies. A common approach is to assess relative stability by considering differences in internal energy by neglecting the PV and TS terms. At 0 K, this approximation is further simplified to equating the difference in Gibbs free energy between two crystals to their lattice energy difference. This approximation is justified by the fact that at 0 K, the lattice energy difference is expected to make the biggest contribution to relative stability. A breakdown of the contributions to internal energy can be found in Section 2.3. The validity of this assumption is examined in this thesis by assessing the effect of thermal contributions (see Section 2.3) on the relative stability of CSP generated structures across a range of temperatures.

2.2 Crystal Structure Prediction framework

The main objective of CSP is to produce a ranked list of plausible crystal structures based solely on the molecular connectivity diagram(s) of the molecule(s) under investigation [7, 11, 17]. The ranking is achieved by ordering crystal structures in terms of ascending (free) energies. A number of CSP methodologies have proven to be effective in examining the crystal landscapes of organic molecules [11]. Successful CSP methodologies have adopted multistage approaches, usually comprising a candidate generation stage and multiple refinement stages, allowing to tackle different problems at each stage [1]. The goal of the candidate generation stage is to generate a sufficient number of diverse crystal structures that will be subsequently used as starting points in the refinement stages.

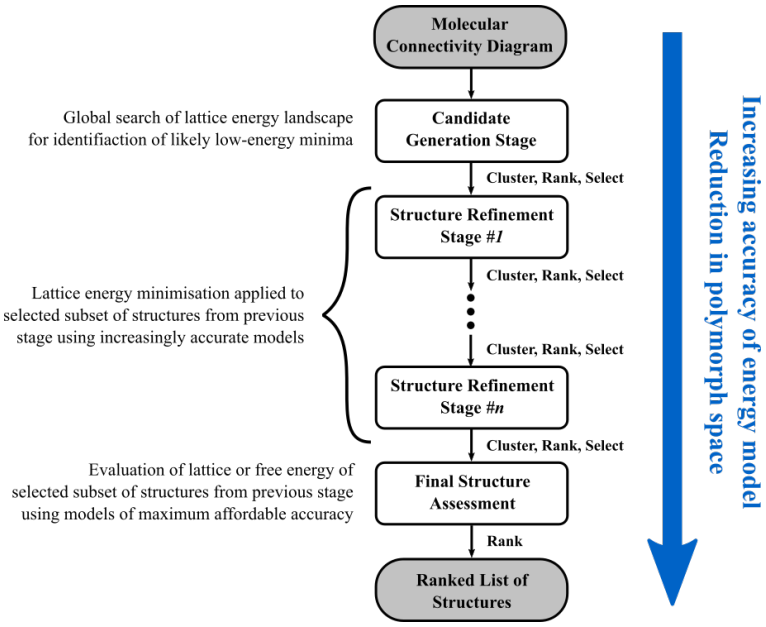


Figure 2.1: A flowchart describing the multistage CSP methodology [1]

The high dimensionality of the crystalline landscapes requires the minimisation of a large number of structures (hundreds of thousands to millions). The predicted structures are generated using simplified and computationally-efficient energy models that allow the thorough exploration of the crystal energy landscape. Our choice to trade off accuracy for efficiency is justified by the fact that a cheaper energy model, while potentially less effective for ranking, can reproduce the crystalline geometry with sufficient accuracy. In the 5^{th} blind test [10], calculations performed with MD simulations using a force field with transferable parameters predicted the experimental forms with satisfactory geometry reproduction, however performed poorly in terms of energy ranking. The most promising structures are those identified as having the lowest energy during the global search stage. Energy cutoffs (thresholds) are applied to narrow down the number of structures that are fed into the subsequent stages and are refined adopting more accurate but computationally expensive energy models. The number of the "low" energy structures a that are fed to the subsequent refinement stages ranges from hundreds to thousands and strongly depends on the computational cost of the utilised methods. For example, structures within 20 kJ/mol from the global minimum of the CrystalPredictor landscape are typically minimised using CrystalOptimizer algorithm [18]. More details for the CrystalPredictor and CrystalOptimizer algorithms can be found in Section 3.6.

The methods employed in the candidate generation stage can be broadly categorised into biased and unbiased approaches. Unbiased methods involve the generation of numerous initial structures with the goal of broadly sampling the space of putative crystal structures. Each initial structure is minimised locally, resulting in a ranked list of potential structures. A notable characteristic of unbiased methods is the equal allocation of computational effort to both low and high energy regions of the crystal energy landscape, even though the latter regions are of less importance. The initial crystal structures are generated by considering a grid in the space

defining the crystal structures. The sampling of the space of crystal structures can be performed in a uniform (homogeneous) manner, however non-uniform approaches have been proven to sample more effectively the phase space. A typical example is the quasi-random sampling based on Sobol' sequence [19] as implemented in the CrystalPredictor algorithm [20, 21, 22] and GLEE [23]. In the CSP workflow of XtalpI [24, 25], the sampling of the space of crystal structures during the candidate generation stage in performed randomly.

Biased methods seek to explore more promising parts of the energy landscape, by allocating computational resources closer to minima in the landscape. In Monte Carlo simulated annealing (MCSA) [26], starting from a random structure a new neighbouring structure in the landscape is accepted unconditionally if it has lower energy. In contrast, if its energy is higher than the previous structure, the new structure is accepted using the Metropolis criterion. The temperature is increasing and decreasing multiple times to allow structures escape local minima. Parallel tempering is an extension MCSA that offers the exchange of configurations between multiple MCSA simulations running in parallel. The latter method has been implemented in GRACE [27] and has demonstrated very good performance in the 5^{th} [10] and 6^{th} [11] CCDC blind tests. Recently, genetic algorithms have been employed in the candidate generation stage, with algorithms like Gator [28] utilising evolutionary niching [29] to enhance the diversity of sampled crystal structures.

The most promising structures, identified in the candidate generation stage, are refined using more detailed energy models. The approaches utilised for CSP refinement can be broadly categorised into force field and electronic structure methods. A force field is a mathematical model that describes the potential energy of the system as a function of interatomic positions. Force fields with transferable parameters assigned to specific atom types have routinely been employed to model interactions in molecular simulation studies. In CSP, the force fields parameters can be either tailor-made or transferable. The term tailored-made CSP force field is used throughout the text to describe force fields parameterised using QM calculations for a specific molecule. Tailor-made parameters are derived for the molecule(s) of interest using quantum mechanical (QM) calculations offering increased accuracy at a higher computational cost. For instance, point charges [30, 25] or multipole moments [31] can be fitted to charge densities obtained from electronic structures methods. The use of tailor-made parameters aims to maintain the accuracy of the fitted charge densities, while keeping the computational cost low. Also, the tailor-made parameters can be used in simulations to evaluate thermodynamic properties using consistent energy models [24].

In electronic structure methods, on the other hand, the crystal is represented at the level of nuclei and electrons. A variety of commercial and academic software exist for performing periodic DFT calculations. The implementations differ mainly on the mathematical representation of the basis used to describe the charge density, whether plane waves [32], atom-centred Gaussians [33] or natural atomic orbitals [34]. The success of DFT relies on an accurate approximation of the exchange-correlation energy based on the electron density of the system. A hierarchy of existing exchange-correlation functionals has been proposed by Perdew & Schmidt [35]. The

proposed methods include local density approximations, generalised gradient approximations (GGAs), meta-GGAs and hybrid functionals, each of which offers a more accurate description of electronic structure at an increased computational cost. The (semi)local nature of standard DFT approaches neglects long-range dispersion interactions and this results in poor performance for molecular crystals. Dispersion corrected density functional theory (DFT-d) has been successfully applied to address this issue for optimising and ranking structures generated in CSP studies. Notably, the DFT-d methods proposed by Tkackenko and Shefller [36] and the D3 correction by Grimme *et al.* [37] adopt a classical treatment of dispersion as oscillations of multipole moments, which offers good balance between accuracy and efficiency. Beyond pairwise effects, many body dispersion (MBD) interactions [38] provide high accuracy energies but are too expensive for geometry optimisation of molecular crystals. DFT-d methods have gained increased popularity in recent years. In the 5^{th} blind test [10] just 2 groups employed such methods, with the number of groups growing to 12 in the 6^{th} blind test [11].

An important consideration in both candidate generation and refinement stages is the treatment of molecular flexibility. Although an atomistic approach allows for a full representation of molecular flexibility, increased efficiency can be achieved by allowing only some fragments of the molecules to vary freely, while keeping the rest of the molecular geometry rigid. DFT-d calculations are performed in an atomistic manner, whereas force field methods may adopt a fully or partially rigid treatment. Most commonly, a smaller number of degrees of freedom (DoFs) is treated as flexible in the candidate generation stage. As an example, in the global search stage of CrystalPredictorII [22, 21], a few flexible DoFs can be treated as optimisation variables accounting for limited molecular flexibility. The global search in the global lattice energy explorer (GLEE) algorithm [23] involves the generation of multipole conformers prior to minimisation, each of which is minimised adopting a rigid-body approach. Force fields can be also employed in the refinement stages, with DMACRYS [39] and CSO-RM [40] employing a fully rigid approach, optimising the interactions between molecules using a polarisable and anisotropic force field. In other force field-based approaches, such as CrystalOptimizer [18] and CSO-FM the conformational DoFs that are treated as flexible are specified as user-defined optimisation variables. In this thesis, the CrystalPredictorII algorithm is utilised in the candidate generation stage, while the lattice energy minimisation (refinement stage I) is carried out using either DMACRYS/CrystalOptimizer or CSO-RM/CSO-FM. More details regarding the CrystOpt/DMACRYS and CSO-RM/FM lattice energy minimisation algorithms can be found in Section 4.2.1.

As discussed in Section 2.1, the Gibbs free energy is the thermodynamic quantity that governs the stability relation of pure crystals at finite temperature and pressure. However, consideration of free energy is not yet widespread in CSP studies due to the often-limited accuracy of the computed free energy values and the high computational cost of obtaining them. We review the free energy methods that have been used as part of multistage CSP methodologies in the remaining of this chapter. The free energy methods used in CSP can be divided into three categories: Lattice dynamics theory for calculating vibrational free energies, free energy calculation

methods based on molecular dynamics (MD) simulations and methods that account for disorder in molecular crystals, which are discussed in Sections 2.3, 2.4 and 2.5, respectively.

2.3 Free energy calculations with lattice dynamics

The vibrational contribution to free energy can be evaluated using the lattice dynamics (LD) theory that was originally developed by Born and Huang [41]. In this section, we introduce the approximations used in LD theory and describe the importance of capturing vibrational contribution to free energy for the case of molecular crystals. We also review existing LD methods that have been applied to rank CSP generated structures using vibrational free energy as the criterion determining relative stability.

2.3.1 Approximations in lattice dynamics theory

LD theory [41, 15] provides a formalism for the description of atomic and molecular vibrations within a crystal lattice. Estimates of both the zero point energy (ZPE) and thermal contributions to vibrational free energy, $A^{\mathrm{vib}}(T)$, can be obtained from phonon frequencies, including the effects of both entropy and thermal energy. Within the LD framework, the Gibbs free energy of a crystal is given by:

$$\begin{aligned}\Delta G(T,P) &= \Delta U^{\mathrm{latt}} + PV + A^{\mathrm{vib}}(T) \\ &= \Delta U^{\mathrm{latt}} + PV + U^{\mathrm{ZPE}} + U^{\mathrm{thermal}}(T) - TS\end{aligned} \tag{2.3.1}$$

where ΔU^{latt} is the lattice energy, which is defined as the energy difference between the crystal and gas states and is discussed in more detail in Section 3.4. A^{vib} collectively denotes the vibrational contributions to free energy. These include the ZPE, U^{ZPE}, thermal energy, $U^{\mathrm{thermal}}(T)$, and entropic contribution to free energy, TS. Note that definition of $\Delta G(T,P)$ given in Equation (2.3.1) is equivalent with Equation (2.1.2), as for a molecular crystal ΔU is equal to the sum of ΔU^{latt}, U^{ZPE} and $U^{\mathrm{thermal}}(T)$. More details regarding the thermal contributions can be found in Subsection 3.3.2.

The harmonic approximation (HA) of LD theory was first introduced by Born and Huang [41]. Within the HA, the potential energy of the crystal is approximated by a quadratic (harmonic) function using a second-order series expansion around an equilibrium position, typically a local minimum on the lattice energy hypersurface. This harmonic potential allows the derivation of analytical expressions for the vibrational free energy and associated thermodynamic quantities based on phonon frequencies. The derivation of the HA equations for molecular crystals and our CSP force field are described in Chapter 3. The HA is restricted to the evaluation of thermodynamic properties at constant volume, such as the isochoric heat capacity (C_{V}). The equilibrium geometry of the unit cell is commonly obtained via a lattice energy minimisation and remains fixed at all temperatures. The HA is expected to perform better at lower temperatures since at elevated temperature the effects of thermal expansion and anharmonic vibrational

contributions become more significant. At temperatures above the melting point, the HA is not applicable as the periodic long-range order of the crystal breaks down.

The quasi-harmonic approximation (QHA) [42, 43, 44, 45] is an extension of the HA that takes into account the effect of temperature on the geometry and volume of the unit cell. In the QHA, multiple HA calculations are performed for different unit cell geometries, allowing the minimisation of $\Delta G(T, P)$ with respect to volume at different temperatures. Each considered geometry corresponds to same crystal structure or polymorph at different temperatures accounting for thermal expansion, which can be either isotropic or anisotropic [46]. QHA facilitates the evaluation of properties such as the isobaric heat capacity (C_{P}). It is expected that QHA is more reliable than HA for determining relative stability and evaluating thermodynamic properties. However, QHA requires knowledge of the dependence of unit cell volume on temperature and comes with a significant increase in cost as multiple HA calculations have to be performed for each considered unit cell volume.

An additional extension of LD is the Anharmonic Approximation (AA), in which anharmonic contributions to the free energy are taken into account [47, 48, 49]. In most cases, the AA involves incorporating higher order terms in the expansion series to capture the non-quadratic shape of the lattice energy minimum. LD is the most widespread approach for computing vibrational free energies in CSP studies. Although the anharmonic approximation can provide more accurate results, HA and QHA are usually preferred due to their relative simplicity and lower computational cost [47, 50, 14]. In the remainder of this section, we highlight the importance of including vibrational contributions to free energy. In addition we present those studies that have performed calculations using the LD theory as part of the CSP multistage procedure, emphasising the use of HA and QHA in the 6^{th} blind test.

2.3.2 The importance of vibrational free energy in CSP

There have been efforts to quantify the contributions of ZPE and temperature to the enthalpy and entropy of polymorphs. Nyman and Day [51] have conducted the largest study to date of rigid-body vibrational free energies. They found that in 9% of the 601 polymorphic pairs examined, the inclusion of these terms caused a re-ranking of polymorphic stability between 0 K and room temperature. The study also confirmed that the differences in ZPE and isochoric heat capacities between polymorphs are small, and highlighted the role of entropy as the main contributor to changes in free energy differences. In a later study [45], the Gibbs free energies of 864 crystal structures (475 polymorphic pairs) were evaluated between 0 K and the melting point using both HA and QHA. Approximately 20% of the polymorphic pairs were re-ranked, highlighting again the importance of temperature effects on polymorphic stability.

Overall, the inclusion of ZPE and thermal contributions is an important consideration if we are to bridge the gap between computed and experimental crystal structure landscapes. In recognition of this, free energy contributions have been incorporated within benchmark test sets such as POLY59 [52] and X23 [53].

2.3.3 Application of lattice dynamics in CSP

To date, there has been a limited number of studies where free energy contributions have been included in the refinement stages of CSP. Most commonly, lattice energy is used instead of free energy as the criterion for determining the relative stability. This approximation is attributed to the increased computational cost associated with computing free energies for a large number of structures generated in CSP studies. In most cases, free energies are only calculated for the most promising structures identified as having the lower lattice energy during an intermediate refinement stage. In [50], the rigid-body vibrational free energies of the 100 lowest energy structures of Chloridazon were evaluated using DMACRYS [39] employing the co-prime linear splitting. In the PhD thesis of Vasileiadis [14], the HA was employed to refine the 200 lowest energy structures of Tetracyanoethylene and Imidazole using an atomistic force field. In these studies, the free energy refinements typically resulted in an improved ranking of the experimental form(s). In [47], the five target molecules of the sixth blind test were revisited, by calculating the harmonic vibrational free energies at the PBE+TS level of theory. For all cases, except the polymorphic target XXIII, the vibrational free energies were calculated for 4-8 structures per molecule, with the experimental forms emerging as the lowest free energy structures. In the case of target XXIII, the HA was applied to 46 structures while QHA and AA were employed for 9, suggesting the existence of a yet undiscovered form as the most stable. In a very recent study [54], a free energy benchmark set has been applied to multiple CSP generated landscapes for evaluating the stability and solubility of molecular crystals. This study, allowed the evaluation of the free energy of various compounds at different percentages of relative humidity.

In the 6$^{\text{th}}$ blind test [11], 6 out of 25 participating groups performed LD calculations to rank structures based on free energies in their submitted lists. The groups of Price & coworkers and Day & coworkers evaluated Helmholtz free energies by adopting a rigid-body treatment utilising anisotropic force fields. The groups of Pickard & coworker, Boese, and Tkatchenko & coworkers used DFT-d, while the group of Brandenburg & Grimme utilised HF-3c methods to perform vibrational free energy calculations within the HA. The number of structures examined in this context was heavily dependent on the approximations and cost of the underlying lattice energy models as well as utilised free energy methods. As an example, using rigid-body force fields, the group of Day & coworkers applied HA on a total of 992 structures using the co-prime linear splitting method [45, 51]. On the other hand, groups utilising DFT-d [55, 37] or HF-3c [56] methods were typically able to evaluate harmonic free energies for just a few structures. The AA was used by the group of Pickard & coworkers to evaluate the ZPE of a single structure by employing the vibrational self-consistent field theory [49]. In general, the ranking of the computed polymorphs based on free energy was found to be more accurate, relative to experimental evidence, compared to lattice energy ranking. The experimental structures of molecules XXII, XXV and XXVI, were predicted as the global minima in the free energy landscape by three of the groups using the HA. For Form B of molecule XXIII, Brandenburg & Grimme predicted the experimental polymorph as the global minimum by including the ZPE in

the lattice energy.

2.4 Free energy calculations with molecular dynamics

Molecular dynamics (MD) is a molecular simulation technique that allows the study of classical systems for various thermodynamic ensembles [57]. Under the assumption of ergodic sampling, the time averages in MD simulations are equivalent to ensemble averages, enabling the calculation of equilibrium and transport properties. At finite temperature and pressure, MD provides a natural method for accounting for the effects of anharmonicity, such as thermal expansion and the temperature dependence of phonon frequencies. The restriction of MD to short time scales, in combination with the presence of high energy barriers in solid-solid transitions, limits the applicability of conventional MD methods to molecular crystals [57]. To overcome this limitation, a number of advanced MD methods have been developed for the study of crystalline materials. Common MD methods used for calculating free energies of molecular crystals include adiabatic free energy dynamics [58, 59], path-integral MD [60], orthogonal space random walk [61] and multistate Bennett acceptance ratio (MBAR) [62, 63]. The accuracy of MD simulations is inherently dependent on the utilised force field. In most cases, force field with transferable parameters, such as OPLS [64], AMBER [65] and DREIDING [66], are used to quantify the intramolecular and intermolecular interactions. More accurate force fields have also been proposed, such as AMOEBA [67] that uses multipole moments and accounts for induction effects. Force fields with transferable parameters perform well for liquids, but are not great for CSP [10]. Tailor-made force fields, fitted to periodic DFT-d calculations [1, 25, 30] perform better in terms of ranking structures generated in CSP studies in comparison to force fields using transferable parameters.

MD simulations have been applied to in predicted structures generated in CSP studies only in a handful of cases [68]. Metadynamics is a powerful method for exploring the free energy landscape of complex systems based on collective variables (CVs) [69, 70]. The sampling of the configuration space is enhanced by adding a history-dependent term as a bias potential that is a function of these CVs. The history-dependent term prevents the system from being trapped in deep local minima, facilitating the study of transitions between forms. The successful application of metadynamics to molecular crystals requires the identification of suitable CVs [71, 68, 72]. Metadynamics were applied to refine the CSP-generated landscapes of 5-fluorouracil [73] and the industrial pigment red (PR179) [74]. In the former case, metadynamics was used to refine the 60 lowest lattice energy structures, leading to 25% fewer free energy minima at ambient temperature and pressure. Form II was predicted to be the most stable structure at 0 K and 0 Pa and remained unchanged without undergoing a phase transition during the metadynamics simulation. Conversely, the experimentally most stable structure, Form I, transitioned into a disordered structure after just a few metadynamics steps. For PR179, a total of 18 candidate structures, generated through CSP at 0 K and 0 Pa, were employed as initial configurations for metadynamics simulations. The most stable form, Form I, emerged as the global minimum in

both the lattice energy and free energy landscapes. In a very recent study [75], the crystallisability of two sulfonated compounds was investigated using a combined CSP and metadynamics approach without embedding any prior knowledge of experimental information. Sulflower is known to be crystallisable, while despite extensive experimental efforts, persulfurated coronene (PSC) appears only in amorphous form. The CSP study involved structure generation and rigid-body refinement using CrystalPredictorII [22] and DMACRYS [39], respectively. The effect of temperature was captured via NPT and NVT MD simulations and well-tempered metadynamics (WTMD) for the 20 lowest energy structures generated in the CSP study. The experimental structure of sulflower was predicted as the lowest energy structure that did not transform into an amorphous form. In contrast for PSC, all the generated crystal structures were found to convert to disordered systems under small amounts of work indicating that it is not crystallisable. In addition, for the two sulfonated compounds, the number of plausible polymorphs predicted in the final landscape was significantly reduced by the application of WTMD. This study is the first time that the propensity of a compound to crystallise was successfully predicted with an entirely *in silico* approach.

In the sixth blind test, only the group of Tuckerman & coworkers carried out MD calculations [11]. They used a Symmetry adapted perturbation theory (SAPT) DFT-generated force field to obtain a free energy ranking for 30 structures for molecule XXII using crystal-Adiabatic Free Energy Dynamics [58]. They found the experimental structure as the fourth lowest energy structure among the set of thirty. The Einstein Crystal method [76, 77] is another promising approach for free energy evaluation via thermodynamic integration [57]. In this method, the ideal Einstein Crystal, in which atoms oscillate around their equilibrium positions via harmonic springs, is used as a reference state with an analytical free energy expression. Free energy is then evaluated along a thermodynamic pathway (cycle) that connects this reference state to the solid of interest via intermediate states. The extended Einstein Crystal method has been used to evaluate the solubility of both organic and inorganic crystals [78, 79], however to our knowledge this method has not yet been applied to CSP studies. The Einstein crystal has been used as the reference state in the workflow of XtalPi [25] with free energy computed using multistage Bennett Acceptance Ratio (MBAR) [80]. In a recent study [25], the polymorphic molecule XXIII, from the 6^{th}blind test [11], was revisited using a CSP approach that includes the evaluation of free energy contributions. The CSP was carried out using Xtalpi's cloud computing protocol [24]. It was found that the tailor-made force field derived from B3LYP-D* accurately captured the observed experimental stability.

2.5 Disorder in molecular crystals

In addition to the effects of thermal and vibrational motion on free energy, disorder plays a significant role in the stability of experimental crystals. Approximately a quarter of all structures deposited in the CSD [81] exhibit disorder. In the case of such disordered structures, the role of configurational entropy contributions to free energy becomes crucial in determining crystalline

stability. The significance is attributed to the exponential rise in the number of available microstates within the ensemble of configurations. The distinct static structures generated in CSP studies might belong to the same ensemble of configurations, often resulting in the identification of metastable structures as stable forms. The main types of disorder that have been examined in CSP studies distinguish between static and dynamic disorder. The former refers to cases where different conformations are adopted by different molecules in the crystalline environment. Dynamic disorder is related to the motion of the molecules within the crystalline lattice, often arising from vibrations, rotations or diffusion. From a computational perspective, the modelling of disorder remains challenging, even with experimental guidance. As a result, there have been few CSP studies that have assessed the effect of disorder on the ranking of predicted structures. In a combined experimental and CSP study, both dynamic and static disorder were investigated for dimethylsulfoxide (DMSO):Carbamazepine solvates [82] using static models of disorder and molecular dynamics simulations. It was found that in this static system disorder is dominant at lower temperatures whereas dynamic disorder becomes more dominant above room temperature.

A promising technique for modelling substitutional and orientational static disorder is the symmetry adapted ensemble technique (SAET) [83]. The importance of configurational free energy in disordered systems has been highlighted through studies of several organic molecular crystals, such as caffeine [84], eniluracil & dichloro/dibromobenzene [85]. In CSP studies, SAET has also been applied alongside experiments for loratadine [86] and gandotinib [87]. For loratadine [86], a CSP study using the GRACE [27] software resulted in the prediction of ordered Form II as the global lattice energy minimum, in contrast with experimental measurements that indicated Form I to be the most stable structure. Consideration of isolated-site disorder analysis and SAET led to the prediction of disordered Form I as the energetically most favourable polymorph. The appearance of the disordered Form II was attributed to non-equilibrium effects (frozen-in crystallisation). For gandotinib [87], the experimental study revealed a polymorph-rich landscape. The CSP study on neat polymorphs predicted an undiscovered polymorph as the lowest lattice energy structure, while Form I matched a highly populated family of similar structures and Form II was not found in the CSP search. Consideration of vibrational free energies, at the PBE(0)-MBD+F_{vib} level, and SAET led to the conclusion that the structure predicted as being the most stable is not considerably more stable than Form I. The case of gandotinib highlights the importance of considering configurational free energy in disordered systems, and illustrates some of the continuing computational and experimental challenges in the exploration of solid form landscapes.

2.6 Overview & Summary

So far we have introduced the general multistage CSP methodology and reviewed existing free energy methods that have been applied in CSP studies. This section aims to address the limitations and approximations of the utilised free energy methods to reveal the gaps in literature and provide motivation for the work shown in this thesis. As discussed in Sections 2.3.3 and 2.4,

there is only a handful of cases where free energies have been utilised to rank non-disordered structures generated in CSP studies. This limited applicability of free energy calculation methods is attributed to the complexity and computational cost associated with evaluating entropic contributions and capturing the effect of thermal and vibrational motion on free energy. A summary of the free energy methods employed in CSP is given in Table 2.1. The free energy calculations methods applied in CSP can be can decomposed into two categories. Those methods that compute vibrational free energies using the approximation of lattice dynamics theory (HA, QHA or AA) and those methods evaluating free energy employing molecular simulations (Meta-Dynamics and MBAR). In the remainder of this section, we focus on lattice dynamics methods have been used in CSP studies to highlight gaps in the literature and provide motivation for the work presented in this thesis.

Table 2.1: Summary of studies that include an evaluation of the free energies of CSP-predicted structures as part of the ranking procedure. Abbreviations not defined in the text: 6BT: Target molecules XXII - XXXVI of the 6$^{\text{th}}$ blind test, DMA: distributed multipoles analysis, PC: point charges, MetadD: Metadynamics.

Authors	Molecule	Method	Energy model	Flexibility	Ref
Price group	6BT (except XXIX)	HA	DMA	No	[60]
Day group	XXII, XXV	HA	DMA	No	[60]
Nyman	chloridazon	HA	DMA	No	[50]
Vasileiadis	imidazole & TCNE	HA	DMA	Yes	[14]
Tkatchenko group	6BT	HA	PBE-TS	Yes	[60]
Hoja *et al.*	6BT	HA	PBE-TS	Yes	[47]
Hoja *et al.*	XXII	QHA & AA	PBE-TS	Yes	[47]
Karamertzanis *et al.*	5-fluorouracil	MetaD	PC	Yes	[73]
Zykova *et al.*	PR179	MetaD	PC	Yes	[74]
Zhang *et al.*	XXIII	MBAR	PC	Yes	[24]

Vibrational free energies have been computed using DFT-d for the target molecules XXII to XXVI of the 6$^{\text{th}}$ blind test [11] in the work of Hoja *et al.* [47] with candidate crystal structures generated with GRACE [27]. The same procedure has been employed by Tkatchenko's group in the 6$^{\text{th}}$ blind test [11], with candidate crystal structures provided by the group of Price and co-workers. The geometry optimisation was performed at the PBE-TS level of theory and calculated energies were supplemented with single point calculations using MBD corrections and the PBE0 hybrid functional. Vibrational free energy contributions were calculated at the PBE-TS level of theory. The computational cost of performing the free energy calculations ranged from 180 CPU hours to 45,000 CPU hours per crystal structure depending on the unit cell size, shape and symmetry. Overall, the cost associated with free energy evaluations at the PBE-TS level of theory was reported to be two to three time larger than that of performing lattice energy

single point calculations at the PBE0 + MBD level of theory. For all the blind test molecules, the use of calculated free energies resulted in experimental structures having the lowest energy among all structures.

Alternatively, vibrational free energies can be calculated using force fields that provide efficiency, allowing the examination of a larger number of structures. The model of intermolecular interactions using DMA [31] has been successfully applied to a wide range of crystal structures e.g. by Price and Day groups in the 6^{th} blind test [11] as well as in the PhD thesis of Nyman [50]. In these mentioned studies, free energies have been computed using DMACRYS adopting a rigid-body approach that neglects the impact of atomic vibrations in calculated free energies. In some cases, the effect of flexibility on the calculated lattice energies (instead of free energies) was taken into account by performing lattice energy minimisations with CrystalOptimizer [18]. A lattice dynamics method that employs DMA and in principle accounts for molecular flexibility has been developed by Vasileiadis [14]. Although this method employs an atomistic treatment, it was applied to refine the CSP landscapes of two rigid molecules imidazole and tetracyanoethylene (TCNE).

In the work presented in this thesis, we extent the applicability of the method developed by Vasileiadis to flexible molecules. We examine the effect of using a lattice energy minimisation algorithm that accounts for molecular flexibility. Our objective is to examine whether the evaluation of free energies can be incorporated into our CSP multistage approach. In the next chapter, we derive the lattice dynamics equations for performing with an energy model that used distributed multipoles.

2.7 Conclusions

The stability of molecular crystals at a given T and P is governed by Gibbs free energy. The goal of CSP is to identify the most stable form of a compound at a given temperature and pressure and other forms in order of descending stability. The state-of-the art CSP methods employ multistage approaches, which involve a candidate generation and multiple refinement stages. Despite the associated computational cost, lattice energy calculations employing DFT-d are expected to yield higher accuracy compared to force field methods. Force fields are usually employed in the candidate generation stages as their efficiency compared to DFT-d methods allows for a broader exploration of the crystal landscape. The development of customised force fields, tailored to a specific molecule and level of theory, facilitates the examination of a larger number of structures while maintaining a high level of accuracy. Two of the most successful tailor-made force fields use either multipoles moments fitted to isolated-molecule QM calculations via DMA [39] or point charges fitted to periodic DFT-d [24, 25].

There are very few studies in the literature in which free energies have been employed to rank predicted structures in CSP generated landscapes. This is attributed to the high computational cost of capturing entropic contributions for a large number of structures. The existing methods include atomistic DFT-d calculations or rigid-body LD with anisotropic force fields. The former

method can provide very accurate results, but is computationally expensive. The anisotropic force fields can provide the necessary efficiency, but the rigid treatment neglects contributions arising from intramolecular vibrations. The method adopted in this thesis offers efficiency while incorporating intramolecular contributions but has so far only been applied to evaluate the free energy of rigid molecules. Our objective is to extend the applicability of the method to molecules that exhibit molecular flexibility.

Chapter 3

Lattice dynamics for a tailor-made CSP force field

In this chapter, we initially present a mathematical description of crystal geometry (Section 3.1) and derive the necessary expressions for performing LD calculations within the HA for the general case of a potential expressed as a function of the atomic positions in the crystal (Section 3.2). Next in Section 3.3, we define vibrational contributions to the free energy and provide analytical expressions for entropy, isochoric heat capacity and thermal energy. We proceed by describing the specific tailor-made force utilised in the lattice energy and free energy refinement stages in our work (Section 3.4). The lattice dynamics equations using our tailor-made force field are then derived from the equations for the general potential. Further details of the derivations presented in this chapter can be found in the thesis of Vasileiadis [14]. Here, we summarise the key points of the LD derivations, provide additional details of the derivations of thermodynamic properties (entropy, isochoric heat capacity and thermal energy). We also correct some typographical errors in [14] related to the derivation of the electrostatic component of the dynamical matrix (see Section 3.5) and the unit vector connecting two interacting atoms.

3.1 Mathematical description of molecular crystals

A crystal structure can be described by means of a lattice and a basis, where a lattice refers to a regular and periodic arrangement of points in space and basis to a group of atoms or molecules that are assigned to the lattice. The lattice and basis can be represented in a unit cell, which describes a repeating unit of the crystal without considering any symmetry operations that, when translated along three dimensions, reproduces the entire crystal structure. As an example, the unit cells of the YN and OP polymorphs of ROY are shown in Figure 1.1. Without any loss of generality, the shape of the unit cell geometry is given by a parallelepiped defined by three lattice lengths (a, b, c) and three lattice angles (α, β, γ). An example of a unit cell is shown in Figure 3.1.

The unit cell matrix, $\boldsymbol{R}$, is determined by the three lattice vectors, $\boldsymbol{a}, \boldsymbol{b}$ and $\boldsymbol{c}$ and given by:

$$\boldsymbol{R} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \begin{pmatrix} a & b\cos\gamma & c\cos\beta \\ 0 & b\sin\gamma & c\frac{\cos\alpha - \cos\beta\cos\gamma}{\sin\gamma} \\ 0 & 0 & \frac{c\sqrt{w}}{\sin\gamma} \end{pmatrix}, \tag{3.1.1}$$

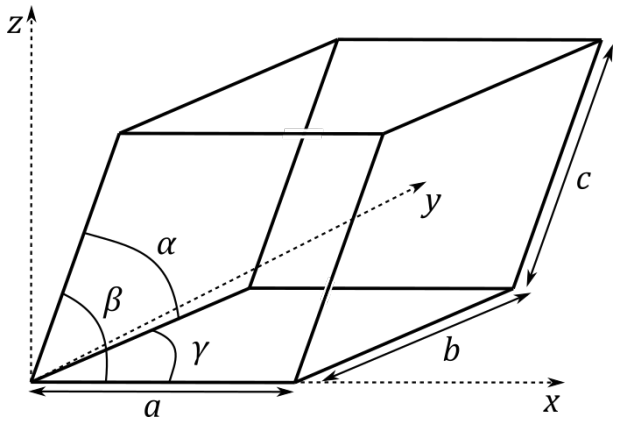


Figure 3.1: A typical diagram of a unit cell

where w is the lattice deformation variable [20, 40] given by:

$$w = 1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma. \tag{3.1.2}$$

The upper triangular form of the unit cell matrix implies that lattice vector $\boldsymbol{a}$ is parallel to the x-axis, lattice vector $\boldsymbol{b}$ lies on the xy-plane, and lattice vector $\boldsymbol{c}$ forms a right-handed coordinate system with respect to the other two vectors. The unit cell volume V can be calculated as $V = \boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})$.

For molecular crystals, the description of atomic positions requires to introduce three different indices, each of which is used to label unit cells, molecules and atoms. Given a unit cell containing Z molecules, the capital case indices $I, I = 1, 2, ..., Z$, and $J, J = 1, 2, ..., Z$, are used to enumerate the molecules in that unit cell. The total number of atoms inside a molecule is denoted by N and lower case indices, $i, i = 1, 2, ..., N$, and $j, j = 1, 2, ..., N$, are used to enumerate atoms in a molecule. A set $\boldsymbol{L}$ of three-dimensional vectors, $\boldsymbol{l}$, with $\boldsymbol{L} = \{\boldsymbol{l}\}$ and $\boldsymbol{l} \in \mathbb{N}^3$, is used to describe the position of an arbitrary unit cell relative to the reference unit cell $\boldsymbol{l}^T = (\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0})$ via translations defined by the elements of $\boldsymbol{l}$. The infinite set $\boldsymbol{r}$ consists of vectors holding the coordinates of each atom in the crystal and is defined as

$$\boldsymbol{r} = \{\boldsymbol{r}_{lIi}, \quad \boldsymbol{l} \in \boldsymbol{L}, \quad i = 1, 2, \ldots N, \quad I = 1, 2, \ldots, Z\}, \tag{3.1.3}$$

where $\boldsymbol{r}_{lIi}$ is the vector describing the atomic coordinates of atom i in molecule I and unit cell $\boldsymbol{l}$. Due to the periodic structure of the crystal, it is convenient to evaluate the coordinates of an arbitrary atom i in molecule I in any periodic unit cell $\boldsymbol{l}$ with respect to the atomic position of the same atom in the reference unit cell

$$\boldsymbol{r}_{lIi} = \boldsymbol{r}_{\boldsymbol{0}Ii} + \boldsymbol{l}\boldsymbol{R}, \tag{3.1.4}$$

where $\boldsymbol{r}_{\boldsymbol{0}Ii}$ is the vector describing the position of atom i in molecule I in the reference unit cell.

3.2 Harmonic approximation

In this section, we will derive the harmonic approximation equations for the case of a general potential, $U(\boldsymbol{r})$, as a function of the positions of atoms within the crystal $\boldsymbol{r}$. The crystal lattice

energy can be approximated using a Taylor series expansion around an arbitrary reference point, $\tilde{\boldsymbol{r}}$, as:

$$\begin{aligned} U(\boldsymbol{r}) \approx U(\tilde{\boldsymbol{r}}) &+ \sum_{\boldsymbol{l}\in\boldsymbol{L}}\sum_{I=1}^{Z}\sum_{i=1}^{N}(\boldsymbol{r}_{\boldsymbol{l}Ii}-\tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii})^T \left.\frac{\partial U}{\partial \boldsymbol{r}_{\boldsymbol{l}Ii}}\right|_{\tilde{\boldsymbol{r}}} \\ &+ \frac{1}{2}\sum_{\boldsymbol{l}\in\boldsymbol{L}}\sum_{I=1}^{Z}\sum_{i=1}^{N}\sum_{\boldsymbol{l}'\in\boldsymbol{L}}\sum_{J=1}^{Z}\sum_{j=1}^{N}(\boldsymbol{r}_{\boldsymbol{l}Ii}-\tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii})^T \left.\frac{\partial^2 U}{\partial \boldsymbol{r}_{\boldsymbol{l}Ii}\partial \boldsymbol{r}_{\boldsymbol{l}'Jj}}\right|_{\tilde{\boldsymbol{r}}} (\boldsymbol{r}_{\boldsymbol{l}'Jj}-\tilde{\boldsymbol{r}}_{\boldsymbol{l}'Jj}) + ..., \end{aligned} \tag{3.2.1}$$

where the dots (...) in the last term of the right hand side of Equation (3.2.1) denote higher-order terms of the expansion series (beyond quadratic terms).

In the harmonic approximation, the expansion point $\tilde{\boldsymbol{r}}$ is chosen to coincide with a local minimum (equilibrium point) in the lattice energy landscape. The harmonic potential, U^{har}, is thus approximated by a quadratic function using a second-order Taylor series expansion around this equilibrium position. At those equilibrium position points, the force acting on each atom is equal to zero and as a result the first-order term in the Taylor series expansion vanishes. Furthermore, the expansion is truncated at the quadratic term by neglecting higher-order terms in the series. U^{har} provides an approximation of the true potential energy near the equilibrium position and is given by the following quadratic equation:

$$U^{\mathrm{har}}(\boldsymbol{r}) = U(\tilde{\boldsymbol{r}}) + \frac{1}{2}\sum_{\boldsymbol{l}\in\boldsymbol{L}}\sum_{I=1}^{Z}\sum_{i=1}^{N}\sum_{\boldsymbol{l}'\in\boldsymbol{L}}\sum_{J=1}^{Z}\sum_{j=1}^{N}(\boldsymbol{r}_{\boldsymbol{l}Ii}-\tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii})^T \left.\frac{\partial^2 U}{\partial \boldsymbol{r}_{\boldsymbol{l}Ii}\partial \boldsymbol{r}_{\boldsymbol{l}'Jj}}\right|_{\tilde{\boldsymbol{r}}} (\boldsymbol{r}_{\boldsymbol{l}'Jj}-\tilde{\boldsymbol{r}}_{\boldsymbol{l}'Jj}). \tag{3.2.2}$$

Using the harmonic potential, the forces exerted on each atom, $\boldsymbol{F}_{\boldsymbol{l}Ii}$, in the crystal can be evaluated from the first derivative of the harmonic potential with respect to the atomic position vectors

$$\boldsymbol{F}_{\boldsymbol{l}Ii} = -\left.\frac{\partial U^{\mathrm{har}}(\boldsymbol{r})}{\partial \boldsymbol{r}_{\boldsymbol{l}Ii}}\right|_{\tilde{\boldsymbol{r}}} = -\sum_{\boldsymbol{l}'\in\boldsymbol{L}}\sum_{J=1}^{Z}\sum_{j=1}^{N}\left.\frac{\partial^2 U}{\partial \boldsymbol{r}_{\boldsymbol{l}Ii}\partial \boldsymbol{r}_{\boldsymbol{l}'Jj}}\right|_{\tilde{\boldsymbol{r}}} (\boldsymbol{r}_{\boldsymbol{l}'Jj}-\tilde{\boldsymbol{r}}_{\boldsymbol{l}'Jj}),\ \forall \boldsymbol{l}Ii, \tag{3.2.3}$$

where the symbol $\forall$ implies that the indices take all their permissible values. The equations of motion, derived from Newton's second law, are used to obtain the dynamical evolution of the crystal:

$$\boldsymbol{F}_{\boldsymbol{l}Ii} = m_{Ii}\frac{\partial^2(\boldsymbol{r}_{\boldsymbol{l}Ii}-\tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii})}{\partial t^2},\ \forall \boldsymbol{l}Ii, \tag{3.2.4}$$

where m_{Ii} is the mass of atom Ii, which does not depend on the unit cell index as the periodic images of an atom have the same mass. By combining Equations (3.2.3) and (3.2.4), the equations of motion become:

$$-\sum_{\boldsymbol{l}'\in\boldsymbol{L}}\sum_{J=1}^{Z}\sum_{j=1}^{N}\left.\frac{\partial^2 U}{\partial \boldsymbol{r}_{\boldsymbol{l}Ii}\partial \boldsymbol{r}_{\boldsymbol{l}'Jj}}\right|_{\tilde{\boldsymbol{r}}} (\boldsymbol{r}_{\boldsymbol{l}'Jj}-\tilde{\boldsymbol{r}}_{\boldsymbol{l}'Jj}) = m_{Ii}\frac{\partial^2(\boldsymbol{r}_{\boldsymbol{l}Ii}-\tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii})}{\partial t^2},\ \forall \boldsymbol{l}Ii, \tag{3.2.5}$$

At this point, it is convenient to introduce the displacement vector, $\boldsymbol{d}_{\boldsymbol{l}Ii} = \boldsymbol{r}_{\boldsymbol{l}Ii} - \tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii}$, defined as the difference between the position of atom $\boldsymbol{l}Ii$ and its equilibrium position. Note that the partial derivatives of the displacement vector and atomic position vector can be used interchangeably

as $\partial \boldsymbol{d}_{\boldsymbol{l}Ii} = \partial(\boldsymbol{r}_{\boldsymbol{l}Ii} - \tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii}) = \partial \boldsymbol{r}_{\boldsymbol{l}Ii}$. Substituting $\boldsymbol{d}_{\boldsymbol{l}Ii}$ in Equation (3.2.5) yields:

$$-\sum_{\boldsymbol{l}' \in \boldsymbol{L}} \sum_{J=1}^{Z} \sum_{j=1}^{N} \frac{\partial^2 U}{\partial \boldsymbol{d}_{\boldsymbol{l}Ii} \partial \boldsymbol{d}_{\boldsymbol{l}'Jj}}\bigg|_{\tilde{\boldsymbol{r}}} \boldsymbol{d}_{\boldsymbol{l}'Jj} = m_{Ii} \frac{\partial^2 \boldsymbol{d}_{\boldsymbol{l}Ii}}{\partial t^2}, \ \forall \boldsymbol{l}Ii, \tag{3.2.6}$$

The solution of the system of second order differential equations stated as Equation (3.2.6) has the general form of a harmonic wave [15]:

$$\boldsymbol{d}_{\boldsymbol{l}Ii} = \tilde{\boldsymbol{d}}_{Ii}(\boldsymbol{k}) e^{i(\boldsymbol{k} \cdot \tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii} - \omega(\boldsymbol{k})t)}, \tag{3.2.7}$$

where $\tilde{\boldsymbol{d}}_{Ii}(\boldsymbol{k})$ is the amplitude of the harmonic wave, which is independent of $\boldsymbol{l}$ as the difference in the motion of atoms in different unit cells is described by the exponential factor. $\omega(\boldsymbol{k})$ is the angular frequency, which depends on the reciprocal space (wave) vector, $\boldsymbol{k}$, taking values in the reciprocal space defined as

$$\boldsymbol{k} = \xi_a \boldsymbol{a}^* + \xi_b \boldsymbol{b}^* + \xi_c \boldsymbol{c}^*, \quad \xi_a, \xi_b, \xi_c \in \mathbb{R}, \tag{3.2.8}$$

where ξ_a, ξ_b and ξ_c are the reciprocal space coordinates and $\boldsymbol{a}^*, \boldsymbol{b}^*$ and $\boldsymbol{c}^*$ are the reciprocal lattice vectors given by:

$$\boldsymbol{a}^* = 2\pi \frac{\boldsymbol{b} \times \boldsymbol{c}}{\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})}, \quad \boldsymbol{b}^* = 2\pi \frac{\boldsymbol{c} \times \boldsymbol{a}}{\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})}, \quad \boldsymbol{c}^* = 2\pi \frac{\boldsymbol{a} \times \boldsymbol{b}}{\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})}, \tag{3.2.9}$$

The unknown parameters of the harmonic wave equation, $\tilde{\boldsymbol{d}}_{Ii}$ and $\omega(\boldsymbol{k})$, are determined by substituting the harmonic wave Equation (3.2.7) into Newton's equations of motion (3.2.6). The resulting expression reads:

$$-\sum_{\boldsymbol{l}' \in \boldsymbol{L}} \sum_{J=1}^{Z} \sum_{j=1}^{N} \frac{\partial^2 U}{\partial \boldsymbol{d}_{\boldsymbol{l}Ii} \partial \boldsymbol{d}_{\boldsymbol{l}'Jj}}\bigg|_{\tilde{\boldsymbol{r}}} \tilde{\boldsymbol{d}}_{Jj}(\boldsymbol{k}) e^{i(\boldsymbol{k} \cdot \tilde{\boldsymbol{r}}_{\boldsymbol{l}'Jj} - \omega(\boldsymbol{k})t)} = m_{Ii} \tilde{\boldsymbol{d}}_{Ii}(\boldsymbol{k}) e^{i(\boldsymbol{k} \cdot \tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii} - \omega(\boldsymbol{k})t)} \omega^2(\boldsymbol{k}), \ \forall \boldsymbol{l}Ii. \tag{3.2.10}$$

After some manipulation involving dividing with $\sqrt{m_{Ii}}$, introducing $\frac{\sqrt{m_{Jj}}}{\sqrt{m_{Jj}}}$, taking into account that the terms $e^{-\omega(\boldsymbol{k})t}$ cancel out and rearranging the exponential term, Equation (3.2.10) can be written in the following equivalent form:

$$-\sum_{\boldsymbol{l}' \in \boldsymbol{L}} \sum_{J=1}^{Z} \sum_{j=1}^{N} \frac{\partial^2 U}{\partial \boldsymbol{d}_{\boldsymbol{l}Ii} \partial \boldsymbol{d}_{\boldsymbol{l}'Jj}}\bigg|_{\tilde{\boldsymbol{r}}} \frac{1}{\sqrt{m_{Ii} m_{Jj}}} \sqrt{m_{Jj}} \tilde{\boldsymbol{d}}_{Jj}(\boldsymbol{k}) e^{i\boldsymbol{k} \cdot (\tilde{\boldsymbol{r}}_{\boldsymbol{l}'Jj} - \tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii})} = \sqrt{m_{Ii}} \tilde{\boldsymbol{d}}_{Ii}(\boldsymbol{k}) \omega^2(\boldsymbol{k}), \ \forall \boldsymbol{l}Ii. \tag{3.2.11}$$

Next, defining a set of polarisation vectors as $\boldsymbol{e} = (\boldsymbol{e}_{11}, \boldsymbol{e}_{12}, ..., \boldsymbol{e}_{1N}, \boldsymbol{e}_{21}, ..., \boldsymbol{e}_{ZN})^T$, where $\boldsymbol{e}_{Ii} = \sqrt{m_{Ii}} \tilde{\boldsymbol{d}}_{Ii}$, yields

$$-\sum_{\boldsymbol{l}' \in \boldsymbol{L}} \sum_{J=1}^{Z} \sum_{j=1}^{N} \frac{\partial^2 U}{\partial \boldsymbol{d}_{\boldsymbol{l}Ii} \partial \boldsymbol{d}_{\boldsymbol{l}'Jj}}\bigg|_{\tilde{\boldsymbol{r}}} \frac{1}{\sqrt{m_{Ii} m_{Jj}}} \boldsymbol{e}_{Jj}(\boldsymbol{k}) e^{i\boldsymbol{k} \cdot (\tilde{\boldsymbol{r}}_{\boldsymbol{l}'Jj} - \tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii})} = \boldsymbol{e}_{Ii}(\boldsymbol{k}) \omega^2(\boldsymbol{k}), \ \forall \boldsymbol{l}Ii. \tag{3.2.12}$$

The next and final step in the derivation involves introducing the dynamical matrix, $\boldsymbol{D}(\boldsymbol{k})$ as the mass weighted Hessian matrix of the second-order derivatives of the lattice energy

with respect to atomic positions (or displacements) multiplied by the exponential phase factor, $e^{i\boldsymbol{k}\cdot(\tilde{\boldsymbol{r}}_{\boldsymbol{l}'Jj}-\tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii})}$. The dynamical matrix is a $3ZN \times 3ZN$ matrix. The dynamical matrix elements, $D_{iI,Jj}(\boldsymbol{k})$, are given by

$$D_{iI,Jj}(\boldsymbol{k}) = -\frac{1}{\sqrt{m_{Ii}m_{Jj}}} \sum_{\boldsymbol{l}' \in \boldsymbol{L}} \frac{\partial^2 U}{\partial \boldsymbol{d}_{\boldsymbol{l}Ii} \partial \boldsymbol{d}_{\boldsymbol{l}'Jj}}\bigg|_{\tilde{\boldsymbol{d}}} e^{i\boldsymbol{k}\cdot(\tilde{\boldsymbol{r}}_{\boldsymbol{l}'Jj}-\tilde{\boldsymbol{r}}_{\boldsymbol{l}Ii})}. \tag{3.2.13}$$

As mentioned above, the differentiation of the lattice energy with respect to atomic positions is equivalent with the differentiation with respect to atomic displacements. In addition, the reference position vector $\tilde{\boldsymbol{r}}$ is replaced with the symbol $\boldsymbol{r}$ to simplify the notation. Furthermore, due to the invariance of the exponential phase factor to the unit cell, we can focus on the reference unit cell without any loss of generality. As a result, the exponential phase factor becomes $e^{i\boldsymbol{k}\cdot(\boldsymbol{r}_{\boldsymbol{l}Jj}-\boldsymbol{r}_{\boldsymbol{0}Ii})}$ and the dynamical matrix expression can be rewritten into the following equivalently expression:

$$D_{iI,Jj}(\boldsymbol{k}) = \frac{1}{\sqrt{m_{Ii}m_{Jj}}} \sum_{\boldsymbol{l} \in \boldsymbol{L}} \frac{\partial^2 U}{\partial \boldsymbol{r}_{\boldsymbol{0}Ii} \partial \boldsymbol{r}_{\boldsymbol{l}Jj}}\bigg|_{\boldsymbol{r}} e^{i\boldsymbol{k}\cdot(\boldsymbol{r}_{\boldsymbol{l}Jj}-\boldsymbol{r}_{\boldsymbol{0}Ii})}. \tag{3.2.14}$$

The system of equations (3.2.10) can be re-written in matrix notation as

$$\boldsymbol{D}(\boldsymbol{k})\boldsymbol{e} = \omega^2(\boldsymbol{k})\boldsymbol{e}. \tag{3.2.15}$$

Equation (3.2.15) describes an eigenvalue problem, wherein a matrix multiplied by a vector is equal to the same vector multiplied by a set of scalar quantities. The eigenvalues, λ_n, of the dynamical matrix are computed via diagonalisation. Since the dynamical matrix is Hermitian ($\boldsymbol{D}^T = \boldsymbol{D}^*$), it has real eigenvalues. The eigenvalues of the dynamical matrix are related to the angular, ω_n, and radial, ν_n, phonon frequencies via the following relation

$$\omega_n(\boldsymbol{k}) = \sqrt{\lambda_n} = 2\pi\nu_n(\boldsymbol{k}) \quad n = 1, 2, ..., 3ZN. \tag{3.2.16}$$

The signs of the phonon frequencies have a well-defined physical interpretation. Positive eigenvalues correspond to stable structures, whereas negative eigenvalues yield imaginary frequencies indicating that the structure in unstable.

3.3 Thermodynamic functions

In this section, we derive the harmonic vibrational free energy expression using the Bose-Einstein distribution of bosons [15]. Additionally, we provide analytical expressions for entropy, heat capacity and thermal energy.

3.3.1 Vibrational free energy

As shown in the Section 3.2, the harmonic potential is a quadratic function of the atomic displacements from their equilibrium positions. This quadratic (harmonic) function means that

atoms lie on potential wells of parabolic shape. The vibrational energy of the harmonic oscillator, $U^{\text{vib}}(\eta_n^{\boldsymbol{k}})$, is derived by adopting a quantum mechanical description.

$$U^{\text{vib}} = \sum_{\boldsymbol{k}} \sum_{n=1}^{3N} \left(\frac{1}{2} + \eta_n^{\boldsymbol{k}}\right) \hbar\omega_n(\boldsymbol{k}), \quad \eta_n^{\boldsymbol{k}} \in \mathbb{N}_0, \tag{3.3.1}$$

where $\hbar = \frac{h}{2\pi}$, with h being the Planck constant. In the HA, the oscillators are not coupled, meaning that the motion in each direction is independent of the others. As can be seen in Equation (3.3.1), the energy states are separated by $\hbar\omega_n(\boldsymbol{k})$. For $\eta_n^{\boldsymbol{k}} = 0$, the lowest energy state (ground state) is occupied, yielding a non-zero contribution to vibrational energy. This contribution arises from the quantum mechanical treatment of the harmonic oscillator and is referred to as zero point energy (ZPE), U^{ZPE}. The free energy expression can be derived by constructing the partition function, Q_{NVT}, in the canonical ensemble (NVT) and evaluating the free energy using $A = -k_{\text{B}}T \ln Q_{NVT}$, where k_{B} is the Boltzmann constant [14]. Alternatively, making use of the fact that phonons are bosons, the free energy expression can be derived using the Bose-Einstein distribution [15]. The expression describing the vibrational contributions to the free energy, $A^{\text{vib}}(T)$, is given by:

$$A^{\text{vib}}(T) = \sum_{\boldsymbol{k}} \sum_{n=1}^{3N} \frac{1}{2}\hbar\omega_n(\boldsymbol{k}) + \sum_{\boldsymbol{k}} \sum_{n=1}^{3N} k_{\text{B}}T \ln\left[1 - \exp\left(\frac{-\hbar\omega_n(\boldsymbol{k})}{k_{\text{B}}T}\right)\right], \tag{3.3.2}$$

where the first term on the right hand side is the ZPE and we will refer to the second term as the thermal contributions to the free energy (not to be confused with thermal energy). The Helmholtz free energy at a given temperature, $A(T)$, is computed as the sum of the static lattice energy and vibrational component:

$$A(T) = U^{\text{latt}} + A^{\text{vib}}(T). \tag{3.3.3}$$

Since the reciprocal space is continuous, the sum over $\boldsymbol{k}$, which appears in Equation (3.3.2) can be expressed as a triple integral along the reciprocal space coordinates ξ_a, ξ_b and ξ_c.

$$A(T) = U^{\text{latt}} + \int_0^1 \int_0^1 \int_0^1 \sum_{n=1}^{3N} \left\{\frac{1}{2}\hbar\omega_n(\boldsymbol{k}) + k_{\text{B}}T \ln\left[1 - \exp\left(\frac{-\hbar\omega_n(\boldsymbol{k})}{k_{\text{B}}T}\right)\right]\right\} d\xi^a d\xi^b d\xi^c \tag{3.3.4}$$

In computational studies, however, the evaluation of $\boldsymbol{k}$ points is performed in a discrete manner. The triple integral of Equation (3.3.4) is calculated using the Gauss-Legendre quadrature (GLQ) numerical integration technique. Details of the GLQ can be found in the PhD thesis of Vasileiadis [14].

The Gibbs free energy, G, can been obtained via the partition function in the isobaric-isothermal ensemble (NPT) and is given by the equation below

$$G(T,P) = U^{\text{latt}} + PV + \int_0^1 \int_0^1 \int_0^1 \sum_{n=1}^{3N} \left\{\frac{1}{2}\hbar\omega_n(\boldsymbol{k}) + k_{\text{B}}T \ln\left[1 - \exp\left(\frac{-\hbar\omega_n(\boldsymbol{k})}{k_{\text{B}}T}\right)\right]\right\} d\xi^a d\xi^b d\xi^c \tag{3.3.5}$$

Note the consistency between Equation (3.3.5) and Equation (2.3.1). In Equation (3.3.5), the vibrational contributions to free energy (A_{vib}), which are namely, ZPE, thermal energy and

entropy, are expressed as a function of phonon frequencies, $\omega_n(\boldsymbol{k})$. As the temperature rises, the unit cell's volume tends to expand due to thermal effects. This expansion significantly influences the position of the lattice energy minimum. In the harmonic approximation, the volume of the unit cell is assumed to be kept constant, neglecting the effects of thermal expansion at temperatures above 0 K. The QHA (see Section 2.3.1) can be utilised to tackle the limitations arising from the harmonic treatment and provide more accurate values for the Gibbs free energy. Throughout this thesis, the Helmholtz free energy is used as the metric for determining the relative stability of crystals at temperatures above 0 K. The Helmholtz free energy is also referred to as vibrational free energy.

3.3.2 Entropy, heat capacity and thermal energy

The vibrational entropy is calculated as the negative of the first derivative of the Helmholtz free energy with respect to temperature

$$S = -\left.\frac{\partial A(T)}{\partial T}\right|_V. \tag{3.3.6}$$

The derivation involves the differentiation of the thermal contributions to free energy, which are temperature-dependent. The resulting expression for entropy reads:

$$S = -k_\mathrm{B}\sum_{\boldsymbol{k}}\sum_{n=1}^{3N}\ln\left[1-\exp\left(\frac{-\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T}\right)\right] + \frac{1}{T}\sum_{\boldsymbol{k}}\sum_{n=1}^{3N}\hbar\omega_n(\boldsymbol{k})\left[\frac{\exp\left(\frac{-\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T}\right)}{1-\exp\left(\frac{-\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T}\right)}\right] \tag{3.3.7}$$

As has already been discussed, within the harmonic approximation the volume of the unit cell is kept constant at its lattice energy minimum value. As a consequence, thermodynamic quantities can be calculated at constant volume. The isochoric heat capacity or heat capacity at constant volume, C_V, is given by

$$C_V(T) = T\left.\frac{\partial S}{\partial T}\right|_V. \tag{3.3.8}$$

Substituting Equation (3.3.7) yields

$$C_V(T) = k_\mathrm{B}\sum_{\boldsymbol{k}}\sum_{n=1}^{3N}\left(\frac{\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T}\right)^2\frac{\exp\left(\frac{\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T}\right)}{\left(1-\exp\left(\frac{\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T}\right)\right)^2}. \tag{3.3.9}$$

The thermal energy, $U^{\mathrm{thermal}}(T)$, can be calculated as:

$$\begin{aligned} U^{\mathrm{thermal}}(T) &= \int_0^T C_V\,dT' \\ &= \int_0^T k_\mathrm{B}\sum_{\boldsymbol{k}}\sum_{n=1}^{3N}\left(\frac{\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T'}\right)^2\frac{\exp\left(\frac{\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T'}\right)}{\left(1-\exp\left(\frac{\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T'}\right)\right)^2}dT'. \\ &= k_\mathrm{B}\sum_{\boldsymbol{k}}\sum_{n=1}^{3N}\int_0^T\left(\frac{\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T'}\right)^2\frac{\exp\left(\frac{\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T'}\right)}{\left(1-\exp\left(\frac{\hbar\omega_n(\boldsymbol{k})}{k_\mathrm{B}T'}\right)\right)^2}dT'. \end{aligned} \tag{3.3.10}$$

The integral shown in Equation (3.3.10) is computed based on the following formula:

$$\int \left(\frac{a}{x}\right)^2 \frac{e^{\frac{a}{x}}}{\left(1-e^{\frac{a}{x}}\right)^2} dx = \frac{a}{e^{\frac{a}{x}}-1} + c. \tag{3.3.11}$$

The final expression for thermal energy is obtained by setting $a = \frac{\hbar\omega_{\boldsymbol{k}}(\boldsymbol{k})}{k_{\mathrm{B}}}$ and $x = T$

$$U^{\mathrm{thermal}}(T) = \sum_{\boldsymbol{k}} \sum_{n=1}^{3N} \frac{\hbar\omega_n(\boldsymbol{k})}{\exp\left(\frac{\hbar\omega_n(\boldsymbol{k})}{k_{\mathrm{B}}T}\right)-1}. \tag{3.3.12}$$

The addition of Equation (3.3.12) and Equation (3.3.7) multiplied with T yields the thermal contributions to free energy shown in the last term of the integral in the right-hand side of Equation (3.3.5).

3.4 Energy model

So far, we have derived the harmonic approximation equations for molecular crystals based on a general potential, $U(\boldsymbol{r})$, as a function of interatomic distances, which themselves depend on the coordinates of each atom in the crystal. As discussed on Section 2.2, force fields are commonly employed in CSP studies as they offer increased efficiency compared to electronic structure methods. In this section, we provide a detailed description of the tailor-made CSP force field used in our lattice energy and free energy refinement stages.

In computational studies, force fields are energy models used to describe the interactions between atoms. In many force fields, the potential energy, U, is decomposed in contributions arising from different types of interactions

$$U = U^{\mathrm{intra}} + U^{\mathrm{inter}} + U^{\mathrm{other}} \tag{3.4.1}$$

where U^{intra}, U^{inter} and U^{other} refer to intramolecular, intermolecular and other interactions, respectively. U^{intra} captures the contributions arising from the interaction between atoms belonging to the same molecule. On the other hand, U^{inter} accounts for interactions between atoms belonging to different molecules and can be further decomposed into electrostatic, U^{ele}, and repulsion-dispersion interactions, $U^{\mathrm{rep/dis}}$. The term U^{other} consists of interactions that have not been accounted for by U^{intra} and U^{inter}. These interactions can include induction effects, and many-body contributions among others. We assume that U^{other} can be neglected and the force field expression is given by

$$\begin{aligned} U &= U^{\mathrm{intra}} + U^{\mathrm{inter}} \\ &= U^{\mathrm{intra}} + U^{\mathrm{ele}} + U^{\mathrm{rep/dis}}. \end{aligned} \tag{3.4.2}$$

For molecular crystals, the lattice energy, ΔU^{latt}, is defined relative to a reference state. The the gas phase energy, U^{gas}, of the isolated molecule is used as the reference state and ΔU^{latt} is given by.

$$\begin{aligned} \Delta U^{\mathrm{latt}} &= U^{\mathrm{intra}} - U^{\mathrm{gas}} + U^{\mathrm{ele}} + U^{\mathrm{rep/dis}} \\ &= \Delta U^{\mathrm{intra}} + U^{\mathrm{ele}} + U^{\mathrm{rep/dis}}, \end{aligned} \tag{3.4.3}$$

where ΔU^{intra} is the intramolecular contribution to lattice energy. As already mentioned, the parameters used in quantifying interactions in force fields can be either transferable or tailor-made.

3.4.1 Intramolecular interactions

In our force field, the intramolecular energy refers to the interactions between atoms belonging to the same molecule. The term gas phase or *in vaccuo* describes an isolated molecule that does not interact with any other molecules. The minimum energy conformation of a molecule *in vacuo* can be different from the conformation adopted in the crystal, as a result of the proximity of molecules arranged in the lattice. The molecule in its gas phase is used as the reference state, with ΔU^{intra} defined as the energy required to deform a molecule from its gas phase conformation to the conformation adopted in the crystalline environment ($\Delta U^{\text{intra}} = U^{\text{intra}} - U^{\text{gas}}$).

The intramolecular contribution to the lattice energy is calculated via isolated molecule geometry optimisations using the Gaussian software [88]. For rigid molecules, the molecular conformation in the gas phase is approximated to be identical to the one in the crystalline phase, yielding $\Delta U^{\text{intra}} = 0$. The geometry of rigid molecules is thus obtained from an unconstrained geometry optimisation. On the other hand, molecules that adopt different conformations *in vacuo* and in the crystal, as a result of packing forces, are known as flexible molecules. Their crystalline geometry is determined by constrained geometry optimisations. Selected DoFs, which are treated as flexible during the local and global optimisation, are allowed to vary during the geometry optimisation.

3.4.2 Pairwise additive potentials

A vector connecting two arbitrary atoms in the crystal describes the difference in positions between atoms i of molecule I in the the central unit cell ($\boldsymbol{l} = \mathbf{0}$) and atom j in molecule J unit cell $\boldsymbol{l}$. Here we introduce a shorthand of notation, with $\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}$ defined as

$$\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj} = \boldsymbol{r}_{\boldsymbol{l}Jj} - \boldsymbol{r}_{\mathbf{0}Ii} = \boldsymbol{r}_{\mathbf{0}Jj} + \boldsymbol{R}\cdot\boldsymbol{l} - \boldsymbol{r}_{\mathbf{0}Ii}. \tag{3.4.4}$$

We assume that intermolecular interactions are pairwise additive and depend on the difference in positions between atoms belonging to different molecules. The intermolecular energy is calculated as

$$U^{\text{inter}} = \frac{1}{2Z}\sum_{\boldsymbol{l}=-\infty}^{+\infty}\sum_{I=1}^{Z}\underbrace{\sum_{J=1}^{Z}}_{J\neq I \text{ if } \boldsymbol{l}=\mathbf{0}}\sum_{i=1}^{N_i}\sum_{j=1}^{N_j} U^{\text{rep/dis}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}(r_{\mathbf{0}Ii,\boldsymbol{l}Jj}) + U^{\text{ele}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}(\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}), \tag{3.4.5}$$

where $r_{\mathbf{0}Ii,\boldsymbol{l}Jj} = ||\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}||$ is the magnitude of the vector connecting the pair of atoms (the distance between them), the factor $\frac{1}{2}$ avoids the double-counting of interactions, and the intermolecular energy is normalised per molecule in the unit cell (Z). The unit vector connecting two interacting sites, $\hat{\boldsymbol{r}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}$, is given by

$$\hat{\boldsymbol{r}}_{\mathbf{0}Ii,\boldsymbol{l}Jj} = \frac{\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{r_{\mathbf{0}Ii,\boldsymbol{l}Jj}}. \tag{3.4.6}$$

The derivative of $r_{\mathbf{0}Ii,\boldsymbol{l}Jj}$ with respect to $\boldsymbol{r}_{\mathbf{0}Ii}$ is related to the unit vector and can be calculated by application of the chain rule as:

$$\frac{\partial r_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial \boldsymbol{r}_{\mathbf{0}Ii}} = \frac{\partial r_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial \boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}} \frac{\partial \boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial \boldsymbol{r}_{\mathbf{0}Ii}} = \frac{\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{r_{\mathbf{0}Ii,\boldsymbol{l}Jj}} \frac{\partial \boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial \boldsymbol{r}_{\mathbf{0}Ii}} = -\hat{\boldsymbol{r}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}. \tag{3.4.7}$$

A more detailed description and the derivation of the expressions for the second derivatives can be found in the PhD theses of Vasileiadis [14] and Bowskill [40].

3.4.3 Repulsion-dispersion interactions

Repulsion interactions arise from Pauli's exclusion principle, which excludes electrons from occupying the same quantum state simultaneously. As atoms move closer to each other, the overlap between the electron clouds results in a repulsive force that prevents atoms from collapsing into each other. Repulsion is a short-range force, the magnitude of which increases as the interatomic distance decreases. Contrary to repulsion, dispersion is a long range force that is always attractive. Dispersion forces originate from instantaneous fluctuations of dipole moments between atoms. It is a common practice to model repulsion and dispersion (repulsion-dispersion) interactions using a combined energy model. This choice can be justified for two reasons. Repulsion and dispersion are calculated in a pair-wise additive manner and are usually treated as isotropic in the sense that they depend on the distance (magnitude) and not in the direction of the vectors connecting each atom within the crystal. In the context of CSP, the Buckingham potential [89] is the most popular choice for modelling repulsion-dispersion parameters. The wide adoption of the Buckingham potential can be attributed to the availability of parameters fitted to a large set of experimental molecular crystal data performed and available as the FIT potential [90]. Recently, the Buckingham potential parameters have been refitted to computational molecular structure data, where the structures have been optimised using DFT-d at various levels of theory [40]. The expression for the Buckingham potential reads:

$$U^{\text{rep/dis}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}(r_{\mathbf{0}Ii,\boldsymbol{l}Jj}) = A_{\kappa,\kappa'} \exp\Big(-\frac{r_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{B_{\kappa,\kappa'}}\Big) - \frac{C_{\kappa,\kappa'}}{(r_{\mathbf{0}Ii,\boldsymbol{l}Jj})^6}, \tag{3.4.8}$$

where $A_{\kappa,\kappa'}$, $B_{\kappa,\kappa'}$ and $C_{\kappa,\kappa'}$ are the Buckingham potential parameters, $\kappa = \kappa(Ii)$ and $\kappa' = \kappa'(Jj)$ are the types of atoms Ii and Jj, respectively. The parameters proposed from Williams [90, 91] (FIT) are commonly used in our energy model to describe like-like interactions. In the case of unlike atom-pair interactions, *i.e.* between two atoms of different types, the Buckingham potential parameters are given by application of the following mixing rules:

$$A_{\kappa,\kappa'} = \sqrt{A_\kappa A_{\kappa'}}, \quad B_{\kappa,\kappa'} = \frac{2B_\kappa B_{\kappa'}}{B_\kappa + B_{\kappa'}}, \quad C_{\kappa,\kappa'} = \sqrt{C_\kappa C_{\kappa'}}. \tag{3.4.9}$$

3.4.4 Electrostatic interactions

There are a number of available formulations for the description of the electrostatic interactions, differing in sophistication and complexity. One of the simplest electrostatic models is

the Coulomb potential, U^{Coulomb}, wherein a point charge is assigned to each atom and the electrostatic energy depends on interatomic distances as

$$U^{\text{Coulomb}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}(r_{\mathbf{0}Ii,\boldsymbol{l}Jj}) = \frac{1}{4\pi\epsilon_0}\frac{Q_{Ii}Q_{Jj}}{r_{\mathbf{0}Ii,\boldsymbol{l}Jj}}, \tag{3.4.10}$$

where ϵ_0 is the electric permittivity *in vacuo*, Q_{Ii} and Q_{Jj} are the fixed point charges of atoms i in molecule I and atom j in molecule J, respectively. In our energy models, point charges are utilised in the CrystalPredictor algorithm, used to conduct the global search stage. The CrystalPredictor point charges are parameterised using the ChelpG fitting scheme [92], by conducting an isolated molecule calculation using the Gaussian software [88].

More detailed models adopt a QM approach that allows for the evaluation of the electron density at different levels of theory. This description can offer higher accuracy albeit at an increased computational cost. An attractive balance between efficiency and accuracy is offered by distributed multipole analysis (DMA) [31], in which permanent multipole moments are fitted to the charge densities obtained from isolated molecule QM calculations [88] using the GDMA program [31]. DMA results in an anisotropic potential that depends both on the magnitudes and the orientations of the vector connecting each pair of interacting atoms. The distributed multiple moments are obtained by a multipole expansion using a Cartesian or Spherical tensor formalism. The latter formalism is adopted in the lattice energy and free energy refinement stages. The expression for the DMA electrostatic energy reads:

$$U^{\text{ele}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}(\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}) = \frac{1}{4\pi\epsilon_0}\sum_{l_{Ii}=0}^{l\leq 4}\sum_{l_{Jj}=0}^{l\leq 4}\sum_{k_{Ii}\in K_{Ii}}\sum_{k_{Jj}\in K_{Jj}}[l_{Ii}l_{Jj}]Q^{k_{Ii}}_{l_{Ii}}Q^{k_{Jj}}_{l_{Jj}}\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\mathbf{0}Ii,\boldsymbol{l}Jj})}{(r_{\mathbf{0}Ii,\boldsymbol{l}Jj})^{l+1}} \tag{3.4.11}$$

where the factor $[l_{Ii}l_{Jj}]$ is given by:

$$[l_{Ii}l_{Jj}] = (-1)^{l_{Ii}}\sqrt{\frac{(2l_{Ii}+2l_{Jj}+1)!}{(2l_{Ii})!(2l_{Jj})!}} \tag{3.4.12}$$

$Q^{k_{Ii}}_{l_{Ii}}$ and $Q^{k_{Jj}}_{l_{Jj}}$ are the multipole moments of atom i in molecule I and atom j in molecule J, respectively. The multipole ranks, l_{Ii} and l_{Jj} accounts for the order of interaction. The multipole moments are named after their rank as charges ($l_{Ii}=0$), dipoles ($l_{Ii}=1$), quadrupoles ($l_{Ii}=2$), octapoles ($l_{Ii}=3$), and hexadecapoles ($l_{Ii}=4$). $l=l_{Ii}+l_{Jj}$ is the combined rank of multipole interaction between moments $Q^{k_{Ii}}_{l_{Ii}}$ and $Q^{k_{Jj}}_{l_{Jj}}$. Interactions of a combined rank up to four ($l=4$) are considered in the summation. Each multipole moment of rank l_{Ii} has $2l_{Ii}+1$ components taking values in between $K_{Ii}=\{-l_{Ii}, l_{Ii}-1, ..., 0, l_{Ii}-1, l_{Ii}\}$.

$S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\mathbf{0}Ii,\boldsymbol{l}Jj})$ is the so-called S-tensor describing the orientation-dependence of multipole moments [93]. Ω_I and Ω_J are matrices defining the local axis systems of molecules I and J, respectively. The unit vector connecting the two interacting sites, $\hat{\boldsymbol{r}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}$, is given by Equation (3.4.6). For charge-charge (monopole-monopole) interactions ($l=0$), there is no orientational dependence as the interaction is isotropic ($S^{00}_{000}=0$). In this case, Equation (3.4.11) reduces to the Coulomb potential given by Equation (3.4.10). Analytical expressions

for $S^{00}_{101}, S^{\pm 10}_{101}, S^{00}_{011}, S^{0\pm 1}_{101}, S^{00}_{110}, S^{\pm 1\pm 1}_{110}$, and S^{00}_{111} have been reported by Price and Stone [94] and S-tensors for higher multipole moments are derived using recurrence relations.

3.5 Dynamical matrix

As discussed in Section 3.4, the lattice energy is decomposed into intramolecular, repulsion-dispersion and electrostatic contributions. Similarly, we can follow the same breakdown for the dynamical matrix expression

$$\boldsymbol{D} = \boldsymbol{D}^{\text{intra}} + \boldsymbol{D}^{\text{rep/disp}} + \boldsymbol{D}^{\text{ele}}. \tag{3.5.1}$$

The intramolecular dynamical matrix is calculated by evaluating the phonon frequencies of the isolated molecules by performing gas phase calculations using GAUSSIAN [88]. The repulsion-dispersion and electrostatic contributions to the dynamical matrix are derived analytically. We place particular emphasis on the generalised Ewald summation method used for calculating the poorly-converging electrostatic energy, as calculated with DMA. In the remainder of this section, we derive the dynamical matrix expression. This requires the evaluation of the second derivatives of the components of the lattice energy with respect to atomic positions.

3.5.1 Derivatives of repulsion-dispersion interactions

As mentioned in Section 3.4.3, the repulsion-dispersion interactions are a function of the scalar distance between any two arbitrary atoms $\mathbf{0}Ii$ and $\boldsymbol{l}Jj$. The derivative of the repulsion-dispersion energy between any two interacting atoms with respect to the distance is given by:

$$\frac{\partial U^{\text{rep/dis}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial r_{\mathbf{0}Ii,\boldsymbol{l}Jj}} = -\frac{A_{\kappa,\kappa'}}{B_{\kappa,\kappa'}} \exp\Big(-\frac{r_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{B_{\kappa,\kappa'}}\Big) + 6\frac{C_{\kappa,\kappa'}}{(r_{\mathbf{0}Ii,\boldsymbol{l}Jj})^7} \tag{3.5.2}$$

In the dynamical matrix expression, shown in Equation (3.2.14), the differentiation of the lattice energy component is performed with respect to the vector connecting the two interacting atoms. The derivative of the repulsion-dispersion energy with respect to $\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}$ is evaluated by application of the chain rule

$$\begin{aligned}\frac{\partial U^{\text{rep/dis}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial \boldsymbol{r}_{\mathbf{0}Ii}} &= \frac{\partial U^{\text{rep/dis}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial r_{\mathbf{0}Ii,\boldsymbol{l}Jj}} \frac{\partial r_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial \boldsymbol{r}_{\mathbf{0}Ii}} \\ &= \frac{\partial U^{\text{rep/dis}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial r_{\mathbf{0}Ii,\boldsymbol{l}Jj}} \frac{\partial r_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial \boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}} \frac{\partial \boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial \boldsymbol{r}_{\mathbf{0}Ii}}\end{aligned} \tag{3.5.3}$$

where the terms on the right hand side are given by Equations (3.5.2) and (3.4.7). The second-order derivative of the repulsion-dispersion interaction with respect to $\boldsymbol{r}_{\mathbf{0}Ii}$ is given by

$$\frac{\partial^2 U^{\text{rep/dis}}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{\partial (r_{\mathbf{0}Ii,\boldsymbol{l}Jj})^2} = \frac{A_{\kappa,\kappa'}}{B^2_{\kappa,\kappa'}} \exp\Big(-\frac{r_{\mathbf{0}Ii,\boldsymbol{l}Jj}}{B_{\kappa,\kappa'}}\Big) - 42\frac{C_{\kappa,\kappa'}}{(r_{\mathbf{0}Ii,\boldsymbol{l}Jj})^8} \tag{3.5.4}$$

3.5.2 Generalised Ewald summation method

The sum shown in Equation (3.4.11) decays poorly due to the inverse relation between U^{ele} and $r_{\mathbf{0}Ii,\boldsymbol{l}Jj}$. In addition the sum is conditionally-convergent as the obtained value depends on the utilised cutoff. The Ewald summation method [57] can be used to calculate poorly- or conditionally-convergent sum. The method involves partitioning the electrostatic contribution into short- and long-range contribution. Short-range contributions converge rapidly and are summed in the direct space. This contribution in known as the real space contribution, U^{real}. The long-range contribution is transformed into an equivalent form using a Fourier transform and the summation is performed in the reciprocal space. This is known as the reciprocal space summation, U^{rec}. The self-correction term, U^{self}, is introduced to remove contributions that are double-counted in the real and reciprocal sums. The partitioned electrostatic energy expression reads:

$$U^{\text{ele}} = U^{\text{real}} + U^{\text{rec}} + U^{\text{self}}. \tag{3.5.5}$$

The full derivation of the generalised Ewald summation method can be found in the PhD thesis of Vasileiadis [14]. Here, we provide some key parts of the derivation and introduce terms appearing in the expression used to evaluate the electrostatic energy through its various contributions. In the work of Nijboer and De Wette [95], the Ewald summation method is derived by splitting the lattice energy expression into an equivalent one using an auxiliary function of the form $f(x) = f(x)\mathcal{F}(x) + f(x)[1 - \mathcal{F}(x)]$, with the auxiliary function defined as

$$\mathcal{F}(r_{\mathbf{0}Ii,\boldsymbol{q}Jj}) = \frac{\Gamma(\frac{1}{2}+l, \alpha\pi(r_{\mathbf{0}Ii,\boldsymbol{q}Jj})^2)}{\Gamma(\frac{1}{2}+l)},\ 1 - \mathcal{F}(r_{\mathbf{0}Ii,\boldsymbol{q}Jj}) = \frac{\gamma(\frac{1}{2}+l, \alpha\pi(r_{\mathbf{0}Ii,\boldsymbol{q}Jj})^2)}{\Gamma(\frac{1}{2}+l)} \tag{3.5.6}$$

Here, the (complete) Gamma function, $\Gamma(s)$, is defined as:

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1} \mathrm{d}t \tag{3.5.7}$$

and can be regarded as an extension of the factorial for non-integer values of s. s can generally be any complex number except a negative integer. It is useful to partition $\Gamma(s)$ into two integrals from 0 to x and from x to $+\infty$. The lower incomplete gamma function, $\gamma(s,x)$, and upper incomplete Gamma function, $\Gamma(s,x)$, are defined as:

$$\gamma(s,x) = \int_0^x e^{-t} t^{s-1} \mathrm{d}t, \quad \Gamma(s,x) = \int_x^\infty e^{-t} t^{s-1} \mathrm{d}t \tag{3.5.8}$$

From the definitions, it follows that $\gamma(s,x) + \Gamma(s,x) = \Gamma(s)$. In the context of DMA we are interested in lattice sums where $s = l + \frac{1}{2}$ and $x = \alpha\pi(r_{\mathbf{0}Ii,\boldsymbol{l}Jj})^2$. The Ewald summation method of of Nijboer and De Wette [95] is applied to poorly- or conditionally-convergent sums of the general form:

$$\underbrace{\sum_{\boldsymbol{l}\in\boldsymbol{L}}}_{\boldsymbol{l}\neq\mathbf{0} \text{ if } I=J} \frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I, \Omega_J, \hat{\boldsymbol{r}}_{\mathbf{0}Ii,\boldsymbol{l}Jj})}{(r_{\mathbf{0}Ii,\boldsymbol{l}Jj})^{l+1}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}, \tag{3.5.9}$$

The conditionally-convergent part of the electrostatic energy expression, which is given by Equation (3.4.11), contains the S-tensor over the interatomic distance and exponential factor, $e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}$. The evaluation of the lattice energy is performed for $\boldsymbol{k}=\mathbf{0}$ (wave vector equal to zero). In the resulting expression, the electrostatic component of the dynamical matrix is partitioned into real, reciprocal and self-correction contributions as well as a molecular correction term.

$$\begin{aligned}
&\underbrace{\sum_{\boldsymbol{l}\in\boldsymbol{L}}}_{\boldsymbol{l}\neq\mathbf{0}\text{ if }I=J}\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\mathbf{0}Ii,\boldsymbol{l}Jj})}{(r_{\mathbf{0}Ii,\boldsymbol{l}Jj})^{l+1}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}=\\
&\quad=\frac{1}{\Gamma(\frac{1}{2}+l)}\Bigg[\underbrace{\sum_{\boldsymbol{l}\in\boldsymbol{L}}}_{\boldsymbol{l}\neq\mathbf{0}\text{ if }I=J}\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\mathbf{0}Ii,\boldsymbol{l}Jj})\Gamma\left(\frac{1}{2}+l,\alpha\pi\left(r_{\mathbf{0}Ii,\boldsymbol{l}Jj}\right)^2\right)}{(r_{\mathbf{0}Ii,\boldsymbol{l}Jj})^{l+1}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}+\\
&\quad+\frac{i^l2^{2-l}\pi^{3/2}}{V}\sum_{\boldsymbol{h}}\left\{||\boldsymbol{h}+\boldsymbol{k}||^{l-2}\exp\left(-\frac{||\boldsymbol{h}+\boldsymbol{k}||^2}{4\alpha\pi}\right)S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\widehat{\boldsymbol{h}+\boldsymbol{k}})e^{-i\boldsymbol{h}\cdot\boldsymbol{r}_{Ii,Jj}}\right\}-\\
&\quad-\delta_{I,J}(1-\delta_{i,j})\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\mathbf{0}Ii,\boldsymbol{l}Jj})\gamma\left(\frac{1}{2}+l,\alpha\pi\left(r_{\mathbf{0}Ii,\boldsymbol{l}Jj}\right)^2\right)}{(r_{\mathbf{0}Ii,\boldsymbol{l}Jj})^{l+1}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\mathbf{0}Ii,\boldsymbol{l}Jj}}-\\
&\quad-\delta_{Ii}\delta_{Jj}T^0_s\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right)\Bigg],
\end{aligned}\tag{3.5.10}$$

where δ_{ij} is the Kronecker delta, which is 1 if the indices i and j are equal, and 0 otherwise. The first term on the right-hand side of equation (3.5.10) is sampled in the real space and is referred to as the direct sum. The second term is the reciprocal sum, which is sampled in the reciprocal space. The remaining two terms on the right-hand side of equation (3.5.10) correspond to the molecular correction and self-correction both of which are sampled in the real space.

For molecular crystals, the molecular correction term [20] is used to avoid double accounting of electrostatic interactions between atoms belonging to the same molecule as these interactions have already been included in the intramolecular part. The self-correction accounts for the interaction of a charged particle with its periodic images, with its main component, T^0_s, is given by

$$T^0_s\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right)=\lim_{r\to0}\left[S^{m_{Ii}m_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}})\frac{\gamma\left(\frac{1}{2}+l,\alpha\pi r^2\right)}{r^{l+1}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\right]=\delta_{l0}2\sqrt{\alpha\pi},\tag{3.5.11}$$

where, δ_{l0} implies that the self-correction is $2\sqrt{\alpha\pi}$ for charge-charge interactions and 0 in any other case.

3.5.3 Dynamical matrix expression

As already mentioned earlier, the dynamical matrix expression consists of the mass-weighed Hessian matrix of the second-order derivatives of the lattice energy with respect to atomic

displacements multiplied with the exponential phase factor, $e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}}$. Referring to Equation (3.4.11), this requires the second-order derivatives of the S-tensor, which cannot be directly derived from Equation (3.5.10) due to the presence of the exponential phase factor on the left-hand side. Instead, we adopt an approach similar to that of Signorini *et al.* [96] and differentiate Equation (3.5.10) with respect to $r_{\boldsymbol{0}Ii}$ using the product rule. The resulting equation is rearranged to yield an expression for the desired term, namely the derivative of S with respect to $r_{\boldsymbol{0}Ii}$ multiplied by $e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}}$:

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{r}_{\boldsymbol{0}Ii}}\left[\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})}{(r_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})^{l+1}}\right]e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}} = \\
&= \frac{\partial}{\partial \boldsymbol{r}_{\boldsymbol{0}Ii}}\left[\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})}{(r_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})^{l+1}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}}\right] - i\boldsymbol{k}\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})}{(r_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})^{l+1}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}}
\end{aligned} \tag{3.5.12}$$

The resulting expression for the electrostatic component of the Jacobian matrix reads:

$$\begin{aligned}
&\underbrace{\sum_{\boldsymbol{l}\in\boldsymbol{L}}}_{\boldsymbol{l}\neq\boldsymbol{0}\text{ if }I=J}\frac{\partial}{\partial \boldsymbol{r}_{\boldsymbol{0}Ii}}\left[\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})}{(r_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})^{l+1}}\right]e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}} = \\
&= \frac{1}{\Gamma(\frac{1}{2}+l)}\left[\underbrace{\sum_{\boldsymbol{l}\in\boldsymbol{L}}}_{\boldsymbol{l}\neq\boldsymbol{0}\text{ if }I=J}\frac{\partial}{\partial \boldsymbol{r}_{\boldsymbol{0}Ii}}\left[\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})\Gamma\left(\frac{1}{2}+l,\alpha\pi\left(r_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}\right)^2\right)}{(\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})^{l+1}}\right]e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}}+\right. \\
&+\frac{i^l 2^{2-l}\pi^{3/2}}{V}\sum_{\boldsymbol{h}}\left\{\frac{-i(\boldsymbol{h}+\boldsymbol{k})}{||\boldsymbol{h}+\boldsymbol{k}||^{2-l}}\exp\left(-\frac{||\boldsymbol{h}+\boldsymbol{k}||^2}{4\alpha\pi}\right)S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\widehat{\boldsymbol{h}+\boldsymbol{k}})e^{-i\boldsymbol{h}\cdot\boldsymbol{r}_{Ii,Jj}}\right\}- \\
&-\delta_{I,J}(1-\delta_{i,j})\frac{\partial}{\partial \boldsymbol{r}_{\boldsymbol{0}Ii}}\left[\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})\gamma\left(\frac{1}{2}+l,\alpha\pi\left(r_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}\right)^2\right)}{(r_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})^{l+1}}\right]e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}}- \\
&\left.-\delta_{Ii}\delta_{Jj}T^1_s\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right)\right],
\end{aligned} \tag{3.5.13}$$

where the self-correction term $T^1_s\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right)$ is given by:

$$T^1_s\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right) = \frac{\partial}{\partial \boldsymbol{r}_{\boldsymbol{0}Ii}}T^0_s\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right) - i\boldsymbol{k}T^0_s\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right). \tag{3.5.14}$$

Following a similar procedure we differentiate Equation (3.5.13) with respect to $\boldsymbol{r}_{\boldsymbol{0}Ii}$ and subtract Equation (3.5.13) multiplied by $i\boldsymbol{k}$. The component of the electrostatic dynamical

matrix reads:

$$\underbrace{\sum_{\boldsymbol{l}\in\boldsymbol{L}}}_{\boldsymbol{l}\neq\boldsymbol{0}\text{ if }I=J}\frac{\partial^2}{\partial\boldsymbol{r}_{\boldsymbol{0}Ii}\partial\boldsymbol{r}_{\boldsymbol{l}Jj}}\left[\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})}{(r_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})^{l+1}}\right]e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}}=$$

$$=\frac{1}{\Gamma(\frac{1}{2}+l)}\left[\underbrace{\sum_{\boldsymbol{l}\in\boldsymbol{L}}}_{\boldsymbol{l}\neq\boldsymbol{0}\text{ if }I=J}\frac{\partial^2}{\partial\boldsymbol{r}_{\boldsymbol{0}Ii}\partial\boldsymbol{r}_{\boldsymbol{l}Jj}}\left[\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})\Gamma\left(\frac{1}{2}+l,\alpha\pi\left(r_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}\right)^2\right)}{\left(\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}\right)^{l+1}}\right]e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}}+\right.$$

$$+\frac{i^l2^{2-l}\pi^{3/2}}{V}\sum_{\boldsymbol{h}}\left\{\frac{(\boldsymbol{h}+\boldsymbol{k})(\boldsymbol{h}+\boldsymbol{k})^T}{||\boldsymbol{h}+\boldsymbol{k}||^{2-l}}\exp\left(-\frac{||\boldsymbol{h}+\boldsymbol{k}||^2}{4\alpha\pi}\right)S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\widehat{\boldsymbol{h}+\boldsymbol{k}})e^{-i\boldsymbol{h}\cdot\boldsymbol{r}_{Ii,Jj}}\right\}-$$

$$-\delta_{I,J}(1-\delta_{i,j})\frac{\partial^2}{\partial\boldsymbol{r}_{\boldsymbol{0}Ii}\partial\boldsymbol{r}_{\boldsymbol{l}Jj}}\left[\frac{S^{k_{Ii}k_{Jj}}_{l_{Ii}l_{Jj}l}(\Omega_I,\Omega_J,\hat{\boldsymbol{r}}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})\gamma\left(\frac{1}{2}+l,\alpha\pi\left(r_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}\right)^2\right)}{(r_{\boldsymbol{0}Ii,\boldsymbol{l}Jj})^{l+1}}\right]e^{i\boldsymbol{k}\cdot\boldsymbol{r}_{\boldsymbol{0}Ii,\boldsymbol{l}Jj}}-$$

$$\left.-\delta_{Ii}\delta_{Jj}T_s^2\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right)\right] \tag{3.5.15}$$

where the self-correction term $T_s^2\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right)$ is given by:

$$T_s^2\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right)=\frac{\partial}{\partial\boldsymbol{r}_i}T_s^1\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right)-i\boldsymbol{k}T_s^1\left(l_{Ii},l_{Jj},k_{Ii},k_{Jj},\frac{1}{2}\right). \tag{3.5.16}$$

The expression for T_s^2 can be found in Appendix B of the PhD thesis of Vasileiadis [14].

3.6 CSP approach adopted in this thesis

The CSP methodology adopted in this thesis is carried out in a multistage manner. This involves conformational analysis, deterministic global optimisation, lattice energy and free energy refinement stages. Each stage is discussed separately below.

Conformational analysis: In the first stage, a conformational analysis is conducted to determine the conformational degrees of freedom (DoFs) that will be treated as flexible in the next stage (global search). The conformational DoFs are divided into bonds, angles and torsions. Conformational analysis is carried out through a series of constrained isolated-molecule QM geometry minimisations. Each DoF under consideration is constrained to various predefined values, while the remaining DoFs are allowed to vary freely during the gas-phase optimisation. The goal is to establish a relationship between the minimised isolated-molecule energy and the values of the DoFs, thus enabling the detection of low-energy regions. For molecules with few independent degrees of freedom, two-dimensional scans are sometimes conducted (see Section 5.3.1).

LAMs: Local approximate models (LAMs) are used to increase the efficiency of the global search by reducing the number of QM calculations needed for the evaluation of the electrostatic

and intramolecular interactions. LAMs are constructed "on the fly", during the minimisation, by interpolation using a second-order Taylor series expansion around points for which an isolated-molecule QM calculation has already been performed. The algorithm uses information from nearby LAMs thus reducing the number of costly gas phase calculations needed for the lattice energy minimisation. An initial LAM database is generated by carrying out isolated-molecule QM calculations with the selected basis at different values of the independent torsions. Uniform LAMs are created through constrained geometry optimisation of isolated molecules for different molecular conformations. The independent DoFs are restricted to a grid of discrete values, while the dependent DoFs are allowed to vary freely. The QM geometry minimisations are carried out for different combinations of basis sets (e.g. HF, PBE or PBE0) and levels of theory (e.g. 6-31G(d,p), 6-311G(d,p), or 6-311+G(d,p)). Those basis sets that do not provide sufficient accuracy, as determined by comparison with the results of calculations with the most extended basis set considered, are eliminated. Those basis sets that yield high accuracy at a comparatively high computational cost per LAM point are also eliminated. Note that in some cases, the choice of DoFs that are treated as flexible and their range is based on chemical intuition.

Global search: The next stage involves global search, which is carried out using the CrystalPredictor algorithm [97, 21, 22, 7]. The aim of the global search is to generate candidate structures as low-energy minima in the lattice energy landscape. The structures are generated based on a quasi-random Sobol' sequence [19]. The global search stage is carried out with CrystalPredictorII [98, 99, 22] version 2.4.4 [97, 21]. Candidate structures are generated using the 61 commonly encountered space groups that account for approximately 97.9% of all the organic crystal structures found in the CSD [100]. Specifically, the space groups sampled are: $P1$, $P\bar{1}$, $P2_1$, $P2_1/c$, $P2_12_12$, $P2_12_12_1$, $Pna2_1$, $Pca2_1$, $Pbca$, $Pbcn$, $C2/c$, Cc, $C2$, Pc, Cm, $P2_1/m$, $C2/m$, $P2/c$, $C222_1$, $Pmn22_1$, $Cmc2_1$, $Aba2$, $Fdd2$, $Iba2$, $Pnna$, $Pccn$, $Pbcm$, $Pnnm$, $Pmmn$, $Pnma$, $Cmcm$, $Cmca$, $Fddd$, $Ibam$, $P4_1$, $P4_3$, $I\bar{4}$, $P4/n$, $P4_2/n$, $I4/m$, $I4_1/a$, $P4_12_12$, $P4_32_12$, $P\bar{4}_22_1/c$, $I\bar{4}2d$, $P3_1$, $P3_2$, $R3$, $P\bar{3}$, $R\bar{3}$, $P3_12_1$, $P3_22_1$, $R3/c$, $R\bar{3}/c$, $P6_1$, $P6_3$, $P6_3/m$, $P2_13$, $Pa\bar{3}$, $P222_1$ and $Pba2$. Each space group is sampled with a relative frequency matching that of the space group occurrences in the CSD. In the global search stage the electrostatic interactions are calculated utilising point charges (see Equation (3.4.10)) fitted to isolated-molecule QM calculations [88]. The repulsion-dispersion interactions are modelled with FIT parameters. Each generated structure is minimised with the aforementioned energy model and all structures are subsequently processed (clustered) to eliminate duplicate structures generated in global search. The similarity between two crystal structures is quantified by means of the root mean squared deviation ($\text{RMSD}_{\mathcal{N}}$) of atomic coordinates between any two assessed structures, where $\mathcal{N}$ is the number of molecules being considered for comparison. The RMSD_{15} is used to evaluate the differences in the atomic coordinates between two shells, each containing fifteen (15) molecules, providing a similarity metric for the packing. RMSD_1 is used to compare the differences in the intramolecular conformations of a single molecule within each compared structure. Lower RMSD values indicate a better match between the structures, and according to the 6$^{\text{th}}$ CCDC

blind test, two structures can be considered a match for $RMSD_{20}$ values below 0.8 Å [11]. The CSD python API [101] is used within the clustering algorithm. The comparisons for clustering are performed for any pair of structures, with an energy difference below 5 kJ/mol and 0.025 g/cm^3. Structures with $RMSD_{15}$ below 0.8 Å are considered crystallographicaly similar and the structure with the largest lattice energy is removed. Additionally, the simulated powder X-ray diffraction (PXRD) patterns are also used to assess the similarity between structures [102]. Regarding the PXRD, matching score values closer to one indicate a better similarity between the compared structures.

Refinement stages In the first refinement stage, the unique lowest energy structures identified in the global search stage are locally minimised using CrystalOptimizer [18] or CSO-FM [40]. The local minimisation involves lattice energy minimisation using a more sophisticated description of the electrostatic interactions, based on the distributed multipole analysis [31]. The energy model used in the first refinement stage is the energy models presented in Section 3.4. In addition to the more detailed description of the electrostatics interactions, the lattice energy minimisation algorithms account for additional molecular flexibility. More details of the lattice energy minimisation are given in Section 4.2. In the next chapter, we describe the CrystalDynamics algorithm used to evaluate the vibrational free energy of molecular crystals and assess its suitability to be used as a second refinement stage.

3.7 Summary

In this chapter, we have presented a mathematical formalism for performing free energy calculations using the harmonic approximation of lattice dynamics theory. A number of approximations have been considered for the development of the methodology described in this chapter. The harmonic approximation assumes that the Taylor series expansion is performed around a local minimum of the lattice energy hypersurface. As a result, the harmonic potential is expressed as a quadratic function of interatomic distances and the effect of higher order terms in the Taylor series expansion is neglected. The lattice lattice dynamics equations have been initially derived for the general case of a potential expressed as a function of the atomic positions in the crystal. The utilised energy model accounts only for pairwise interactions, i.e. the energy of the crystal is calculated as the sum of interactions between pairs of atoms neglecting the effect of many body interactions. In turn, the tailor-made CSP force field used in this thesis has been described in detail. Our tailor-made force field partitions lattice energy into intramolecular, repulsion-dispersion and electrostatic interactions. Special emphasis is placed on the poorly-convergent sums appeared in the electrostatic expression, which is calculated by employing the generalised Ewald summation method. The analytical derivation of the dynamical matrix expression provides robustness and efficiency as it allows for the evaluation of phonon frequencies at various points in the reciprocal space. In the next chapter, we describe the CrystalDynamics algorithm and apply it to evaluate the vibrational free energy of molecular crystals.

Chapter 4

Assessment of CSP workflows & local energy minimisation algorithms

In this chapter, we begin with presenting the CrystalDynamics algorithm that evaluates the vibrational free energy of molecular crystals. We provide details regarding the implementation, which is based on the mathematical formulation and methodology described in Chapter 3. The objective of this chapter is to introduce a reliable workflow, capable of calculating the Helmholtz free energy of rigid and flexible molecules. Our free energy workflow involves lattice energy minimisation using one of the two optimisation algorithms, CrystalOptimizer or Crystal Structure Optimizer - Flexible Molecules (CSO-FM), followed by vibrational free energy evaluation using CrystalDynamics. The performance of the two utilised geometry optimisation algorithms is compared based on their ability to converge tightly to a local minimum in lattice energy. Additionally, we assess whether minimised crystal structures satisfy the essential condition of the harmonic approximation requiring the crystal structure to lie on a local minima in the lattice energy landscape. First derivatives are utilised to identify the proximity to lattice energy minima, while the sign of phonon frequencies is used to determine the structure's stability in the free energy landscape. We also examine the sensitivity of the workflow to numerical parameters and to assumptions, such as alternative treatments of molecular flexibility and utilisation of LAMs. The workflow is applied to the two tetracyanoethylene (TCNE) polymorphs, Target XXIX from the 7^{th} CSP blind test, and the lowest 100 predicted structures of the Benzoic Acid landscape generated using CrystalPredictor.

4.1 The CrystalDynamics algorithm

The CrystalDynamics algorithm [14] evaluates the Helmholtz free energy of molecular crystals by employing the harmonic approximation of lattice dynamics theory. The flowchart for computing vibrational free energies is presented in Figure 4.1, where each Box is numbered and discussed throughout the text. The step-by-step algorithmic procedure is shown in Figure 4.2. Prior to the free energy calculation, the geometry of the crystal structure must be minimised using either CrystalOptimizer or CSO-FM lattice energy minimisation algorithms [Box 1]. The CrystalDynamics algorithm calculates each contribution to lattice energy, as well as first and second derivatives. The utilised energy model is consistent with the one used in the lattice en-

ergy minimisation and incorporates intramolecular and intermolecular interactions to the lattice energy. Isolated-molecule QM calculations [Box 2] are performed with the GAUSSIAN 16 software [88] for the evaluation of the intramolecular contribution and its derivatives, as well as the gas phase electron density. The electrostatic contributions are modelled via a distributed multipole analysis, with the multipole moments generated using GDMA [31] program [Box 3]. The repulsion-dispersion interactions, first and second derivatives are calculated using an anisotropic Buckingham potential [89]. The dynamical matrix is computed as the mass weighted Hessian matrix of the second order derivatives of the lattice energy multiplied by an exponential phase factor [Box 4]. Phonon frequencies [Box 5] are obtained from diagonalisation of the dynamical matrix at various sampled $\boldsymbol{k}$ points and can be plotted using dispersion curves. A histogram of vibrational states within given frequency ranges is utilised to generate the phonon density of states (pDOS) from the dispersion curves [Box 6]. The free energy is then evaluated using the Gauss-Legendre quadrature (GLQ) numerical integration method. The CrystalDynamics algorithm evaluates the vibrational contributions to free energy, namely, entropy, zero point energy (ZPE) and thermal energy [Box 7].

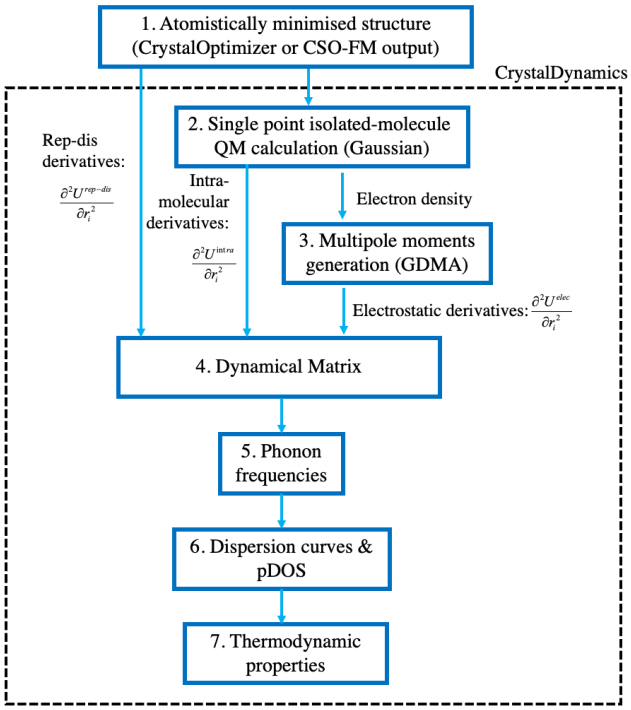


Figure 4.1: Diagrammatic flowchart of CrystalDynamics algorithm. Names inside parentheses denote calls to external algorithms. Each box is discussed in the text and Figure 4.2.

The first version of CrystalDynamics was implemented by Manolis Vasileiadis [14] and has been further extended in this work. The analytical expressions for the electrostatic interactions and their second derivatives with respect to atomic positions have been validated against the corresponding expressions implemented in the CrystalDynamics algorithm. In addition, the

thermodynamic expressions for entropy, isochoric heat capacity and thermal energy have been derived analytically and implemented in CrystalDynamics. In the remainder of this section, we initially describe the structure of the input file and then provide details of the implementation. Finally, we describe the output files and the developed extensions of the code.

CrystalDynamics algorithm

Step 1: Perform a lattice energy minimisation using CrystalOptimizer or CSO-FM [Box 1]

Step 2: Read the geometry of the crystal structure at the minimum lattice energy and the input file of CrystalDynamics code

Step 3: Evaluate the repulsion-dispersion interactions, first and second derivatives

Step 4: Evaluate the electron density using GAUSSIAN software. Also call GAUSSIAN to evaluate the intramolecular interactions, first and second derivatives wrt atomic displacements [Box 2]

Step 5: Generate multipoles by calling GDMA and evaluate the electrostatic interaction, first and second derivatives [Box 3]

Step 6: Calculate the dynamical matrix at various $\boldsymbol{k}$ points [Box 4]

Step 7: Diagonalise the dynamical matrix to obtain phonon frequencies [Box 5] as well as dispersion curves and pDOS [Box 6]

Step 8: Calculate the vibrational contribution to free energy and associated thermodynamic properties [Box 7]

Figure 4.2: CrystalDynamics algorithm in a step-by-step procedure.

4.1.1 Input file & Supported types of calculation

The code begins by reading the input file named "vibrations.in". The input file is divided into 7 sections shown in Figure 4.3. The first section contains information regarding the cutoffs of real, r_c^{real} and reciprocal, k_c^{rec}, electrostatic interactions. The repulsion-dispersion interactions are modelled in the real (direct) space with the cutoff given by r_c^{dir}. In addition, the Ewald summation convergence parameter, α, maximum rank of multipole moments, l, and the step size used in numerical derivatives are specified. The second section is used to define options related to the isolated-molecule QM calculation. Specifically, the input file includes details of the level of theory and basis set to be used, a keyword (frequencies) to enable second-derivative

evaluation, as well as a specification of the numbers of processors and memory to be used by GAUSSIAN. In the van der Waals interactions section, the FIT or WILLIAMS type of parameterisation is specified alongside a keyword that enables the execution of lattice dynamics calculations using parameters specifically developed at the PBE0/6311+G(d,p) level of theory. Section 4 is related to the settings associated with the free energy calculation, such as the number of points to be used for the GLQ. Also, the upper, $T_{\max}$, and lower, $T_{\min}$, temperatures as well as the temperature increment, T_{step}, are specified to indicate the temperature values at which the thermodynamic properties are to be computed. In section 5, the number of atoms within a molecule as well as the number of molecules in the asymmetric unit cell and unit cell are defined. Section 6 is dedicated to specifying settings related to the desired type of calculation. These settings include the type of calculation to be employed, number of sampled wave vectors for the density of states, and dispersion curves limits. In the last section, the paths of the executables of the two external programs (GAUSSIAN and GDMA) used by CrystalDynamics are specified. Table 4.1 provides a brief summary of the default parameters values used in the CrystalDynamics algorithm. The crystal geometry is provided via the "crystal.res" file that adopts the format of the SHELXS convention. This input file contains the values of the cell lengths and angles, lattice type, symmetry operation elements, unique types of atoms and fractional coordinates of each atom in the asymmetric unit cell.

Table 4.1: Default parameters values used in the CrystalDynamics algorithm.

Parameter name	Parameter symbol	Default value
real electrostatic cutoff	r_c^{real}	10 Å
reciprocal electrostatic cutoff	k_c^{rec}	0.5 Å^{-1}
repulsion-dispersion cutoff	r_c^{dir}	30 Å
Ewald convergence parameter	α	0.16
combined rank of multipoles moments	l	4
functional/basis set	-	HF/6-31G(d,p)
van der Waals parameterisation	-	FIT
lower temperature	$T_{\min}$	0 K
upper temperature	$T_{\max}$	300 K
increment of temperature	T_{step}	10 K

Within the framework of the CrystalDynamics implementation, five distinct types of calculations are available. These types include: (a) Lattice energy, (b) Dispersion curves, (c) Density of states, (d) Helmholtz free energy and (f) Energy derivatives. The full calculation of the free energy is conducted by specifying "Helmholtz free energy" as the designated calculation type. For calculation types (a)-(c), the code is terminated without evaluating the free energy. These are useful for testing and debugging purposes. The final type of calculation, named "Energy derivatives", involves the analytical and numerical evaluation of the lattice energy derivatives

Section 1: **Lattice energy calculation**
specify: r_c^{real}, k_c^{rec}, r_c^{dir} α, l, FD step size

Section 2: **Quantum mechanics**
functional/basis set
processors and memory used by GAUSSIAN

Section 3: **van der Waals interactions**
FIT or WILLIAMS parametrisation
parameter set derived at specific level of theory

Section 4: **Entropy**
T_{min}, T_{max} and T_{step}
number of points for the GLQ

Section 5: **Asymmetric unit information**
Z, Z', N

Section 6: **Type of calculation**
see Figures 4.4 and 4.5

Section 7: **Paths of external programs**
set paths of GAUSSIAN and GDMA

Figure 4.3: Breakdown of CrystalDynamics input file named "vibrations.in"

and is also useful for validating the analytical derivatives.

4.1.2 Implementation

The code is written in the object oriented FORTRAN 90 programming language making use of linear algebra package (LAPACK) library. The main program is named "CrystalDynamics.f90" and calls a number of subroutines depending on the desired type of calculation. For all implemented types of calculation, the code begins by setting the global variables to their initial values. Next, the two input files are read and the unit cell is constructed. This process is described in Figure 4.4. The CrystalDynamics algorithm reads the values of the lattice lengths and angles and uses them to re-construct the size and shape of the unit cell. This is achieved by the "build_unit_cell" subroutine. The fractional coordinates, $\boldsymbol{r}_{Ii}^{\text{fractional}}$, of symmetry dependent molecules are calculated by application of the symmetry operations found in the "crystal.res" input file. Details of this process can be found in Bowskill's PhD thesis [40]. The unit cell volume is evaluated and the matrix representing the unit cell (See Section 3.1) is used in converting the fractional coordinates into their Cartesian counterparts. In the case that the option "Translate

all atoms in the unit cell" is enabled, all molecules are placed in the central unit cell. Lastly, the body-fixed coordinates for each molecule are evaluated trough the "spacefixed2bodyfixed" subroutine.

1. Call **initialise_global**
2. Call **read_data**
   - 2.1 read input file (.in) and crystal geometry file (.res)
   - 2.2 apply space group considerations
   - 2.3 evaluate $\boldsymbol{r}_{Ii}^{\text{fractional}}$ for symmetry dependent atoms
3. Call **build_unit_cell**
   - 3.1 Call **matrix_representation_unitcell**
     calculate $\boldsymbol{R}$, $\boldsymbol{V}$ and $\boldsymbol{R}^{-1}$
   - 3.2 Call **cart_from_frac**
     convert fractional to Cartesian coordinates $\boldsymbol{r}_{Ii} = \boldsymbol{R}\boldsymbol{r}_{Ii}^{\text{fractional}}$
   - 3.3 Call **report_input_information**
     print values on output file (.out)
   - 3.4 Call **spacefixed2bodyfixed**
     evaluate body fixed from space fixed coordinates

Figure 4.4: CrystalDynamics Initial subroutines called for each type of calculation

Once the input information is read and the unit cell containing the molecules is constructed, the algorithm proceeds by evaluating each lattice energy contribution. A breakdown of the subroutines called for this type of calculation is given in Figure 4.5. Supercells are created by replicating the unit cell along the three dimensions and used in the calculation of all inter-atomic interactions. A more detailed description of supercells is provided in Section 4.2.2. The inter-atomic distances are calculated for each molecule in the supercell and the repulsion-dispersion interactions and derivatives are evaluated in a pair-wise manner. An isolated-molecule QM calculation is performed with GAUSSIAN [88] to obtain the intramolecular first and second derivatives, as well as the charge density. Multipole moments are generated using GDMA [31] using the charge density obtained from the isolated-molecule calculation. The electrostatic interactions and gradients are calculated by employing the generalised Ewald summation method adopting the spherical harmonics formulation. Once all contributions and derivatives have been evaluated, the Hessian matrix is constructed and diagonalised to assess whether the structure lies at a local minimum. For the case of stable structures, the dynamical matrix is sampled and diagonalised at various $\boldsymbol{k}$ points for the evaluation of phonon frequencies and the generation of

1. Call **rep_disp_interaction**
    - 1.1 evaluate the number of layers that supercell consists
    - 1.2 distance vector for each pair of atoms r_{ij} in the unit cell
    - 1.3 evaluate rep-dis interactions and derivatives
2. Call **read_intra_second**
    - 2.1 obtain charge density via a single point QM calculation
    - 2.2 submit a frequency calculation for Hessian evaluation
3. Call **read_multipoles**
    - 3.1 read charge density
    - 3.2 generate multipole moments files using GDMA
4. Call **electrostatics**
    - 4.1 evaluate the direct and reciprocal electrostatics (See Section 4.5.2)
5. Call **dynamical_matrix**
    - 5.1 initialise the dynamical matrix
    - 5.2 construct the dynamical matrix at $\boldsymbol{k}$=0
    - 5.3 obtain phonon frequencies via diagonalisation
6. Call **dispersion_curves**
    - 6.1 construct and diagonalise the dynamical matrix at various $\boldsymbol{k}$ points
7. Call **density_of_states**
    - 7.1 bin dispersion curves with respect to the frequencies
    - 7.2 obtain density of states from the histogram
8. Call **hfec** (Helmholtz free energy calculation)
    - 8.1 evaluate the vibrational Helmholtz free energy using GLQ

Figure 4.5: CrystalDynamics subroutines for free energy calculation

dispersion curves and pDOS. The thermodynamic properties, including vibrational free energies, are evaluated using the GLQ. The number of $\boldsymbol{k}$ points to be sampled in each direction by the GLQ are determined by the user. We have found that 4 or 5 points per direction can capture

the pDOS accurately, resulting in the evaluation and diagonalisation of the dynamical matrix with 64 or 125 $\boldsymbol{k}$ points, respectively.

The first section in the output file contains the information provided in the input file. The code outputs the repulsion-dispersion parameters, as well as fractional and Cartesian coordinates of all molecules in the unit cell. The next part is dedicated to the output of the QM calculations, which is used as input for the generation of multipole moments using GDMA. The atom-centred multipole moments are evaluated for each atom and molecule in the unit cell. The intramolecular second order derivatives are calculated alongside isolated-molecule ZPE and phonon frequencies. The last part of the output file contains a breakdown of all lattice energy contributions. Finally, the sampled dynamical matrices are outputted. For each sampled point, the eigenvalues of the dynamical matrix are printed to validate that they exhibit positive values. The computed thermodynamic properties include Helmholtz free energy, entropy, isochoric heat capacity, and thermal energy.

A number of extensions have been implemented in the CrystalDynamics code to enhance its capabilities. These extensions include the evaluation of entropy, isochoric heat capacity and thermal energy. Moreover, numerical second-order derivative and Hessian evaluations have been integrated into the code. Furthermore, the code now accommodates the treatment of crystals with up to 99 molecules in the unit cell enabling the analysis of a broader variety of experimental or theoretical compounds.

4.2 Lattice energy minimisation

The proposed workflow for the free energy evaluation requires a crystal structure that corresponds to a minimum in the lattice energy. This section is dedicated to the generation of such a structure. We assess a number of available options for performing geometry optimisation with molecular unit cells. We use DMACRYS [39] as accessed via CrystalOptimizer [18] to optimise unit cells comprised of rigid molecules and Crystal Structure Optimizer - Rigid Molecules (CSO-RM) [40] wrapped inside CSO-FM for flexible molecules. In the next subsection, the bilevel lattice energy minimisation problem is formulated. The lattice energy of the minimised structures is also calculated using CrystalDynamics to compare the consistency between the energy models employed by the lattice energy minimisation and free energy evaluation algorithms.

4.2.1 Bilevel minimisation of lattice energy

As discussed in Section 3.4. the lattice energy, U^{latt}, can be decomposed into intramolecular, ΔU^{intra}, and intermolecular, U^{inter}, energy contributions. ΔU^{intra} captures the energy required to deform the molecule from its most stable gas phase conformation. U^{inter} refers to the intermolecular energy arising from the interactions between atoms on separate molecules in the crystal. The lattice energy expression is given by

$$U^{\text{latt}} = \Delta U^{\text{intra}}(\boldsymbol{\theta}) + U^{\text{inter}}(\boldsymbol{X}, \boldsymbol{\theta}), \tag{4.2.1}$$

where vector $\boldsymbol{\theta}$ denotes a set of conformational degrees of freedom (DoFs) and vector $\boldsymbol{X}$ collectively denotes the lattice variables (cell angles and lengths) and the rigid body coordinates (positions of centres of mass and Euler angles) of the molecules in the asymmetric unit cell. The description of a molecule using atomic coordinates (described in Chapter 3) is equivalent to the representation utilising conformational DoFs. Details of the transformation between atomic coordinates ($\boldsymbol{r}$) and DoFs ($\boldsymbol{\theta}$) can be found in Appendix A of Kazantsev's PhD thesis [103]. While the two descriptions are equivalent, numerical tolerances may mean that that gradients may differ between the two coordinates systems following and minimisation

The lattice energy minimisation can be formulated as a bilevel optimisation problem, which involves optimising the objective function at two levels: an inner and an outer loop [18, 103, 104]. In the inner loop the lattice variables, centres of mass and Euler angles are treated as optimisation variables to minimise an objective function that consists intermolecular interactions at fixed values of the DoFs. In the outer loop, the optimisation variables are conformational DOFs including bonds angles and dihedrals. Using this formalism the minimisation of lattice energy can be written as:

$$\min_{\boldsymbol{\theta},\boldsymbol{X}} U^{\text{latt}} = \min_{\boldsymbol{\theta}} \left[\Delta U^{\text{intra}}(\boldsymbol{\theta}) + \min_{\boldsymbol{X}} U^{\text{inter}}(\boldsymbol{X};\boldsymbol{\theta})\right]. \tag{4.2.2}$$

The optimum packing of the molecules in the crystal is thus found by performing a rigid body optimisation in the inner problem. The solutions obtained from the inner problem are used to evaluate the objective function of the outer problem. The inner problem is solved for each set of decision variables generated in the outer problem. In the outer problem, the intramolecular contributions to the lattice energy are optimised for each conformational DoFs that are treated as flexible. This is accomplished by performing a constrained gas phase optimisation using the GAUSSIAN software [88]. For each newly visited conformation in the outer problem, a new gas phase optimisation is required for the evaluation of the objective function. These isolated-molecule optimisations are the most computationally expensive step in the minimisation process. To enhance the efficiency of the calculation, local approximate models (LAMs) [97, 21, 22, 7] are used to reduce the amount of QM calculations needed at the outer loop. LAMs are constructed "on the fly", during the minimisation, by interpolation using a second-order Taylor series expansion around points for which an isolated-molecule QM calculation has already been performed. The algorithm uses information from nearby LAMs thus reducing the number of costly gas phase calculations needed for the lattice energy minimisation.

Two different algorithms can be utilised to solve the bilevel optimisation problem described in equation 4.2.2. For the inner problem, the rigid body optimisation can be conducted using either DMACRYS [39] or CSO-RM [40]. The algorithm for outer problem is chosen based on the choice for the inner problem. CrystalOptimizer [18] calls DMACRYS to perform the geometry optimisation. Alternatively, CSO-FM can be employed in the outer problem, with CSO-RM as the inner problem algorithm. First and second derivatives of the intermolecular interactions with respect to DoFs are computed using central finite differences. This implies that each DoF is perturbed forward and backwards by a finite difference step size specified by

user. Single point calculations of the intermolecular interactions are carried out to calculate the energy of the perturbed conformations and estimate the derivatives. The Hessian matrix of second-order derivatives of the intramolecular interactions with respect to the flexible DOFs is stored in the database used to construct LAMs. CrystalOptimizer and CSO-FM use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [105] and the variables for each successive iteration are determined using line search. CrystalOptimizer has two termination conditions. The code converges when the calculated gradients are below a user-specified tolerance threshold or when a lower energy crystal structure cannot be found in the current search direction indicating termination on negligible displacements. This termination criterion is disabled to generate input structures fro CrystalDynamics that are lying as close to a true lattice energy minimum as possible.

4.2.2 Consistency in computed lattice energies

In this section, we examine the differences between the two utilised lattice energy minimisation algorithms, namely DMACRYS/CrystalOptimizer and CSO-RM/CSO-FM. Both algorithms can account for extensive molecular flexibility and use distributed multipoles to model the electrostatic interactions. Interactions between multipole moments, with a combined rank of up to four ($l = 4$), have been implemented in all lattice energy minimisation algorithms and CrystalDynamics. However there are some differences in the handling of electrostatic interactions in the reciprocal space.

As mentioned in section 4.2.1, the rigid-body minimisation is carried out using DMACRYS or CSO-RM. In the DMACRYS program [39], the Ewald summation method is applied to charge–charge, charge–dipole, dipole-charge and dipole–dipole interactions. The rest of the electrostatic interactions, up to a combined multipole rank of four, are calculated in the real space due to fast converging expression derived by Leslie [106]. Contrary in CrystalDynamics, the generalised Ewald summation method [14] is utilised for all electrostatic interactions between multipole moments. In addition, DMACRYS is accessed as an external program being called several times during the CrystalOptimizer optimisation procedure. DMACRYS is called for the minimisation of the intermolecular energy and for performing single point calculation for perturbed structures for the evaluation of intermolecular gradients via finite differences. The file containing the crystal geometry is repeatedly utilised both as input and output yielding truncation errors in numerical precision. These inconsistencies are not present in the CSO-FM/CSO-RM minimisation algorithm, where the input file describing the crystal geometry is read once. In contrast to DMACRYS, CSO-RM is implemented based on the generalised Ewald summation method described in Section 3.5.2 and in Vasileiadis thesis [14] and is used for all multipole interactions. A comparison between multipole interaction calculated with DMACRYS and CSO-RM is shown in Bowskill's thesis [40]. It was found that the energy contributions displayed good consistency, with calculated valued not exceeding 10^{-3} kJ/mol when using extended numerical settings.

We examine whether CSO-RM/CSO-FM and CrystalDynamics yield equivalent contributions to the calculated lattice energies. Before we proceed to the comparison, we elucidate two notable inconsistencies between CSO-RM and CrystalDynamics energy models. These are related to the application of splines and the treatment of periodicity within the two algorithms. One difference between the CSO-RM and CrystalDynamics energy models is that the former algorithm makes use of splines [40] in the evaluation of intermolecular contributions to lattice energy. Splines are used to ensure that the interaction decays smoothly to zero at the cutoff value. This avoids discontinuities in the interaction energy potential thus improving the robustness of the minimisation procedure. Each intermolecular contribution to the lattice energy is multiplied with a quintic function for distances between the spline cutoff, x_s, and the cutoff of the relevant interaction, x_c. Here, we have introduced a shorthand in notation describing the cutoffs of real electrostatics, repulsion-dispersion, and reciprocal electrostatics (see Section 4.1.1). These cutoff values have been collectively denoted as $x_c \in \{r_c^{\text{real}}, r_c^{\text{dir}}, k_c^{\text{rec}}\}$. In CSO-RM, the spline cutoff is calculated through the user-defined spline fraction, ζ_s, as

$$x_s = (1 - \zeta_s)x_c. \tag{4.2.3}$$

The quintic splines are applied at distances closer to x_c. The default value is $\zeta_s = 0.1$, with lower ζ_s values resulting in smaller portions of the potential to be splined. In CrystalDynamics, interactions between atoms with distances beyond the defined cutoff limits are omitted from the summations. This allows to maintain the integrity of the lattice energy expression, as well as its first and second derivatives thought the function's domain. It is expected that reducing the spline fraction in CSO-RM would lead to more consistent energies between the two algorithms.

Another difference between CSO-RM and CrystalDynamics lies in the handling of periodicity in both codes. CSO-RM is a lattice energy minimisation algorithm that optimises the position and orientation of molecules, as well as lattice lengths and angles in the asymmetric unit cell. This allows to avoid calculating interactions between symmetry-dependent molecules, thus reducing the computational cost. Conversely in CrystalDynamics, the inter-atomic interactions are calculated for a supercell instead of an asymmetric unit cell. Although the calculation does not benefit from efficiency reduction, it allows to capture collective modes of vibrations between molecules belonging to different unit cells. The supercells are generated by replicating the unit cell in three dimensions. The sizes of the supercells depend on the user specified cutoff values. $r_c^{\text{real}}, r_c^{\text{dir}}$ and k_c^{rec} are specified independently resulting in a different number of supercells constructed for each type of calculation. The number of unit cells to be constructed for each dimension and direction is evaluated as the nearest integer that is greater or equal to the value of the fraction between x_c and the minimum lattice length of the unit cell. Usage of increased cutoff values results in an increased supercell size.

We aim to identify the combinations of r_c^{real} and ζ_s that yield consistent energies between the two codes. For this reason, we firstly perform lattice energy minimisation with CSO-FM and then calculate the lattice energies of the optimised structure with CrystalDynamics. The intermolecular energies are compared, as the final gas phase single point calculation at the end

of the lattice energy minimisation ensures that the same intramolecular energies are used in both codes. To perform the comparison, three different spline fractions $\zeta_s = \{1, 0.1, 0.01\}$ and three different cutoff values for the direct electrostatics, $r_c^{\text{real}} = \{10$ Å, 15 Å, 20 Å$\}$ are utilised. The reciprocal space cutoff and cutoff for repulsion-dispersion interactions are kept constant at $k_c^{\text{rec}} = 0.5$ Å^{-1} and $r_c^{\text{dir}} = 15$ Å, respectively. The utilisation of different cutoff values for the direct electrostatics interactions affects the number of unit cells to be generated in each dimension and direction. Specifically, the three direct electrostatics cutoff values correspond to supercells with 2, 3, and 4 unit cells.

Table 4.2: Compounds studied in this Chapter. More details can be found in the relevant Sections in Chapter 5.

Compound	CSD RefCode	Section in Thesis
Cubic TCNE	TCYETY02	5.1
Monoclinic TCNE	TCYETY	5.1
Benzoic Acid	BENZAC02	5.2
Molecule XXIX	FASMEV	5.3

Table 4.2 provides a brief summary of the molecules examined in Chapter 4. The comparison is performed for the lowest energy structure of CrystalPredictor landscape of Target XXIX. Details of the candidate generation of Target XXIX can be found in Section 5.3.1. The intermolecular energy is decomposed into: repulsion, U^{rep}, dispersion U^{dis}, charge-charge $U^{\text{elec}}(l = 0)$, charge-dipole and dipole-charge $U^{\text{elec}}(l = 1)$, and higher multipole moments $U^{\text{elec}}(l > 1)$. A breakdown of each contribution to the intermolecular energy calculated with CrystalDynamics and CSO-FM is summarised in Table 4.3. It can be observed that altering ζ_s has a more significant impact on the dispersion interactions. This can be attributed to the fact that for the case of non-splined potentials, the dispersion interaction does not decay to zero smoothly. Irrespective of the chosen r_c^{real} value, a tenfold reduction in ζ_s consistently leads to a decrease in the difference between the calculated U^{dis} values obtained from the two codes, by an order of magnitude. The use of different r_c^{real} values has a relatively minor effect on the computed electrostatic interactions evaluated with both codes. We can conclude that $\zeta_s = 0.1$ or $\zeta_s = 0.01$ can provides adequate consistency. In addition, since the effect of altering the r_c^{real} values is less prominent on the differences between multipole interaction energies a value of 10 Å is used as default for both CSO-RM and CrystalDynamics.

Table 4.3: Comparison of intermolecular energy contributions calculated with CSO-FM and CrystalDynamics for various real space cutoffs (r_c^{real}) and splines fractions (ζ_s). Calculations are performed for the lowest energy structure of the CrystalPredictor landscape of Target XXIX. Energies are reported in kJ/mol.

	CrystalDynamics	CSO-FM	Difference	CrystalDynamics	CSO-FM	Difference	CrystalDynamics	CSO-FM	Difference
	$r_c^{\text{real}} = 10$ Å, $\zeta_s = 1$			$r_c^{\text{real}} = 10$ Å, $\zeta_s = 0.1$			$r_c^{\text{real}} = 10$ Å, $\zeta_s = 0.01$		
U^{rep}	60.9116	60.9122	-0.0007	61.2168	61.2173	-0.0005	61.0831	61.0860	-0.0030
U^{dis}	-110.7379	-110.5482	-0.1898	-110.4638	-110.4495	-0.0143	-110.2402	-110.2422	0.0020
$U^{\text{elec}}(l=0)$	-16.5099	-16.5100	0.0001	-16.8940	-16.8941	0.0001	-16.8975	-16.8974	-0.0001
$U^{\text{elec}}(l=1)$	5.0240	5.0237	0.0003	5.5374	5.5371	0.0003	5.6111	5.6109	0.0003
$U^{\text{elec}}(l>0)$	-24.0498	-24.0510	0.0012	-24.4591	-24.4598	0.0007	-24.4661	-24.4671	0.0010
U^{inter}	-90.3861	-90.1969	-0.1892	-90.6001	-90.5861	-0.0140	-90.5207	-90.5206	-0.0001
	$r_c^{\text{real}} = 15$ Å, $\zeta_s = 1$			$r_c^{\text{real}} = 15$ Å, $\zeta_s = 0.1$			$r_c^{\text{real}} = 15$ Å, $\zeta_s = 0.01$		
U^{rep}	60.5289	60.5278	0.0011	60.5721	60.5711	0.0010	60.4791	60.4813	-0.0021
U^{dis}	-109.4833	-109.2946	-0.1887	-109.5332	-109.5167	-0.0164	-109.4278	-109.4292	0.0014
$U^{\text{elec}}(l=0)$	-16.6268	-16.6274	0.0006	-16.6215	-16.6222	0.0007	-16.6188	-16.6193	0.0005
$U^{\text{elec}}(l=1)$	5.3371	5.3374	-0.0003	5.3131	5.3135	-0.0003	5.3448	5.3451	-0.0003
$U^{\text{elec}}(l>0)$	-24.3552	-24.3544	-0.0008	-24.3454	-24.3445	-0.0009	-24.3350	-24.3346	-0.0004
U^{inter}	-89.9364	-89.7486	-0.1878	-89.9280	-89.9124	-0.0156	-89.9025	-89.9018	-0.0006
	$r_c^{\text{real}} = 20$ Å, $\zeta_s = 1$			$r_c^{\text{real}} = 20$ Å, $\zeta_s = 0,.1$			$r_c^{\text{real}} = 20$ Å, $\zeta_s = 0.01$		
U^{rep}	60.9397	60.9411	-0.0014	61.2177	61.2170	0.0008	-110.2583	-110.2552	-0.0031
U^{dis}	-110.6974	-110.5095	-0.1879	-110.4643	-110.4488	-0.0155	61.1137	61.1117	0.0020
$U^{\text{elec}}(l=0)$	-16.5564	-16.5561	-0.0003	-16.8941	-16.8943	0.0002	-16.9300	-16.9303	0.0003
$U^{\text{elec}}(l=1)$	5.0904	5.0901	0.0003	5.5374	5.5370	0.0003	5.6469	5.6466	0.0003
$U^{\text{elec}}(l>0)$	-24.0982	-24.0992	0.0010	-24.4593	-24.4597	0.0004	-24.5032	-24.5032	0.0000
U^{inter}	-90.4124	-90.2238	-0.1886	-90.6001	-90.5858	-0.0143	-90.5779	-90.5770	-0.0008

4.3 Application of free energy workflow

The goal of this section is to illustrate the use of our workflow in refining structures generated in CSP studies and pave the way for the sensitivity analysis to be presented in Section 4.4. For a given set of algorithmic parameters, we perform base calculations to evaluate the vibrational free energy of TCNE polymorphs. Initially, we assess whether performing flexible-body lattice energy minimisation using DMACRYS/CrystalOptimizer can provide structures that can be identified as local minima by CrystalDynamics This is achieved by evaluating the gradients and the second-derivative matrix of the lattice energy with respect to atomic positions. Next, for structures emerging as minima in the lattice energy landscape, the dynamical matrix is constructed and diagonalised at various $\boldsymbol{k}$ points to construct the dispersion curves, which are used to compute the pDOS. Finally, the vibrational free energy and other thermodynamic properties of the two TCNE polymorphs are calculated across a range of temperatures.

4.3.1 Lattice energy gradients

We turn our focus on assessing whether the application of our free energy workflow satisfies the essential condition of the HA. The HA is based on the assumption that the crystal structure lies at a local minimum in the lattice energy hyper-surface. Given an expansion point that coincides with a local minimum (equilibrium position), all the gradients of the lattice energy with respect to the atomic displacements (or equivalently positions) should have values equal to zero. Gradients of the lattice energy can be also interpreted as forces acting on atoms. The curvature of a function at a given point is associated with the eigenvalues of the Hessian matrix. A positive-definite Hessian matrix indicates a local minimum with respect to all dependent variables. In the case where at least one eigenvalue is negative and the rest are positive, the structure lies at a saddle point instead of a local minimum.

Table 4.4: Parameters values used in CrystalDynamics algorithm for calculations shown in Section 4.3.

Parameter name	Parameter symbol	Default value
real electrostatic cutoff	r_c^{real}	15 Å
reciprocal electrostatic cutoff	k_c^{rec}	1.05 Å^{-1}
repulsion-dispersion cutoff	r_c^{dir}	30 Å
Ewald convergence parameter	α	0.08
combined rank of multipoles moments	l	4
functional/basis set	-	M06/6-31G(d,p)
van der Waals parameterisation	-	FIT
lower, upper and increment of temperature	T_{min}, T_{max} & T_{step}	0 K, 400 K & 20 K
number of points for the GLQ	-	4

We apply our workflow to the two structures matching the cubic and monoclinic experimental forms of TCNE in the CrystalPredictor landscape. Details for the global search and lattice energy refinement for TCNE can be found in Sections 5.1.2 and 5.1.3. The TCNE molecule consists of ten atoms ($N = 10$), which results in $3N - 3 = 27$ DoFs in total. These DoFs correspond to 10 bonds, 9 angles and 8 dihedrals, all of which are treated as variables during the lattice minimisation using CrystalOptimizer. Both the predicted TCNE polymorphs are atomistically minimised using CrystalOptimizer. CrystalOptimizer normally has two termination conditions. The code converges either due to negligible differences in geometry between successive minimisation steps or when all gradients with respect to $\boldsymbol{\theta}$ are less than a user defined threshold value. We have disabled the option to terminate on negligible displacements to ensure a tighter convergence to a local minimum in lattice energy. We have used the default threshold value of 1.14 kJ mol^{-1} rad^{-1} for angles and dihedrals and 1.14 kJ mol^{-1}Å^{-1} for bonds as the termination criterion for convergence based on satisfactory gradients.

The next step of our free energy workflow involves the application of the CrystalDynamics algorithm. The utilised algorithmic parameters are shown in Table 4.4 and are consistent with the parameters used in the lattice enrgy minimisation carried out with CrystalOptimizer. Contrary to CrystalOptimizer, CrystalDynamics uses Cartesian coordinates instead of conformational degrees of freedom ($\boldsymbol{\theta}$). For a crystal containing Z molecules in the unit cell and N atoms per molecule, there are $3NZ$ lattice energy gradients with respect to atomic positions (1 in each direction). The monoclinic form has $Z = 3$, resulting into 60 gradients with respect to atomic positions. The distribution of the gradients is plotted using a histogram as shown in Figure 4.6 for the monoclinic polymorph of TCNE. The lattice energy gradients with respect to the atomic positions exhibit low absolute values that in all cases do not exceed 1.0 kJ mol^{-1}Å^{-1}. Most of the peaks are very close to zero indicating that most of the atoms are very close to their equilibrium positions. Notably, the distribution of gradients is symmetrical around zero. This symmetrical distribution implies that although the actual values of the gradients have non-zero values, the sum of all gradients is equal to zero indicating a rigid body minimum. This means that the atoms inside molecules are not in equilibrium positions, however the molecular centres of mass lie at a local minimum. As discussed in Section 4.2.1, the inner loop in the lattice energy minimisation algorithm is performed with DMACRYS as accessed by CrystalOptimizer [18]. The distribution of the gradients demonstrates that the inner loop minimisation process has successfully converged.

The second requirement for identifying a local minimum is the existence of a positive definite Hessian matrix. The Hessian matrices of the CrystalOptimizer minimised structures are evaluated using CrystalDynamics, with eigenvalues computed using the linear algebra package (LAPACK) library [107] for FORTRAN. The minimised structures exhibit Hessian matrices with positive eigenvalues, confirming that they lie at a local minimum. Alternatively, the curvature at the local minimum can be identified by calculating the phonon frequencies at $\boldsymbol{k} = \boldsymbol{0}$, where the exponential phase factor is equal to one and the dynamical matrix reduces to the mass-weighted Hessian matrix (see Equation (3.2.14)). Since the atomic masses always have

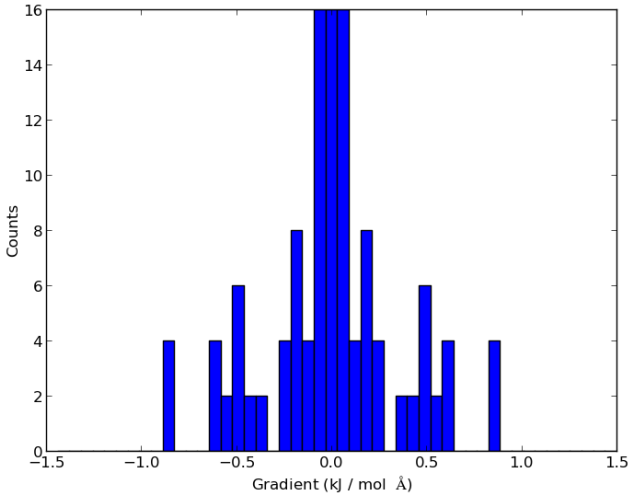


Figure 4.6: Histogram of lattice energy gradients with respect to atomic positions per each direction for the monoclinic form of TCNE (Z=2, N=10) minimised with CrystalOptimizer.

positive values, the eigenvalues of the Hessian matrix share the same sign with the eigenvalues of the dynamical matrix at $\boldsymbol{k} = \mathbf{0}$. Therefore stable structures should have dispersion curves with strictly positive frequencies. We note that Crystal structures of flexible molecules, such as glycine and have been minimised with CrystalOptimizer but did not display positive phonon frequencies. These negative frequencies indicate an unstable structure, which can be attributed to the inconsistencies between the two energy models. The dispersion curves describe the phonon relation with respect to the reciprocal space vector and are discussed in the next section.

4.3.2 Dispersion curves & pDOS

The dynamical matrix is constructed and diagonalised at various $\boldsymbol{k}$-points to calculate the dependence of phonon frequencies on the reciprocal space vector. The frequencies of the $3NZ$ modes of vibration as a function of reciprocal space vector are known as the dispersion curves. The dispersion curves and free energies for TCNE polymorphs were calculated by Vasileiadis [14] and compared with available experimental data. Here, we re-evaluate the dispersion curves to generate the phonon density of states and other thermodynamic properties by employing our free energy workflow that considers molecular flexibility during the minimisation of lattice energy.

Figure 4.7 shows the dispersion curves for the low frequency modes along the three components of the reciprocal space vector. As discussed at the end of Section 4.3.1 the phonon frequencies are related to the eigenvalues of the Hessian matrix at $\boldsymbol{k} = \mathbf{0}$. All phonon frequencies display positive values at $\boldsymbol{k} = \mathbf{0}$, confirming that the optimised structure is a local minima in lattice energy. In addition, the phonon frequencies are positive for all sampled $\boldsymbol{k}$ points indicating that the structure is stable in the free energy landscape. Note that Figure 4.7 displays only the low frequency modes within 10 THz as these modes are expected to have a more significant

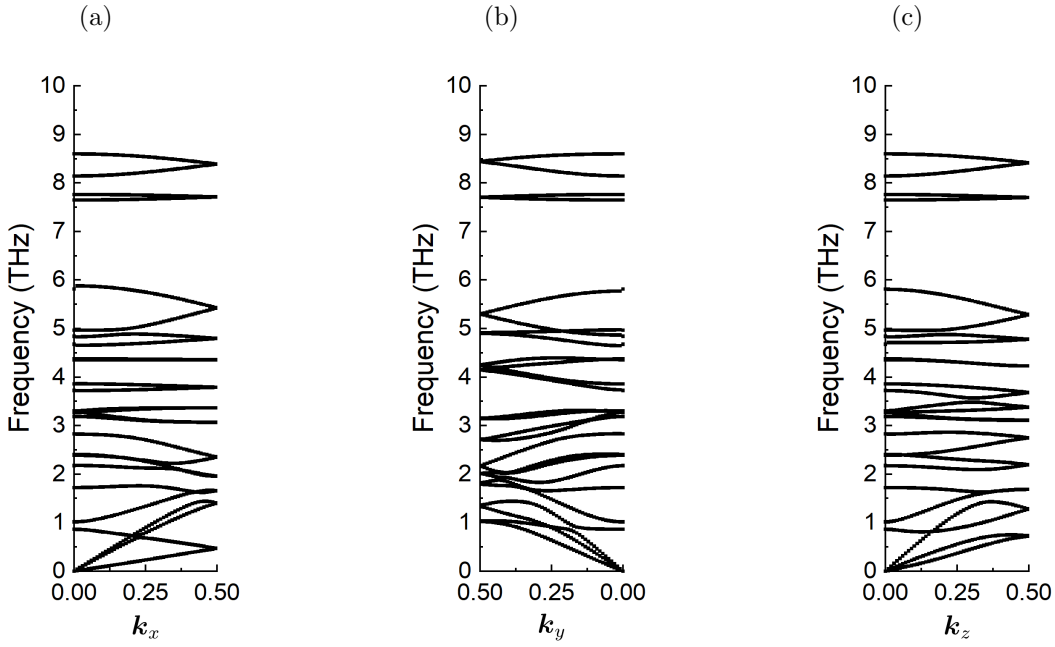


Figure 4.7: Dispersion curves of the monoclinic form describing the relation of phonon frequencies along each component of the reciprocal space vector (a) $\boldsymbol{k}_x$, (b) $\boldsymbol{k}_y$ and (c) $\boldsymbol{k}_z$ for frequencies less than 10 THz.

impact on vibrational free energy in comparison to their higher frequency counterparts [51]. The decreased effect of high frequency modes on vibrational free energy is attributed to the form of the vibrational free energy expression (see Equation (3.3.4)) which tends to zero with increasing frequency. The dispersion curves along the magnitude of $\boldsymbol{k}$ is shown in Figure 4.8a. Notably, the three lowest eigenvalues of the Hessian matrix have values very close to zero indicating the existence of acoustic modes. An acoustic mode of vibration has zero phonon frequency at $\boldsymbol{k} = \mathbf{0}$ and displays an almost linear dependence on each component of the $\boldsymbol{k}$ vector. These acoustic modes correspond to the lowest frequency modes of vibrations arising from all atoms vibrating with the same phase. The rest of the modes have non-zero frequency at $\boldsymbol{k} = \mathbf{0}$ and are known as optical modes.

The pDOS is calculated as a histogram of the dispersion curve generated by binning the frequency. The pDOS of the monoclinic form of TCNE is shown in Figure 4.8b. In the region below 1 THz, the pDOS also displays a linear trend, which is attributed to the contributions of the acoustic modes of vibration. At frequencies between 1 and 5.8 THz, the pDOS is more dense followed by the absence of modes of vibration between 5.8 and 7.4 THz. Two distinct peaks can be observed at approximately 7.5 and 8.4 THz. The pDOS calculated with CrystalDynamics (Figure 4.8b) has some qualitative similarities with the pDOS calculated using DFT-d in Schatschneider *et al.* [2]. In that study harmonic phonon frequencies were evaluated at the PBE-TS level of theory by employing the density functional perturbation theory (DFPT) as implemented in CASTEP [108]. Although the DFT-d energy model varies significantly from

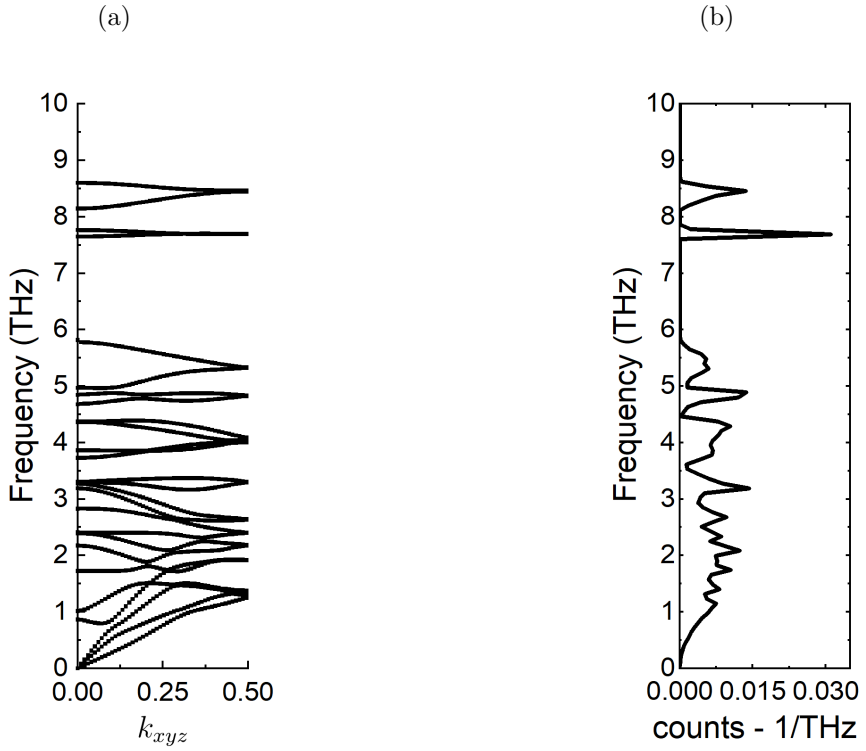


Figure 4.8: (a) Dispersion curves along the magnitude of reciprocal space vector (b) pDOS of the monoclinic form of TCNE calculated as a histogram of (a).

the one used in this thesis, the two calculated phonon density of states exhibit certain common characteristics. In particular, the DFT-d pDOS follows a linear trend in the lowest frequency region similarly to that obtained from CrystalDynamics. In addition, there are no modes of vibration within the frequency range of 5.8 and 7.7 THz, which is in good agreement with the DFT-d pDOS. Knowledge of the pDOS allows the evaluation of thermodynamic properties. The vibrational free energy and other thermodynamic properties calculated for the two TCNE polymorphs are presented in the subsequent section.

4.3.3 Free energy & Thermodynamic properties

We turn our focus on evaluating the Helmholtz free energy and other thermodynamic properties for the two polymorphs of TCNE. As discussed in Section 4.3.1, our free energy workflow is applied to the two structures generated with CrystalPredictor that are matching the cubic and monoclinic forms of TCNE. The generated structures are "atomistically minimised" using CrystalOptimizer meaning that all bonds, angles and dihedrals are treated as optimisation variables, and the Helmholtz free energies are computed using CrystalDynamics. The evaluation of the triple integral appearing in the free energy expression in Equation (3.3.4) is performed using the Gauss-Legendre quadrature (GLQ) numerical integration method. The relative stability of TCNE polymorphs is calculated as the free energy difference, $\Delta\Delta A$, with respect to the cubic form (see Figure 4.9) as a function of temperature. A positive $\Delta\Delta A$ value, indicates that the

monoclinic form has higher free energy than the cubic form and thus is less stable. In contrast, the monoclinic forms is more stable for negative $\Delta\Delta A$ values. Between 0 K and 280 K, the cubic form has a lower free energy than that of the monoclinic form and thus is predicted to be more stable. At temperatures above 280 K, the monoclinic polymorph becomes the most stable form through an enantiotropic transition. The calculated polymorphic stability and enantiotropic transition temperature are in good agreement with experimental and computational studies as discussed in more detail in Section 5.1.

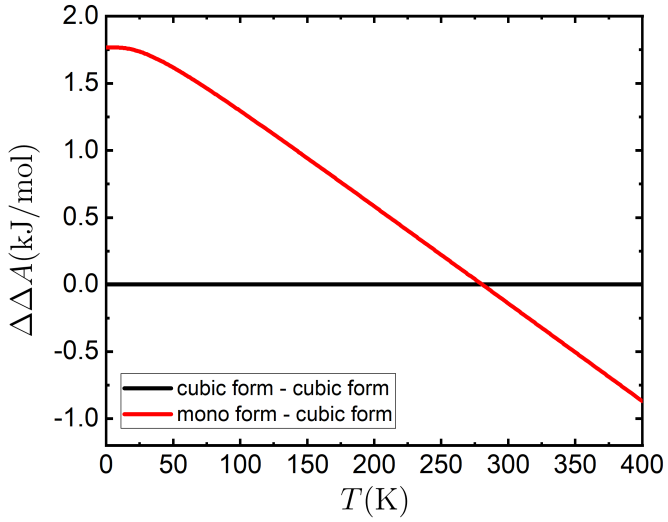


Figure 4.9: Calculated relative Helmholtz free energy between the cubic and monoclinic forms of TCNE, relative to the cubic form, as a function of temperature.

As mentioned in Section 3.3.2, entropy is computed as the negative first derivative of the Helmholtz free energy with respect to temperature at constant volume. In the free energy expression, the entropic contributions are represented as the product of entropy and temperature (TS). Higher entropy results in a lower Helmholtz free energy since both temperature and entropy are always positive, and the entropic term has a negative sign. It is essential to note that this is true in the NVT ensemble, where the number of atoms, volume, and temperature are kept constant. In the NPT ensemble, the Gibbs free energy should be used as the metric to determine stability, taking enthalpic effects into account. In this thesis, calculations are performed using the HA, in which the volume of the unit cell is fixed. The QHA can be employed to account for thermal expansion by performing multiple HA calculations at various unit cell volumes [45, 44], but this is not investigated here. The entropic contributions, calculated for the two TCNE polymorphs, are shown in Figure 4.10. The entropy of the monoclinic form is higher than that of the cubic form over the temperature range between 0 and 400 K. Our calculations are compared with DFT-d calculations performed at the PBE-TS level of theory [2]. Despite the smaller difference obtained in comparison to DFT-d, entropy is seen to be the leading factor in the stabilisation of the monoclinic form at elevated temperatures.

We proceed with our analysis by evaluating the isochoric heat capacity for the two TCNE

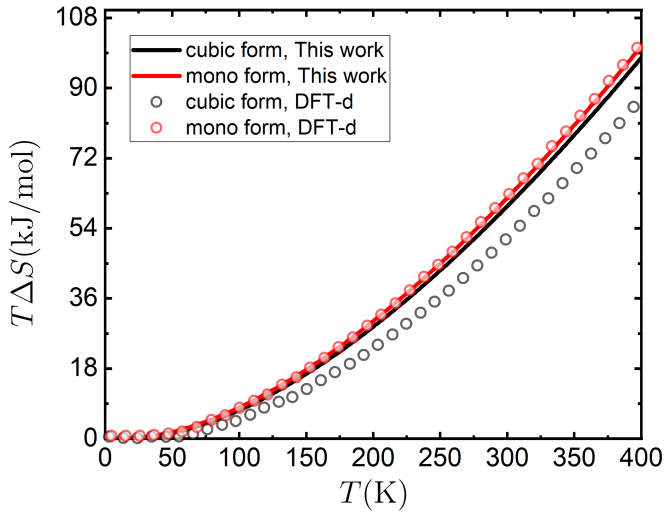


Figure 4.10: Entropic contributions (Entropy multiplied with temperature) of the two polymorphs of TCNE as a function of temperature. Open circles indicate the entropic contributions evaluated using DFT-d at the PBE-TS level of theory [2].

polymorphs. The isochoric heat capacity can be calculated as the derivative of the internal energy with respect to temperature at constant volume using Equation (3.3.8). The calculated heat capacities are shown in Figure 4.11. As expected, minor differences between the computed heat capacities of the two forms are observed. The heat capacities obtained with CrystalDynamics are compared with the heat capacity calculated using the equipartition theorem. Adopting a classical treatment, the internal energy of a system can be calculated using the equipartition theorem. A polyatomic molecule, consisting of N atoms, has $3N$ coordinates that are decomposed into 3 translations, 3 rotations and $3N-6$ vibrations. Translational and rotational degrees of freedom contribute $1/2k_{\mathrm{B}}T$ as they store kinetic energy, whereas vibrations store both kinetic and potential energy contributing, $k_{\mathrm{B}}T$. The number of atoms in a TCNE molecule is equal to 10, so 3 translations and 3 rotations contribute $3k_{\mathrm{B}}T$ collectively and 24 vibrations contribute $24k_{\mathrm{B}}T$. The isochoric heat capacity calculated using the equipartition theorem is thus equal to $27k_{\mathrm{B}} = 0.224$ kJ mol^{-1} K^{-1}. The determination of this uniform value relies on the consideration that translations, rotations, and vibrations are all active in the system. As can be seen in Figure 4.11, the heat capacity calculated using the equipartition theorem is in closer agreement with heat capacities calculated using CrystalDynamics at elevated temperatures.

Thermal energy is computed as the integral of the isochoric heat capacity with respect to temperature (see Equation (3.3.10)). The thermal energy of the TCNE polymorphs is shown in Figure 4.12. Similarly to heat capacity, the thermal energy values computed for each form are fairly similar across the entire temperature range. The thermal energies calculated with CrystalDynamics are compared with DFT-d calculations [2], demonstrating a close agreement for the case of the monoclinic form. The value for the cubic form is slightly overpredicted in comparison to DFT-d. Overall, the impact of thermal energy on polymorphic stability is less

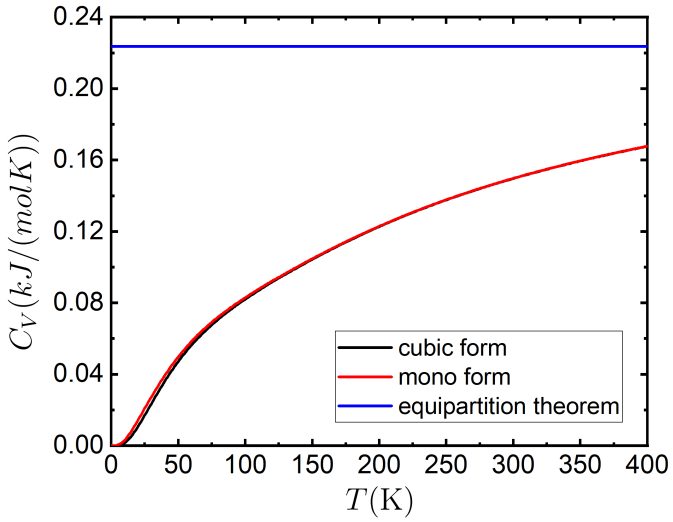


Figure 4.11: Isochoric heat capacity of the two TCNE polymorphs. Blue line represents the heat capacity value calculated using the equipartition theorem.

dominant than the entropy This finding is also in agreement with Nyman [51] who found that differences in the entropy between polymorphs have the biggest impact on stability.

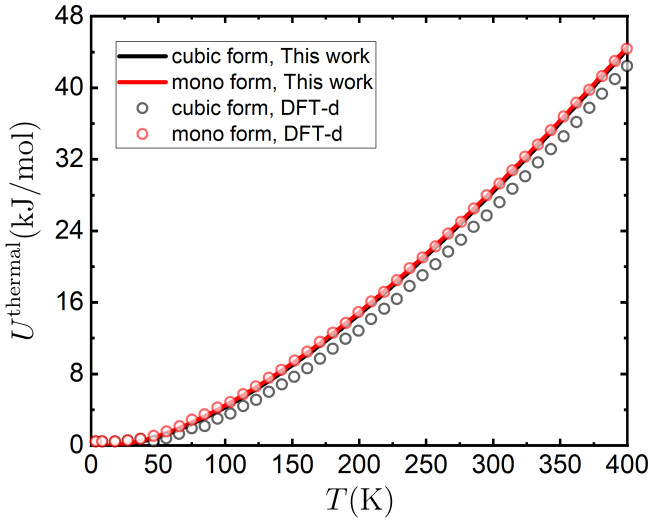


Figure 4.12: Calculated thermal energy for the two TCNE polymorphs. Open circles indicate the thermal energy evaluated using DFT-d at the PBE-TS level of theory [2].

4.4 Sensitivity of the CSP workflow to algorithmic parameters

In the previous section, we demonstrated the successful application of our workflow to the two polymorphs of TCNE, which is a fairly rigid molecule. In this section, we apply our free energy workflow to flexible molecules. This is the first time that CrystalDynamics is used to evaluate

the free energy landscape of a molecule that exhibits flexibility. In the work of Vasileiadis [14], the free energy has been evaluated for rigid structures minimised using DMACRYS [39]. Here, we examine how the number and type of DoFs treated as flexible during the minimisation influence the computed stability of structures on the free energy landscape. We further investigate the effect of altering the finite difference step size in calculating the free energies. The workflow is applied to the 100 lowest energy crystal structures of the Benzoic Acid CrystalPredictor landscape using two minimisation schemes that employ different treatments of molecular flexibility. Lastly, we investigate the impact of performing the lattice energy minimisation, with or without utilising LAMs, on the calculated free energies of Target XXIX. Further details of the CSP studies conducted for Benzoic Acid and Target XXIX are shown in sections 5.2 and and 5.3.

4.4.1 Number of flexible DoFs during minimisation

The impact of the number of flexible DoFs is assessed by considering two minimisation schemes. A minimisation scheme in which all the dihedrals (all torsions) are treated as flexible, as well as a fully atomistic minimisation (all DoFs) scheme are utilised to evaluate the number of structures emerging as stable in the free energy landscape. The 100 lowest energy structures of the CrystalPredictor landscape are locally minimised with CSO-FM employing both the all torsions and atomistic schemes. Following the minimisation, the vibrational free energy is calculated for all structures using CrystalDynamics. The Hessian matrices of all the CSO-FM optimised structures are evaluated to determine whether the minimised structures represent local minima in lattice energy. Among the structures that emerge as local minima, the dynamical matrix is sampled at various $\boldsymbol{k}$ points to evaluate the dispersion curves. Structures that exhibit positive phonon frequencies for each sampled $\boldsymbol{k}$ point are considered stable in the free energy landscape. In contrast, structures with negative eigenvalues of the dynamical matrix at any sampled point are considered unstable. We distinguish between two types of unstable structures. Unstable structures with a negative definite Hessian do not correspond to "true" local minima indicating a saddle point rather than a minimum in the lattice energy surface. On the other hand, structures with a positive definite Hessian matrix that have negative phonon frequencies at any other $\boldsymbol{k}$ point correspond to local lattice energy minima that are not stable in the free energy landscape.

The Benzoic Acid molecule consists of 15 atoms yielding 12 torsions that are treated as flexible in the all torsions minimisation scheme. In the atomistic scheme, a total of 39 conformational DoFs (14 bonds, 13 angles, 12 torsions) were optimised. Quantum mechanical calculations were performed at the PBE/6-31G(d,p) level of theory alongside FIT parametrisation of the Buckingham potential. A step size of 0.001 rad was used to calculate the energy derivatives with respect to dihedrals or angles and a step size of 10^{-5} Å was used for the bonds. These sizes are the default steps sizes used in the CSO-FM input file and are routinely used in our CSP studies. The $RMSD_{15}$ between the experimental (BENZAC02) and the structure minimised with all torsions treated as flexible is equal to 0.222 Å and the $RMSD_1$ is 0.043 Å. Similarly, the

values of $RMSD_{15}$ and $RMSD_1$ for the structure atomistically minimised are 0.209 Å and 0.048 Å. It can seen that the experimental structure is reproduced well by both schemes.

Our findings are summarised in Table 4.5. Out of the 100 structures, 56 emerge as stable using the all torsions minimisation scheme, whilst 80 structures are stable using the atomistic treatment. The all-torsion minimisation scheme leads to 30 structures that are not local minima in the free energy landscape. This number is greatly reduced in the atomistic scheme, to 8. It is evident that the atomistic treatment greatly increases the number of stable structures emerging as stable in the free energy landscape. This increase is attributed to the lower number of unstable structures with negative definite Hessian matrix for the case of atomistic scheme.

The number of structures identified as minima that are unstable in the free energy landscape are 13 and 10 for the cases of all torsions and atomistic schemes, respectively. It is worth comparing a structure that is stable with once scheme and unstable with the other. Of particular interest, the predicted structure matching the experimental form exhibits positive phonon frequencies when subjected to the atomistic scheme but is found unstable in the all torsions scheme. Although the Hessians of both structures are positive-definite, the structure minimised with all torsions has negative phonon frequencies at the 4^{th} sampled $\boldsymbol{k}$ point. A quantification of the geometrical differences between the structures optimised with both schemes is performed. The $RMSD_{15}$ value is found to be 0.045 Å, and the $RMSD_1$ value is determined to be 0.016 Å, illustrating that the geometries obtained by the two schemes are very close. This indicates the high sensitivity of the method and the need of an accurate minimisation model.

Table 4.5: Impact of the number of flexible DoFs on the stability of the 100 lowest energy structures of Benzoic Acid on the CSO-FM landscape

Flexible DoFs	Stable structures	Unstable structures with negative definite Hessian	Unstable structures with positive definite Hessian
all torsions	57	30	13
all DoFs	82	8	10

4.4.2 Finite difference step size & LAMs

In the previous sections we have established that the atomistic minimisation provides a greater number of stable structures in the landscape. Here, we attempt to further increase the number of stable structures by altering the FD step sizes used for the evaluation of the derivatives of lattice energy with respect to conformational DoFs. In previous CSP studies [98, 109] performed with CrystalOptimizer mostly torsions and angles were treated as the DoFs to be treated flexible, while atomistic minimisation schemes were also considered. In the more recent implementation of CSO-FM lattice energy minimisation, the FD step size for bonds has not been tested thoroughly. In this section, we examine whether altering the default value of the bond FD step size has an impact on the number of structures emerging as stable in the free energy landscape. To

make systematic comparisons, the same energy model used in section 4.4.1 is maintained. Local minimisations with CSO-FM are performed for the 100 lowest energy structures of CrystalPredictor landscape of Benzoic Acid. Bonds step sizes of 10^{-4} Å, 10^{-5} Å and 10^{-6} Å are utilised to make comparisons. The number of stable and unstable structures is given in Table 4.6. The use of different bond step sizes has a small impact on calculated lattice and free energies with differences being below 0.5 and 0.3 kJ/mol, respectively. The number of stable structures with positive phonon frequencies is very similar for all used FD bond steps. An additional structure emerges as stable using the shortest bond step of 10^{-6} Å. This structure is identified as local energy minima by all bond FD step sizes, however it displays positive phonon frequencies for all sampled points in the reciprocal space only in the case of the shortest bond size.

Table 4.6: Effect of altering the FD step size for bonds in the landscape of Benzoic Acid.

FD step size (Å) for bonds	Stable structures	Unstable structures with negative definite Hessian	Unstable structures with positive Hessian
10^{-4}	82	8	10
10^{-5}	82	8	10
10^{-6}	83	8	9

In addition to FD step size for bonds, we assess the sensitivity of the CSP workflow to the utilisation of LAMs during the lattice energy minimisation. As discussed in Section 4.2.1, LAMs are employed to increase the efficiency of the lattice energy minimisation algorithm by reducing the number of quantum mechanical calls. For each converged minimisation, a single point isolated-molecule calculation is conducted neglecting the influence of nearby LAMs. In this case, it is expected that the impact of using LAMs will be minimal. The effect of the utilisation of LAMs on the number of structures emerging as stable structures in the free energy landscape is performed for Target XXIX from the 7^{th} CSP blind test. More details regarding the generation of the landscape, lattice energy refinement and free energy evaluation can be found in Section 5.3. The 100 lowest energy structures from the CrystalPredictor landscape of Target XXIX are minimised both with and without employing LAMs. The minimisations are carried out with CSO-FM treating all torsions as the flexible DoFs. A summary of the calculations, including the breakdown of unstable structures, is provided in Table 4.7. The number of stable structures with positive phonon frequencies is similar in both cases. Overall, the calculations can be accelerated by employing LAMs with negligible impact on the calculated free energies.

Table 4.7: Effect of using LAMs on the number of stable structures in the landscape of Target XXIX.

LAMs used	Stable structures	Unstable structures with negative definite Hessian	Unstable structures with positive Hessian
Yes	81	9	10
No	82	8	10

4.5 Conclusions

In this chapter, we have initially presented the CrystalDynamics algorithm, which evaluates the Helmholtz free energy of molecular crystals. The algorithm employs the HA of LD theory to calculate the vibrational contributions to free energy and is based on the mathematical formulation described in Chapter 3. The proposed workflow for performing vibrational free energy calculations consists of lattice energy minimisation followed by free energy evaluation using a consistent energy model. We have assessed the performance of different lattice energy minimisation algorithms to evaluate the Helmholtz free energy of crystals comprised of rigid or flexible molecules. For rigid molecules, DMACRYS/CrystalOptimizer can provide structures that emerge as stable. Conversely for flexible molecules, lattice energy minimisation with CSO-FM/CSO-RM is required to provide structures that display phonons with positive frequencies. This is the first time that the CrystalDynamics algorithm has been applied to evaluate the free energy of crystals that exhibit molecular flexibility.

The consistency between the calculated lattice energy contributions across across energy minimisation and lattice dynamics has been investigated by varying the direct cutoff and spline fraction values. We have found that intermolecular energy contributions are in good agreement for lower spline fractions, while altering the direct space cutoff has a relatively low impact on the differences between the calculated energies. Having established a satisfactory consistency, we proceeded to validate the essential condition behind the HA. In the HA, the structure must lie at a local minimum in lattice energy therefore having zero gradients and a positive definite Hessian matrix. The calculated lattice energy gradients with respect to atomic positions have been found to exhibit low values and to be within the threshold used as the convergence termination criterion indicating that the lattice energy minimisation has successfully converged. In addition, the gradients have been found to yield a symmetrical distribution and their total sum is equal to zero suggesting that although some atoms are not converged tightly, the centre of mass of each molecule lies at an equilibrium position. This equilibrium position is identified as a rigid-body local minimum, as it displays a positive definite Hessian matrix of lattice energy with respect to atomic positions.

The successful application of the workflow to a rigid molecule has been illustrated for the two TCNE polymorphs. For both polymorphs, the dynamical matrix has been constructed

and diagonalised at various $\boldsymbol{k}$ points to evaluate the phonon frequencies as the square roots of the eigenvalues. All calculated eigenvalues of the dynamical matrix, across all sampled $\boldsymbol{k}$ points, were found to exhibit positive values demonstrating that structures are stable in the free energy landscape. The phonon frequency of each mode of vibration along the reciprocal space vector has been displayed in the dispersion curves. The pDOS, calculated as a histogram of the dispersion curves, has been used in evaluating the vibrational free energy and other thermodynamic properties. The pDOS of the monoclinic form has been compared with the pDOS obtained from DFT-d calculations. Some distinct characteristics between the pDOS calculated with CrystalDynamics and the one with DFT-d are found to be similar. However, the differences in energy models and sampling procedures allowed only a qualitative comparison. The free energy evaluation of TCNE polymorphs has demonstrated that our workflow predicts the enantiotropic transition from the cubic to the monoclinic form. The LD approach adopted in this thesis allows the evaluation of thermodynamic properties, namely entropy, heat capacity and thermal energy. Our calculations suggest that the entropic contributions have a more significant impact on polymorphic stability in comparison to thermal energy. This finding is in agreement with the DFT-d calculation conducted for the same molecule as well as other studies that have evaluated the entropic contributions of polymorphic molecules.

The evaluation of the eigenvalues of the Hessian matrix and the dynamical matrix at different $\boldsymbol{k}$ points allows to distinguish between structures that are not local minima and structures that although lie in local minima are unstable in the free energy landscape as they exhibit negative phonon frequencies. Two minimisation schemes, differing in the number and types of degrees of freedom treated as flexible, have been examined. The atomistic minimisation resulted in a greater number of structures that emerged as stable in the free energy landscape compared to the all-torsions scheme. Additionally, the sensitivity to algorithmic parameters has been assessed by altering the bond FD step size in order to obtain the value that produces the larger number of stable structures. Finally, the usage of LAMs during lattice energy minimisation is shown to have minimal effect on the obtained phonon frequencies. In the next chapter, we present our findings for the CSP studies conducted for three molecules of increasing size and flexibility.

Chapter 5

Free energy for refining CSP landscapes

We perform full CSP studies and apply our workflow for evaluating free energies on three molecules with increasing molecular flexibility and size. Initially, we evaluate the vibrational free energy of tetracyanoethylene (TCNE). Our findings are compared with experimental and computational data in the literature as well as a previous CSP study that has adopted a rigid-molecule minimisation scheme [14]. Next, we apply our workflow to Benzoic Acid, a molecule for which only one crystal structure has been reported. The goal of this study, which is the first application of our LD method to a molecule that exhibits conformational flexibility, is to assess the suitability our of free energy workflow in calculating the free energy landscape of a non-rigid molecule. Finally, we present our calculations for Target XXIX, which was part of the CSP study performed for 7^{th} CCDC blind test. This constitutes a greater challenge for LD due to the greater degree of molecular flexibility. The presented CSP studies were performed to evaluate the feasibility of incorporating the proposed workflow in our CSP multistage methodology. Additionally, we assessed the performance of our approach in ranking CSP landscapes across a range of temperatures using free energy as the stability metric. Our findings demonstrate that our proposed CSP workflow is applicable to a variety of molecular systems and highlight the importance of including thermal contributions to lattice energy.

To quantify the impact of the temperature on the ranks, we employed a statistical analysis by calculating Spearman's rank correlation coefficient [110] for the free energies landscapes at 5 temperatures (0, 100, 200, 300, and 400 K). Spearman's rank correlation coefficient, R_S, is a statistical measure that assesses how well the relation between the two sets of variables can be described by a monotonic equation regardless of whether the relationship is linear. The Spearman rank correlation coefficient is computed by first ranking the values of each variable, then calculating the Pearson correlation coefficient on the ranks given by

$$R_S = 1 - \frac{6 \sum D_S^2}{N_S \left(N_S^2 - 1\right)}, \tag{5.0.1}$$

where D_S is the difference between the rank at the considered temperature and 0 K and N_S is the number of points in the dataset, which in our case is equal to the number of structures that emerge as stable in the free energy landscape. Pearson's correlation coefficient describes how well the relation between the two variables can be represented by a linear equation. Pearson correlation coefficient values range from -1 to +1. The lowest and highest coefficient values

indicate that the two examined variables lie on a perfect line with positive or negative slope, respectively. Coefficient values close to 0 correspond to no linear dependency between the variables. Spearman's rank correlation coefficient assesses the monotonicity of the variables by evaluating the linear relationship of their ranks at different temperatures. Similarly to Pearson coefficient, Spearman's rank correlation coefficient also ranges between -1 and 1. Absolute values at 1 indicate a perfect monotonic relationship, whereas a value close to 0 indicates no monotonic relationship between the variables in the dataset.

5.1 Tetracyanoethylene CSP study

In Chapter 4, we have shown the application of our free energy workflow to the two polymorphs of TCNE, with a chemical formula C_6N_4. In addition, we have examined the workflow's sensitivity in altering the FD step sizes for bonds and the usage for LAMs. In this section, we present our full CSP study on TCNE molecule that involves global search and two refinement stages. The structure of the CSP study for TCNE is as follows. We begin with providing a brief literature review on the experimental and computational studies of TCNE crystals. Next, we describe the candidate generation stage performed with CrystalPredictor. In the last two sections of this chapter, we refine the landscape generated in the global search stage using lattice energy and free energy as stability metrics.

TCNE is a planar and rigid molecule used in charge transport carrier applications, such as organic conductors [111] and molecular magnets [112]. The molecular connectivity diagram of TCNE is illustrated in Figure 5.1. Four polymorphs of TCNE have been reported, however only the structures of the two stable forms has been resolved. The stable polymorphs are the cubic form [113] and the monoclinic form [114] (see Figure 5.2). A third polymorph, commonly known as form III, is considered metastable and can be observed as an intermediate phase during the monoclinic to cubic transition under the application of pressure [115]. Another metastable polymorph, form IV, is reported to be obtained by methylene chloride/ethyl acetate solution at room temperature [2]. To the best of our knowledge, there have not been studies describing or reproducing the structure of form IV. The TCNE crystal structures found in the CSD are deposited under the RefCodes family TCYETY. The cubic form has eight molecules in the unit cell (Z=8) with RefCodes: TCYETY01, TCYETY02, TCYETY12. The monoclinic form has two molecules in the unit cell (Z=2) and is deposited under RefCodes: TCYETY, TCYETY03, TCYETY04 and TCYETY11.

The stability relation between the cubic and monoclinic forms, as a function of temperature and pressure, has been examined thoroughly in experimental studies. Murgich and Pissanetzky [116] reported a reversible transition from the cubic to the monoclinic form at 292±0.5 K using the nuclear quadrupole resonance technique. On the other hand, an irreversible transition from the cubic at low temperatures to the monoclinic form at high temperatures was found to occur at 326±2 K and 318±5 K using differential scanning calorimetry (DSC) and neutron diffraction, respectively [117]. These slight differences in the reported values can be attributed to the

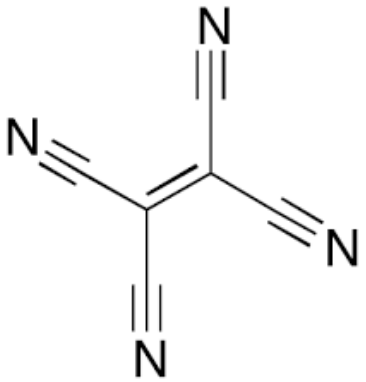


Figure 5.1: Molecular connectivity diagram of TCNE.

dependency of the transition temperature on the heating rate used in DSC experiments. This transition was found to be irreversible as the monoclinic form remained stable upon cooling down to 5 K [118]. The effect of pressure on polymorphic stability has also been examined. The monoclinic form undergoes a transition to the cubic form via the intermediate form III at pressures higher than 5 GPa, whereas the cubic form does not undergo a transition for pressures up to 10.8 GPa [117]. The influence of pressure is less pronounced on the cubic form due to its smaller volume when compared to the monoclinic form.

In the published literature, the stability of TCNE polymorphs has also been investigated computationally. Chaplot *et al.* have performed free energy calculations using the HA [118] and QHA [119], adopting a semi-rigid treatment that allows for the evaluation of the low frequency modes of vibration. Their force field considered only the repulsion-dispersion interactions, modelled using the Buckingham potential, without taking into account any electrostatic contributions. Schatschneider *et al.* [2] have calculated harmonic phonon frequencies at the PBE-TS level of theory, using the DFPT [120]. Calculations performed with DFT-d suggested that the enantiotropic transition from the cubic to the monoclinic form occurs at 160 K. A CSP study for TCNE has been conducted by Vasileiadis [14]. In that study, the molecular conformation was obtained from unconstrained isolated-molecule geometry optimisation and an rigid body minimisation was conducted with DMACRYS [39]. The free energy were calculated at the HF/6-31G(d,p) and M06/6-31G(d,p). In the former case, the cubic form was identified as the global minimum in both the lattice energy and free energy landscapes. The monoclinic form as the 2^{nd} lowest energy structure at 0 K. The enantiotropic transition from the cubic to the monoclinic form was found to occur at approximately 74 K. The free energies calculated at the M06/6-31G(d,p) level of theory ranked the cubic and monoclinic forms, 3^{rd} and 4^{th}, respectively. The monoclinic form became more stable than the cubic form at 79.5 K, whereas above 276 K an undiscovered structure becomes the most stable.

Overall, it is generally agreed that the cubic form is more stable than the monoclinic form at low temperatures and that the order of polymorphic stability changes with increasing temperature. The enantiotropic transition has been reported to occur anywhere between 5 and 328 K. The effect of pressure in polymorphic stability is expected to be negligible at pressures below

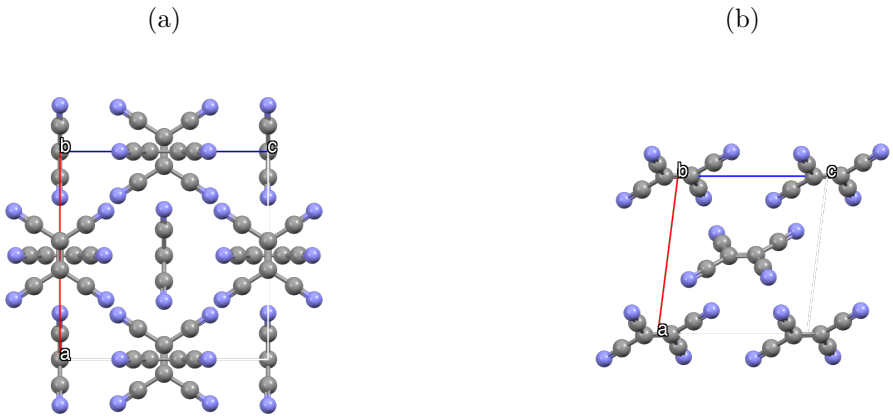


Figure 5.2: Unit cells of (a) cubic form and (b) monoclinic TCNE polymorphs.

1 GPa. All the computational studies concurred that the enantiotropic transition is caused by the higher entropy of the monoclinic form at elevated temperatures. In the section below, we carry out full CSP studies, involving candidate generation with CrystalPredictor, lattice energy minimisation with CrystalOptimizer and free energy evaluation with CrystalDynamics. The goal is to perform lattice energy minimisation accounting for molecular flexibility and calculate free energy using a consistent energy model. Although the impact of flexibility is expected to be small for the rigid TCNE case, we apply our CSP workflow to assess differences in calculated free energies and rankings adopting rigid and flexible treatments.

5.1.1 Global Search

The first step of the CSP multistage approach, discussed in chapter 2, involves determining the conformational degrees of freedom that will be treated as flexible in the candidate generation stage. For TCNE, a fully rigid treatment is employed throughout the global search stage. Adopting a fixed molecular geometry can be justified by the planar structure of TCNE and the presence of triple bonds between the carbon and nitrogen atoms (C≡N). Adopting a rigid treatment results in omitting the intramolecular contribution to the lattice energy throughout the candidate generation stage. The conformation used in the global search stage is obtained by performing an unconstrained gas phase geometry optimisation of the TCNE molecule at the PBEPBE/6-31G(d,p) level of theory. The *ab initio* charge density is also calculated as part of the gas phase minimisation. The CHELPG [92] method is utilised for fitting point charges at the above mentioned charge density. A cutoff of 12 Å is used for the electrostatic interactions quantified via the Coulomb potential. The repulsion-dispersion interactions are modelled via the Buckingham potential employing a FIT parametrisation [90] with a cutoff of 15 Å. More details for the global search stage conducted with CrystalPredictor [97, 21, 22, 7] can be found in Section 3.6.

Half a million initial structures with one molecule in the asymmetric unit cell (Z'=1) are

generated using pseudo-random Sobol' point sequencing [19]. The initial crystal structures are minimised using CrystalPredictor's energy model resulting in 498,159 stable crystals. The next step of the search stage involves analysing the stable structures and results into 5,718 unique structures within 20 kJ/mol from the global lattice energy minimum. These 5,718 structures are clustered in order to identify crystallographically-duplicate structures belonging to an equivalent lattice energy minimum. Clustering is performed with COMPACK [121] and further reduces the number of structures from 5,718 to 1,188. The lowest energy structure within each cluster is selected as a representative of that particular cluster. The 1,188 clustered structures within 20 kJ/mol of the global minimum of the CrystalPredictor landscape form the final set of structures generated in the global search stage. The clustered CrystalPredictor landscape is shown in Figure 5.3.

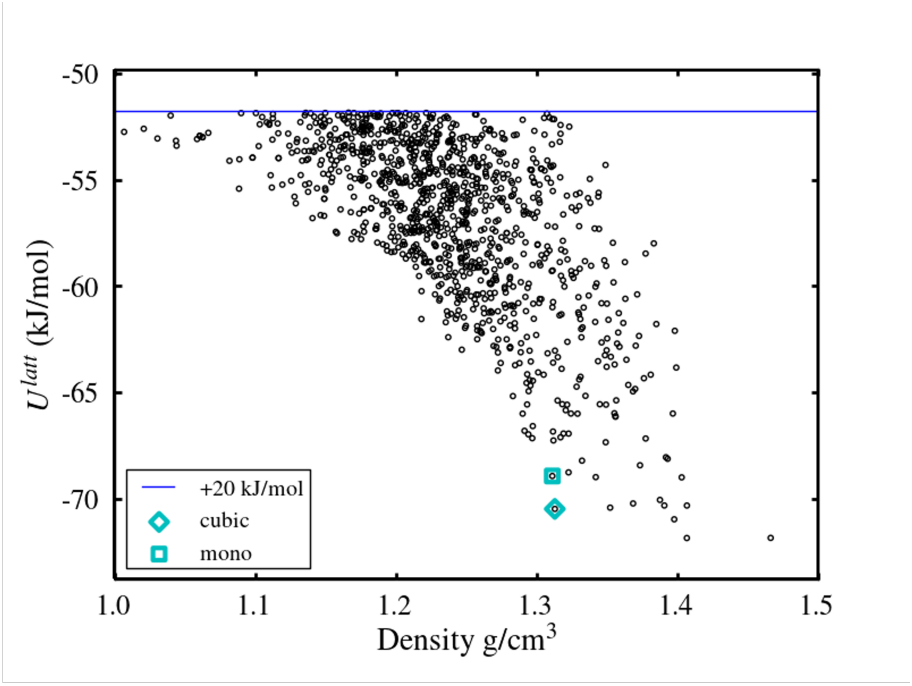


Figure 5.3: CrystalPredictor landscape for TCNE molecule. Open circles indicate the structures obtained at the global search stage after performing clustering. The predicted forms that match the experimental ones are highlighted with a cube and rhombus. The blue line indicates the 20kJ/mol threshold from the global minimum.

Both the experimental forms are predicted as low energy minima in the CrystalPredictor landscape. The 3rd lowest energy structure matches the cubic form and the 11th lowest energy structure the monoclinic form. The similarity matching tool of CCDC Mercury [122, 101] are used to compare the predicted crystals with their counterparts deposited in the CSD (RefCode family TCYETY). The geometry comparison between experimental and predicted crystals is summarised in Table 5.1. Accurate geometries for the two forms are generated as the predicted structures display an average $RMSD_{15}$ below 0.6 Å. The simplistic energy model of CrystalPredictor is able to identify the two experimental forms within just -2.914 kJ/mol from the global

Table 5.1: Geometry comparison of the cubic and monoclinic form generated at the CrystalPredictor stage with the experimental TCNE crystal structures in CSD.

Polymorph	RefCode	R-factor (%)	$RMSD_1$ (Å)	$RMSD_{15}$ (Å)	PXRD
Cubic	TCYETY01	4.80	0.036	0.118	0.924
	TCYETY02	2.70	0.035	0.118	0.925
	TCYETY12	4.26	0.032	0.182	0.851
Monoclinic	TCYETY	8.30	0.057	0.393	0.873
	TCYETY03	3.50	0.050	0.397	0.871
	TCYETY04	2.20	0.054	0.398	0.868
	TCYETY11	3.00	0.053	0.395	0.868

minimum in the lattice energy landscape (-71.777 kJ/mol). In the following sections, low energy structures generated in the global search stage are refined using increasingly sophisticated models and methods.

5.1.2 Refinement stage I - Lattice energy

The next stage of the CSP study involves refining the lowest energy structures obtained in the candidate generation stage using more detailed energy models. The refinement stage I is performed using the CrystalOptimizer algorithm [18], which performs geometry optimisation of molecular crystals by minimising their lattice energy. In total, TCNE has 27 DoFs: 10 bonds, 9 angles and 8 dihedrals. All DoFs are treated as flexible during the lattice energy refinement stage. Note that the terms "atomistic" and "fully flexible" are used interchangeably throughout the text. In the outer loop, finite differences were used for the evaluation of the intermolecular derivatives. The default values for FD step sizes of CrystalOptimizer are used in the global search stage. FD step sizes of 0.03 Å, 0.35° and 0.50° are utilised for bonds, angles and dihedrals, respectively.

Quantum mechanical calculations are employed at the M06/6-31G(d,p) level of theory to populate the intramolecular and intermolecular LAM database. To model electrostatic interactions, a distributed multipole approach is adopted, considering multipoles up to a combined rank of four ($l = 4$). Multipole interactions are calculated with a real space cutoff of 15 Å and a reciprocal space cutoff of 1 $Å^{-1}$. Repulsion-dispersion interactions are calculated with Buckingham potential and FIT parameterisation using a 30 Å cutoff. At the end of each minimisation, a single point energy calculation is performed to obtain the intramolecular and electrostatic gas phase energies without considering the effect of LAMs.

The 1,188 clustered structures within 20 kJ/mol from the global minimum of the CrystalPredictor landscape are refined with CrystalOptimizer [18]. All the locally minimised crystals are then clustered with COMPACK resulting in 837 unique structures that constitute the final ranked set of refinement stage I, shown in Figure 5.4. The ranking of the predicted structures

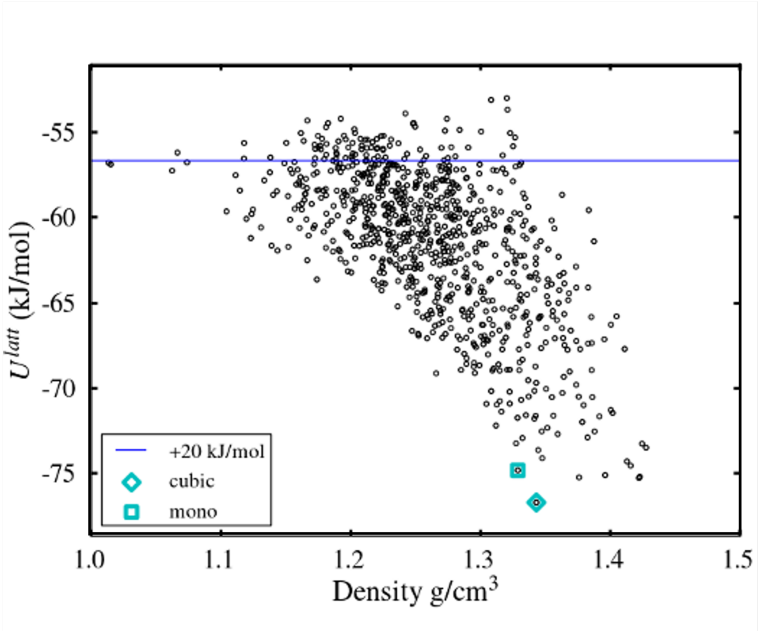


Figure 5.4: CrystalOptimizer landscape. Open circles indicate minimised crystal structures. Structures matching the experimental cubic and monoclinic form are highlighted with a rhombus and a cube symbol. The blue line indicates the 20kJ/mol threshold from the global minimum.

matching the two experimental structures is improved in the CrystalOptimizer landscape. The global minimum of the CrystalOptimizer landscape is identified as the cubic form with an energy of -76.663 kJ/mol. Furthermore, the structure ranked 6$^{\text{th}}$, exhibiting an energy of -74.780 kJ/mol, matches the experimental monoclinic form. The geometry comparison between the structures locally minimised with CrystalOptimizer and experimental forms is performed with CSD python API [122, 101] and can be found in Table 5.2.

Table 5.2: Geometry comparison between the cubic and monoclinic form at refinement stage I performed with CrystalOptimizer.

Polymorph	RefCode	RMSD$_1$ (Å)	RMSD$_{15}$ (Å)	PXRD
Cubic	TCYETY01	0.021	0.066	0.969
	TCYETY02	0.020	0.066	0.969
	TCYETY12	0.022	0.132	0.900
Monoclinic	TCYETY	0.032	0.297	0.909
	TCYETY03	0.026	0.302	0.908
	TCYETY04	0.029	0.301	0.904
	TCYETY11	0.029	0.298	0.905

The energy model of CrystalOptimizer takes into account molecular flexibility and utilises distributed multipoles, instead of point charges, providing more accurate geometry and ranking predictions compared to CrystalPredictor. In particular, the rank of the cubic and monoclinic

forms is improved from 3rd and 11th in the CrystalPredictor to 1st and 6th lowest energy structures in the CrystalOptimizer landscape. Our findings are in good agreement with the above mentioned experimental and computational studies that suggest that the cubic form is the most stable polymorph at 0 K. The monoclinic form is also found at a very low rank and with very similar energies with the rest of the predicted low energy structures just -1.883 kJ/mol above the global minimum in lattice energy. Vasileiadis [14] found the cubic and the monoclinic forms as the 2nd and 6th lowest energy structures, respectively. The improved ranking in our study can be attributed to the flexible treatment and to usage of extended cutoffs for the evaluation of the electrostatics and repulsion-dispersion interactions. In the subsequent stage of our CSP study, we employ vibrational free energy calculations to investigate the transition from the cubic to the monoclinic form. Additionally, we aim to assess the impact of temperature on the TCNE lattice energy landscape.

5.1.3 Refinement stage II - Free energy

The vibrational free energies of the 100 lowest energy structures of the CrystalOptimizer landscape are evaluated using the CrystalDynamics algorithm. The energy model used for the free energy evaluation is consistent with the energy model adopted during the lattice energy refinement stage. Vibrational free energy calculations are performed at the M06/6-31G(d,p) level of theory and with the FIT potential. Out of 100 examined structures, 76 emerged as stable in the free energy landscape as they displayed positive phonon frequencies along all sampled $\boldsymbol{k}$ points. The remaining 24 structures that displayed imaginary frequencies were eliminated from the free energy landscape. In the previous CSP study [14], the lattice energy landscape contained 148 structures (122 with Z'=1, 25 with Z'=2 and 1 with Z'=8) out of which 24% emerged as stable for calculations carried out with the HF functional. In the remainder of this section, we present the vibrational free energies and other thermodynamic properties of all structures that displayed positive phonons frequencies.

At 0 K, the thermal contributions to free energy (entropy and thermal energy) are zero therefore the free energy of a system is equal to the sum of lattice energy and ZPE. The ZPE landscape of TCNE is shown in Figure 5.5a. The difference between the ZPE of the forms matching the experimental structures is equal to 0.361 kJ/mol. This difference is in relatively close agreement with the polymorphic stability study presented in [51], which found that differences in ZPE calculated using rigid-body phonon frequencies between polymorphs do not exceed the 0.33 kJ/mol in the 95% of polymorphs pairs. The calculated ZPE values fall within a narrow range for all stable structures in the landscapes, with the difference between the lowest and highest calculated values being equal to 1.136 kJ/mol. The cubic form has the 4th highest ZPE energy value, however it remains the global minimum of the free energy landscape at 0 K shown in Figure 5.5b. Overall, the free energy landscape at 0 K and the lattice energy landscape have very similar ranking (see Figure 5.5b). This finding is in agreement with the previous CSP study conducted by Vasileiadis [14] that reported minor ranking alteration due to ZPE contributions.

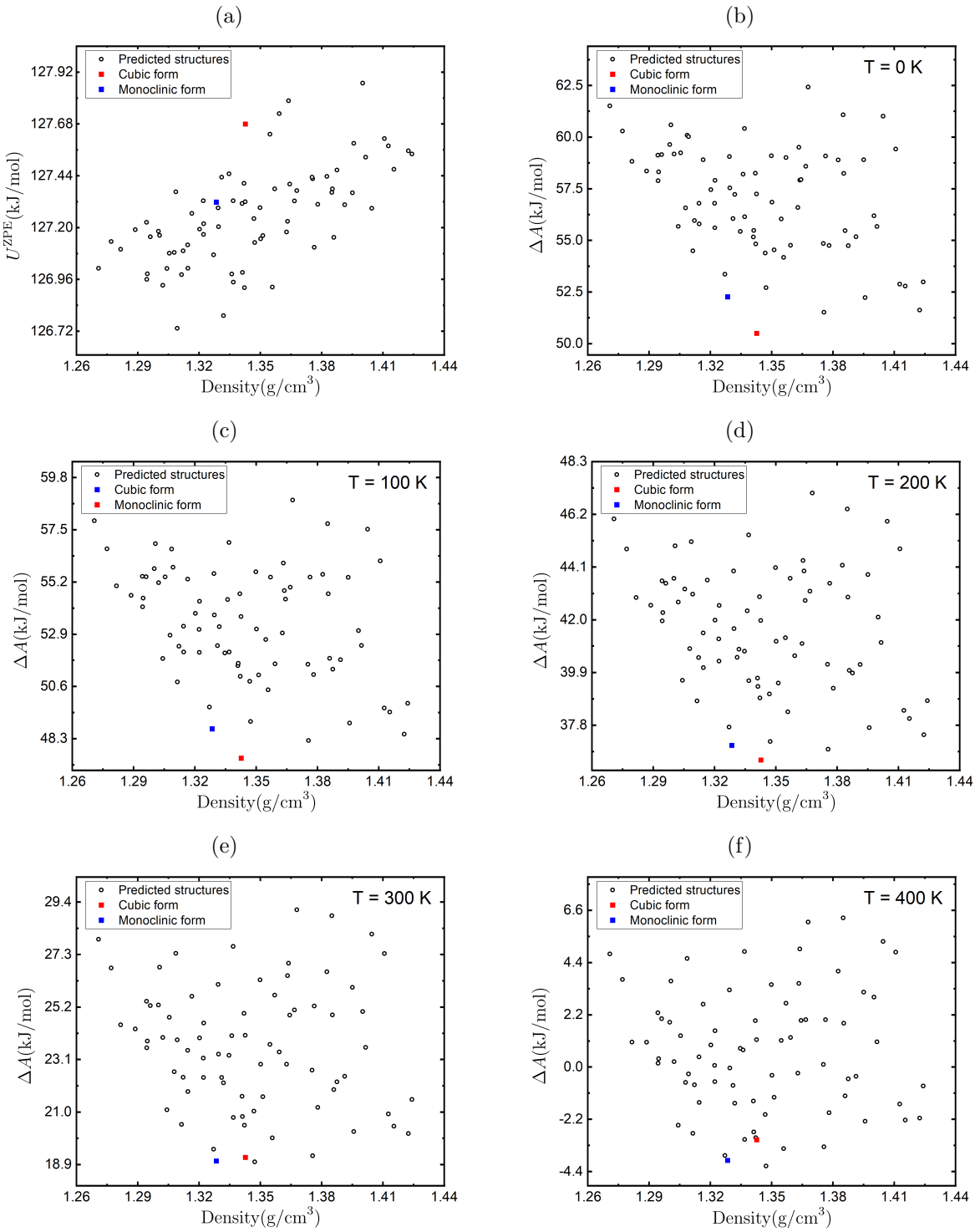


Figure 5.5: (a) ZPE energy landscape of TCNE. Free energy landscape at (b) 0 K, (c) 100 K, (d) 200 K, (e) 300 K and (d) 400 K. Structures shown as red and blue squares represent the cubic and monoclinic forms, respectively.

The relatively small differences in ZPE between predicted structures provide justification for the common approximation to use lattice energy as a metric for ranking stability at 0 K.

The cubic form is identified as the global minimum in free energy at 0 K, (lattice energy and ZPE). The monoclinic form emerges as the 4^{th} lowest free energy structure. Note that the 2^{nd} and 5^{th} lowest energy structures in the lattice energy landscape are unstable and therefore are not considered in the ranking. The free energy landscapes at 0 K, 100 K, 200 K, 300 K and 400 K are shown in Figures 5.5b-5.5f. The cubic form remains the global minimum in free energy between 0 K and ≈275 K indicating that is the most stable structure at lower temperatures in line with experiments and the DFT-d study. At 275 K, an enantiotropic transition occurs, resulting in a change in the stability of the two forms. At 300 K the monoclinic form becomes the global minimum and the cubic form the 3^{rd} lowest energy structure. At 400 K, another predicted structure ranked becomes the global minimum and the monoclinic form becomes the second lowest free energy structure.

We proceed with examining the vibrational contribution that has the biggest impact on the relative stability across different temperatures. As discussed on Chapter 3, the vibrational contributions to free energy are decomposed into ZPE, S (entropy), and U^{thermal} (thermal energy), with the latter two depending on temperature. We evaluate the difference in entropic contributions, $T\Delta S$, calculated relative to the lowest entropy structure in the free energy landscape. For TCNE, the structure with lowest entropy is the structure ranked 35^{th} in the CrystalOptimizer landscape The calculated $T\Delta S$ values at 100 K, 200 K, 300 K and 400 K are shown in Figures 5.6a-5.6d. Overall, the entropic contributions become more significant with increasing temperature. At 100 K, the difference between the lowest and biggest calculated $T\Delta S$ among all structures in the landscape is 1.756 kJ/mol. At 400 K, that difference increases to 8.883 kJ/mol, indicating the increased effect of temperature on entropic contributions.

In general, the lower density structures have bigger entropic contributions than their higher density counterparts. This finding is in agreeent with the previous study conducted for TCNE [14] as well as other studies in the literature [17, 123] that have reported that vibrational entropy favours the stabilisation of less dense structures. The cubic form of TCNE has the 3^{rd} at 100 K and 4^{th} lowest entropy values above 200 K. This low entropy leads to the destabilisation of the cubic form with increasing temperature. As a result, the difference in free energy between the cubic and monoclinic forms decreases with increasing temperature, yielding the enantiotropic transition at 275 K.

The other temperature temperature-dependent vibrational contribution is U^{thermal}. Similarly, the difference in thermal energy, $\Delta U^{\text{thermal}}$, is calculated relative to the structure having the lowest U^{thermal} in the free energy landscape. The $\Delta U^{\text{thermal}}$ values for the 76 stable structures are shown in Figure 5.6e at 100 K and in Figure 5.6e at 400 K. The rank of entropic term and thermal energy does not change with increasing temperature. The magnitude of the above mentioned quantities, however changes with increasing temperature yielding changes in relative stability and the thermal energy is found to have a lower impact on relative stability than entropy.

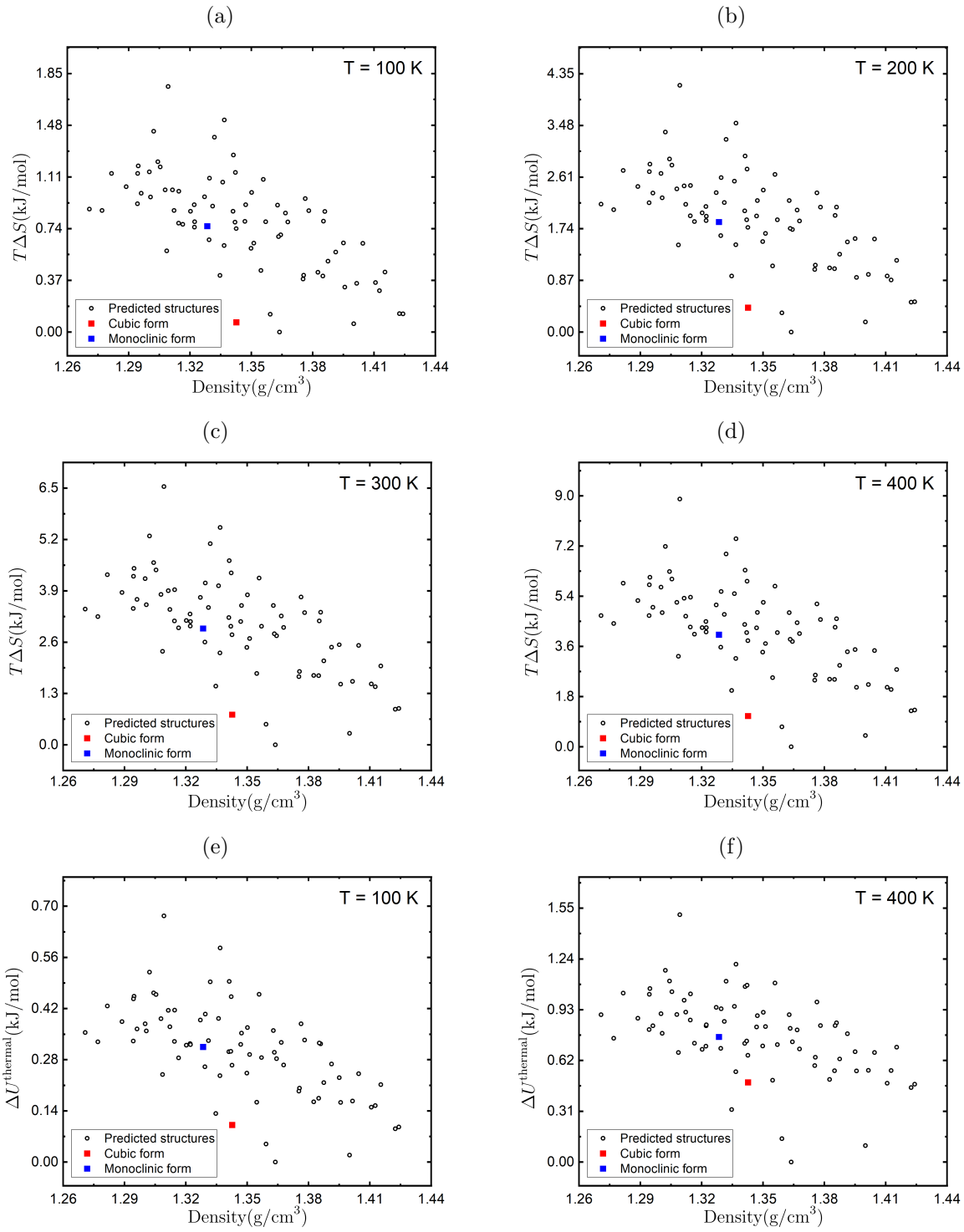


Figure 5.6: Calculated relative entropic contributions to free energy for the TCNE landscape (a) 100 K, (b) 200 K, (c) 300 K and (d) 400 K. Calculated thermal energy for the TCNE landscape at (e) 100 K and (f) 400 K. Structures shown as red and blue squares represent the cubic and monoclinic forms, respectively.

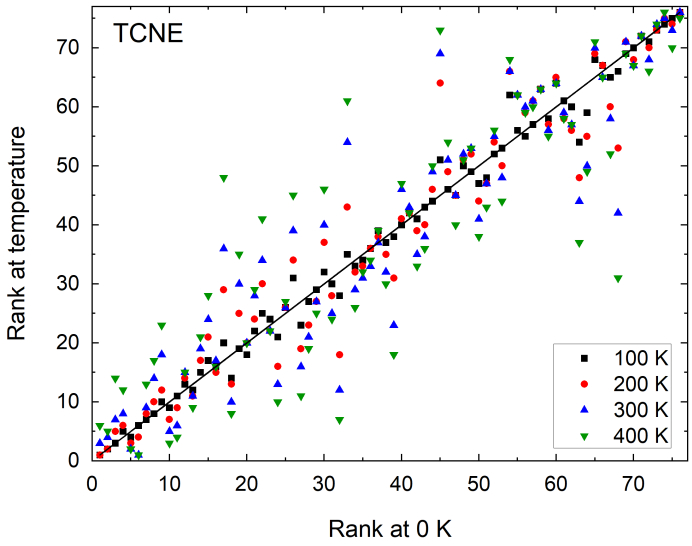


Figure 5.7: Rank of the 76 structures of the TCNE free energy landscape at various temperatures relative to their rank at 0 K.

The rank of the stable structures in the free energy landscape at various temperatures is shown as a parity plot in Figure 5.7 relative to their rank at 0 K. At 100 K, the effect of temperature on the stability is less significant as the structures in the parity plot are distributed close to the x=y line. The increase of temperature results in an increasing scatter of the data on the parity plot. To obtain a metric quantifying the effect of temperature on the ranks of structures, we calculated the Spearman rank correlation coefficient for each temperature relative to the ranks at 0 K. The calculated values are summarised in Table 5.3. The $R_S(T)$ values decrease at elevated temperatures indicating that the deviation of the ranks of structures relative to their rank at 0 K is increased with increasing temperature.

Table 5.3: Spearman rank correlation coefficient at various temperatures relative to 0 K.

	100 K	200 K	300 K	400 K
$R_S(T)$	0.994	0.967	0.925	0.860
$R_S(T) - R_S(0)$	0.006	0.033	0.075	0.140

5.2 Benzoic Acid CSP study

The current section summarises our full CSP study on Benzoic Acid involving global search, lattice energy and free energy refinements. Benzoic Acid has the chemical formula $C_7H_6O_2$ and consists of a benzene ring substituted with a carboxyl group (-COOH). The molecular connectivity diagram of Benzoic Acid is shown in Figure 5.8. The structure of Benzoic Acid crystals is stabilised by the formation of hydrogen bonds between the carboxylic acid groups of adjacent molecules. Despite being a compound that has been crystallised widely over the centuries, only one crystal structure has been observed experimentally [124]. There are 21 crystal structures deposited in the CSD (17 with atomic coordinates) under the family name BENZAC. This large number of deposited structures represents the collective efforts and studies dedicated to understanding the properties and behaviour of Benzoic Acid crystals and serves as an indication that another polymorph is unlikely to be discovered.

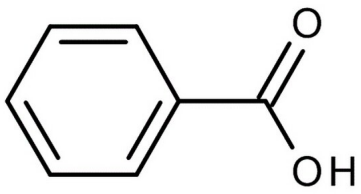


Figure 5.8: Molecular connectivity diagram of Benzoic Acid.

To the best of our knowledge, state-of-the art CSP methods have not been applied for the study Benzoic Acid single-crystals. Benzoic Acid co-crystals with caffeine [125] have been examined using CSP methods. CSP studies for of benzoic acid derivatives, such as chloro and methyl ortho-benzoic acids and their co- crystals [126] as well as nitrobenzoic acid derivatives and benzoic acid isomers [127] have been conducted using CrystalPredictor and CrystalOptimizer. In the former study, interchanging the chloro and methyl groups yields more stable experimental forms than their predicted counterparts. In the latter study, the experimental forms are predicted low in energy in the CSP landscapes, with many cases the experimental structures being the global minimum.

5.2.1 Global Search

The first step in the candidate generation stage involves the identification of the flexible DoFs and the generation of LAMs. The LAMs used in the global search stage have been generated in the work of Sugden *et al.* [97] using the adaptive LAM algorithm. For Benzoic acid, we treat the torsion describing the carboxylic group is treated as flexible and the geometry of the benzene ring is kept rigid to its gas phase conformation. LAMs have been generated at the M06/6-31G(d,p) level of theory across the full range of 360° with a step of 10°. Point charges, fitted to the same level of theory, are calculated with a real space cutoff of 12 Å. Repulsion-dispersion interactions are modelled via the Buckingham potential, employing the FIT parametrisation

with a cutoff of 15 Å. The global search is performed for one molecule in the asymmetric unit cell (Z'=1) using CrystalPredictor algorithm and yields 6,807 stable structures within 20 kJ/mol from the global minimum in lattice energy. Clustering with the CSD python API reduces the number of structures to 6,304. The clustered CrystalPredictor landscape consists of the final set of structures generated in the candidate generation stage (See Figure 5.9).

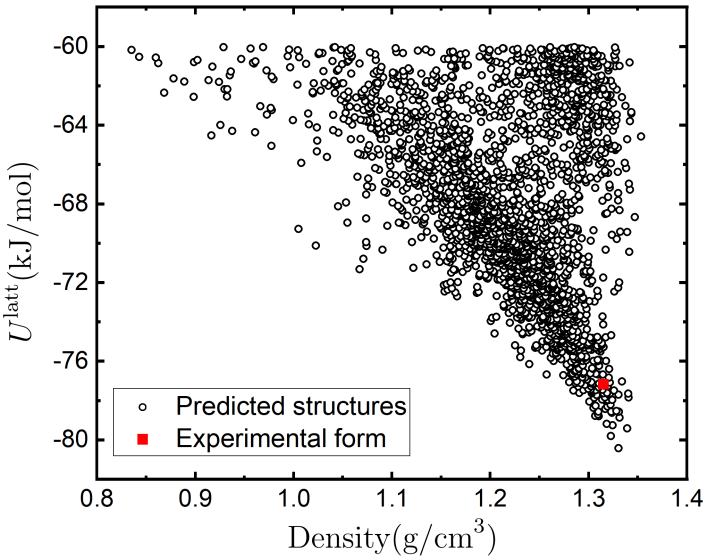


Figure 5.9: CrystalPredictor Z'=1 landscape of Benzoic Acid after clustering. The structure predicted 52nd in the lattice energy landscape (red square) matches the experimental form with RefCode BENZAC02.

The structure corresponding to the global minimum in CrystalPredictor landscape is found to have an energy of -80.419 kJ/mol. The experimental structure is ranked 52nd in the landscape, with an energy of 3.247 kJ/mol above the global minimum. This suggests that the experimental structure is found low in energy as it falls within 20 kJ/mol from the global minimum, which is the usual energy threshold for structures considered in subsequent steps. Structural similarity analysis was employed to assess the differences in geometry between the experimental and predicted structures. The $RMSD_{15}$ between the predicted structure and experimental form with CSD RefCode BENZAC02 was 0.307 Å and the $RMSD_1$ = 0.043 Å. PXRD pattern similarity was also found to have a high matching score of 0.965. The CrystalPredictor energy model provides a satisfactory energy and geometry prediction despite the usage of point charges for the electrostatic interaction. In the subsequent sections, the set of structures obtained in the global search stage is refined by calculating the lattice energy with more detailed potentials.

5.2.2 Refinement stage I - Lattice energy

The application of our free energy workflow has been presented in Chapter 4. In this section, we show calculations performed for the combination of settings that produced the most stable

structures. We remind the reader that two treatments of molecular flexibility were examined. An all-torsions as well as a fully atomistic treatment are employed to assess the minimisation scheme that yields the larger number of structures with positive phonons frequencies in the free energy landscape. In addition to the number of DoFs treated as flexible, we examine the effect of varying the step size used for calculating derivatives via FD. It was found that atomistic treatment in conjunction with FD steps for bonds lengths equal to 10^{-6} Å produced the largest number of structures with positive phonon frequencies. The latter step size is used for bonds lengths, alongside 10^{-3} rads for angles and dihedrals. In total, the Benzoic Acid molecule has 39 conformational DoFs, including 14 bond lengths, 13 angles and 12 dihedrals all of which were treated as flexible during minimisation. Quantum mechanical calculations were conducted at the PBE/6-31G(d,p) level of theory. A cutoff of 10 Å was utilised for the direct electrostatic interactions, modelled using DMA up to a combined rank of four ($l = 4$). The repulsion-dispersion interactions were quantified with FIT parameterisation using a cutoff of 15 Å. Refinement using CSO-FM is performed for the 100 lowest energy structures of the CrystalPredictor landscape generated in the previous stage.

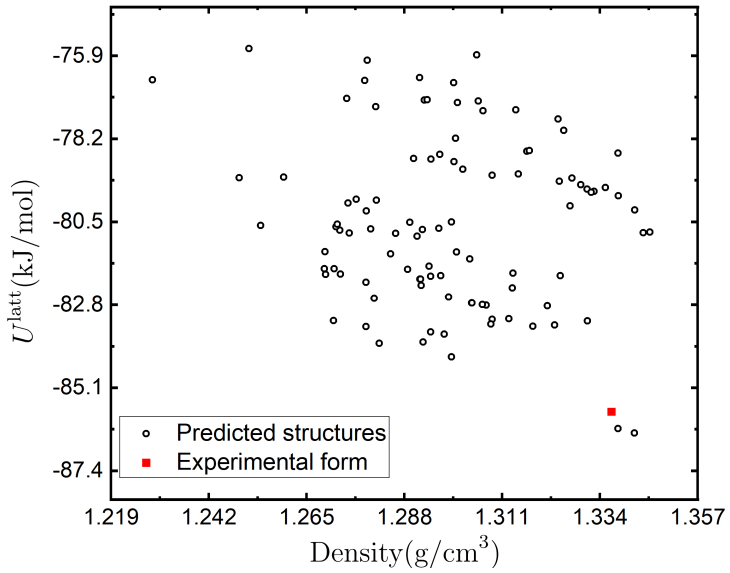


Figure 5.10: Landscape of the 100 lowest energy structures for Benzoic Acid minimised with CSO-FM. The structure matching the experimental form (BENZAC02) is highlighted with a red square and.

The landscape minimised using CSO-FM is depicted in Figure 5.10. The structure ranked 3rd is found to match the experimental crystal structure of Benzoic Acid (CSD RefCode BENZAC02). The structure corresponding to the global minimum in the CSO-FM landscape is found to have an energy of -86.353 kJ/mol with the experimental structure having an energy difference of just 0.585 kJ/mol. This improved ranking can be attributed to the usage of a different functionals in the refinement stage (PBE) compared to the one used in the global search (M06). In addition, CSO-FM makes use of a more accurate energy model that employs multipoles and incorporates molecular flexibility. Regarding geometry prediction, the experimental structure

with CSD RefCode BENZAC02 is found to match the structure ranked 3rd, with an $RMSD_{15}$ = 0.209 Å and $RMSD_1$ = 0.048 Å. The geometry reproduction of the experimental crystal structure is improved in comparison to CrystalPredictor, with very minor differences in the molecular conformation obtained by both algorithms.

5.2.3 Refinement stage II - Free energy

The final stage of the CSP study involves the evaluation of the vibrational contributions to free energy for all 100 structures in the CSO-FM lattice energy landscape. The free energies are computed using CrystalDynamics employing an energy model that is consistent with the one adopted in the lattice energy refinement stage (see Section 5.2.2). The application of CrystalDynamics has been presented in Section 4.4.2, where it was found that 83 structures displayed positive phonon frequencies across all sampled $\boldsymbol{k}$ points. Here, we present the free energy landscapes and discuss ZPE, entropy and thermal energy.

The ZPE landscape is shown in Figure 5.11a. Selecting the free energy at 0 K as the stability ranking criterion by considering ZPE does not alter the lattice energy ranking. Although the structure matching the experimental form has a high ZPE value, its ranking remains unaffected at 0 K. The free energy at temperatures between 0 K and 400K are shown in Figures 5.11b-5.11f. For all structures, the vibrational free energy decreases with increasing temperature. The three lowest energy structures in the lattice energy landscape are stable as they display positive phonon frequencies. As mentioned in the previous section, the experimental form was found to be the 3rd lower energy structure in the CSO-FM lattice energy landscape. The same ranking is maintained in the free energy landscape across the 0-400 K range.

We evaluate the impact of entropic contribution in relative stability. The differences in entropic contributions, $T\Delta S$ at 100 K, 200 K, 300 K and 400 K are shown in Figures 5.12a-5.12d. The difference between the minimum and maximum calculated $T\Delta S$ value in the free energy landscape increases with increasing temperature. Specifically, the difference is 1.064 kJ mol^{-1} K^{-1} at 100 K and reaches 4.679 kJ mol^{-1} K^{-1} at 400 K. The structure matching the experimental form exhibits the 3rd lowest entropy at 100 K and the 2nd lowest at higher temperatures. This low relative entropy indicated a tendency for destabilisation of the experimental structure. Despite that stabilisation, no difference in ranking is observed for temperature up to the melting point of Benzoic Acid, which is 396 K [128].

The calculated $\Delta U^{\text{thermal}}$ at 100K and 400 K are shown in Figures 5.12e-5.12f. The $\Delta U^{\text{thermal}}$ difference between the minimum and maximum values for the 100 lowest energy structures are 0.319 and 0.703 kJ/mol for 100 K and 400 K, respectively. Our results confirm the significance of entropic contributions in crystalline stability at elevated temperatures. Although free energy does not improve the ranking, in this case, the study shows the feasibility of taking free energy into account within CSP.

The parity plot of the ranks of the stable structures in the Benzoic acid landscapes at various temperatures relative to their rank at 0 K is shown in Figure 5.13. The rise of temperature does

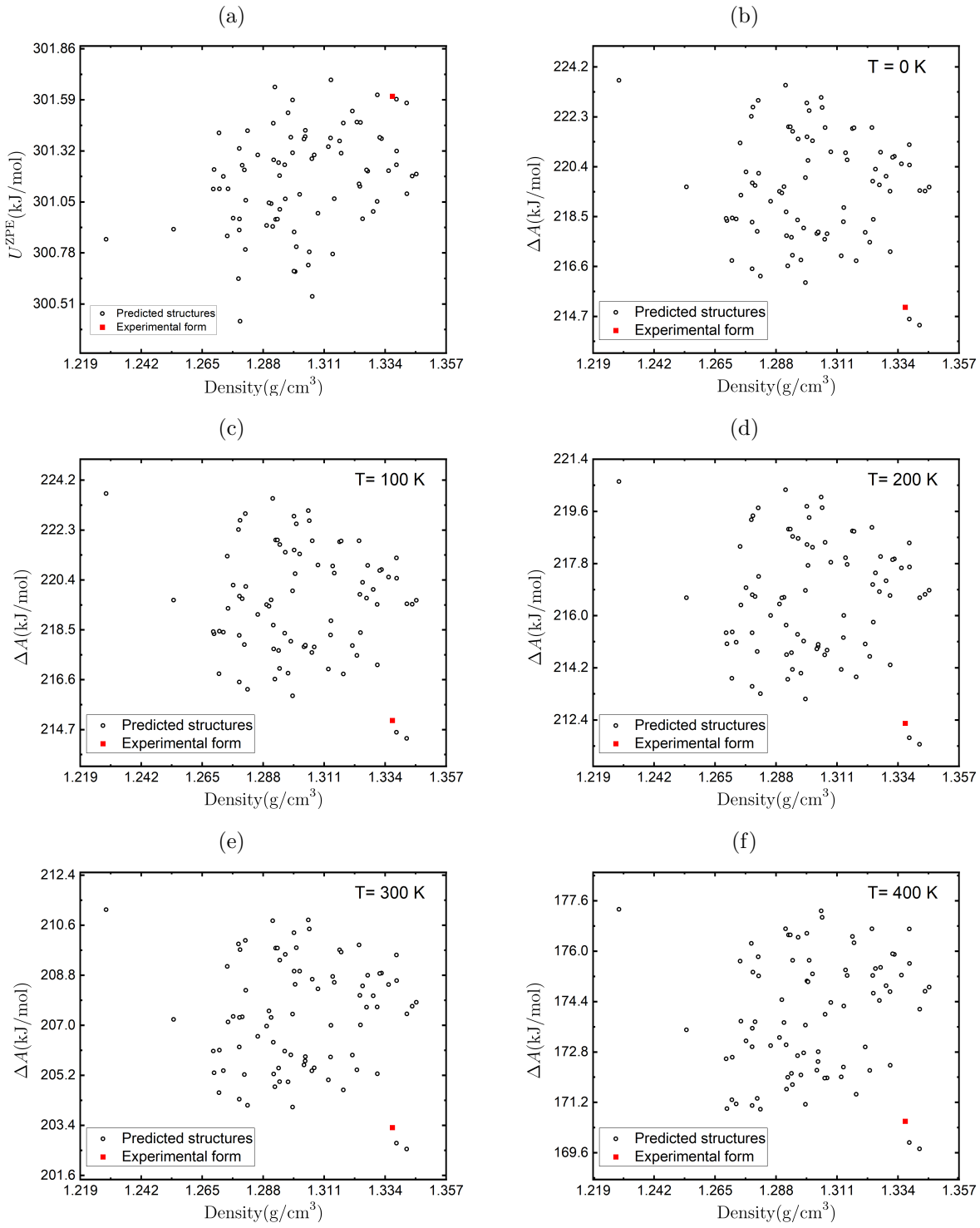


Figure 5.11: (a) ZPE landscape of Benzoic Acid and free energy landscapes at (b) 0 K, (c) 100 K, (d) 200 K, (e) 300 K and (d) 400 K. The predicted structure matching the experimental form (BENZAC02) is highlighted with a red square.

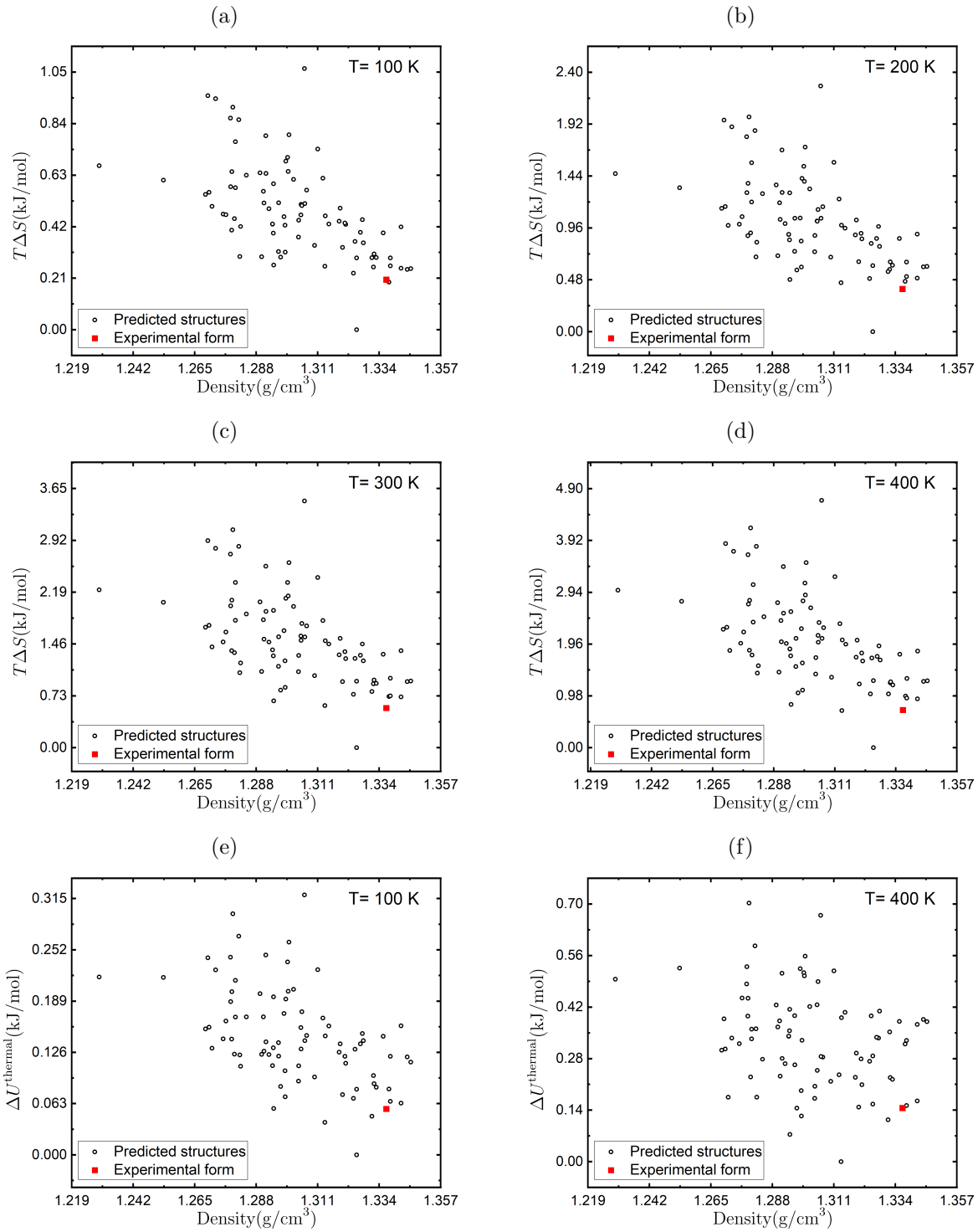


Figure 5.12: Entropic contributions of the Benzoic acid landscape at (a) 100 K, (b) 200 K, (c) 300 K, (d) 400 K. Thermal energy for the Benzoic acid at (e) 100 K and (f) 400 K. The predicted structure matching the experimental form (BENZAC02) is highlighted with a red square.

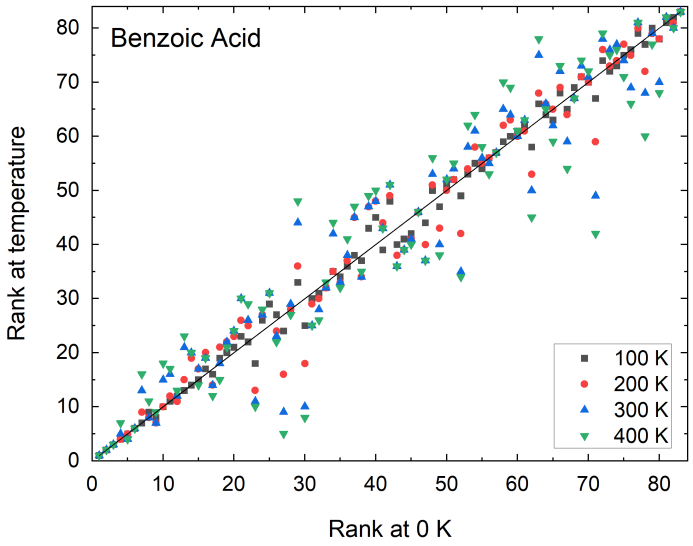


Figure 5.13: Rank of the 83 structures of the Benzoic acid free energy landscape at various temperatures relative to their rank at 0 K.

not affect the rank of the three lowest energy structures including the experimental form, which is ranked 3rd. The Spearman rank correlation coefficient is calculated for each temperature relative to the ranks at 0 K and (See Table 5.4). The value of 0.992 calculated for the difference in rank between 0 K and 100K confirms that the impact of temperature in ranking is negligible at lower temperatures. At elevated temperatures, the calculated Spearman rank correlation coefficients are close to 1, indicating that the effect of temperature on the relative stability is relatively low for the case of Benzoic acid.

Table 5.4: Spearman rank correlation coefficient at various temperatures relative to 0 K for Benzoic acid.

	100 K	200 K	300 K	400 K
$R_S(T)$	0.997	0.984	0.959	0.935
$R_S(T) - R_S(0)$	0.003	0.016	0.041	0.065

5.3 Target XXIX from 7th CSP blind test

This section provides a summary of the CSP study conducted for Target XXIX of the 7th CSP blind test organised by CCDC. For Target XXIX, CCDC disclosed the chemical connectivity diagram alongside the experimental PXRD pattern in the so-called "experimentally assisted challenge". The CSP study shown in this section aimed to produce a ranked list of plausible crystal structures for Target XXIX of the blind test. Here, we report the Z'=1 CSP study that involves global search and application of our workflow to evaluate the lattice and free energy landscapes of Target XXIX. Limited Z'=2, 3 and 4 searches have been also carried out but were incomplete. The experimental structure has three molecules in the asymmetric unit cell (Z'=3) and was not found in the search. In addition, we present the post-test analysis of the experimental structure of Target XXIX, after its release with CSD RefCode FASMEV. This analysis includes local minimisation of the experimental form using the energy models of CrystalPredictor and CrystalOptimizer, and free energy evaluation using CrystalDynamics.

5.3.1 Global Search

A global search for Target XXIX, with chemical formula $C_6H_4(NH_2)(CO_2H)$ or $C_8H_9O_2N$ has been carried out to generate a list of plausible crystal structures. The molecular connectivity diagram alongside the three DoFs treated as flexible during the candidate generation stage are shown in Figure 5.14. This choice was based on chemical intuition in order to generate sufficient candidate structures at the global search stage. The flexible DoFs include the dihedral O1-C2-C3-C4 (dih4) of the carboxylic group connected to the benzene, the dihedral H5-C1-O1-C2 (dih5) describing rotation of the terminal methyl group linked to one of the oxygens, and the dihedral H1-N1-C8-C7 (dih16) of the amine group linked to the benzene. Initially, an unconstrained geometry optimisation was carried out using the hybrid PBE0 functional and three progressively more diffuse basis sets to evaluate the gas phase energy and geometry. The basis sets used include 6-31G(d,p), 6-311G(d,p) and 6-311+G(d,p). The values of the flexible dihedrals for the minimised *in vacuo* conformations can be found in Table 5.5. Isolated-molecule calculations reveal that similar values of the dihedrals are obtained in the gas phase for various basis sets and the same level of theory. Consequently, we proceed with LAM generation using the most diffuse basis set 6-311+G(d,p), which is expected to be more accurate and which remains affordable despite the higher computational cost.

The next step of the CSP study involves generating LAMs that will be used in the global search stage. Uniform LAMs are created through constrained geometry optimizations of isolated molecules, considering various molecular conformations. The flexible DoFs are restricted to a grid of discrete values, while the rigid DoFs are allowed to vary freely. To generate a comprehensive set of LAMs for each flexible DoF, calculations are performed spanning the entire range of 360°, with a step size of 60°. This ensures a thorough exploration of the conformational space, enabling the identification of low energy regions. This systematic procedure results in

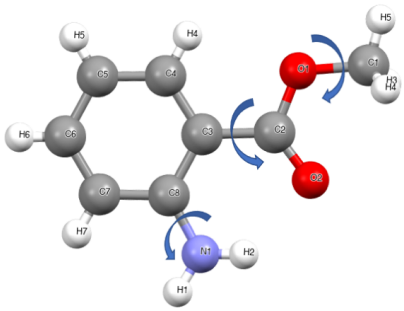


Figure 5.14: The molecular connectivity diagram of Target XXIX. The blue arrows represent the dihedrals: O1-C2-C3-C4 (dih4), H5-C1-O1-C2 (dih5), and H1-N1-C8-C7 (dih16), which have been treated as flexible during the global search stage.

Table 5.5: Values of flexible dihedrals obtained from atomistic geometry optimisations of TCNE with the PBE0 functional and three different basis sets

Torsion/Basis Set	6-31G(d,p)	6-311G(d,p)	6-311+G(d,p)
O1-C2-C3-C4 (dih4)	180.4130	180.3371	180.2872
H5-C1-O1-C2 (dih5)	-180.8213	-179.2112	-178.8522
H1-N1-C8-C7 (dih16)	-8.0871	-8.5269	-8.4391

the creation of 216 uniform LAM points illustrated by 2D heat maps (see Figures 5.15a, 5.15c and 5.15e). The adaptive LAM generation algorithm [97, 21] was utilised to determine unpopulated conformational regions of low confidence. An intramolecular energy cutoff of 30 kJ/mol and a maximum LAM energy difference of 5 kJ/mol are used to determine regions requiring additional sampling. This process suggests the generation of 97 additional adaptive LAM points (see Figures 5.15b, 5.15d and 5.15f). In total, 313 uniform and adaptive LAM points were used in the global search stage. Comparing Figures 5.15a and 5.15b reveals that the adaptive LAM generation algorithm identifies a spurious minimum located around dih4=330° and dih5=150°. Similarly, a low energy region, around dih4=330° and dih16=40°, disappears with the application of adaptive LAM algorithm (see Figures 5.15c and 5.15d). It can be seen from figure 5.15 that the energetically preferable conformations are close to the gas phase values reported in Table 5.5.

The global search is performed using the CrystalPredictor algorithm [98, 99, 22] version 2.4.4 [97, 21] with LAMs generated at the PBE0/6-311+G(d,p) level of theory. A cutoff of 25 Å was used for the electrostatic interactions modelled with point charges. FIT potential parameters are used to model the repulsion-dispersion interactions with a cutoff of 25 Å and a quintic spline [20] cutoff of 22 Å. In total, one million initial crystal structures with one molecule

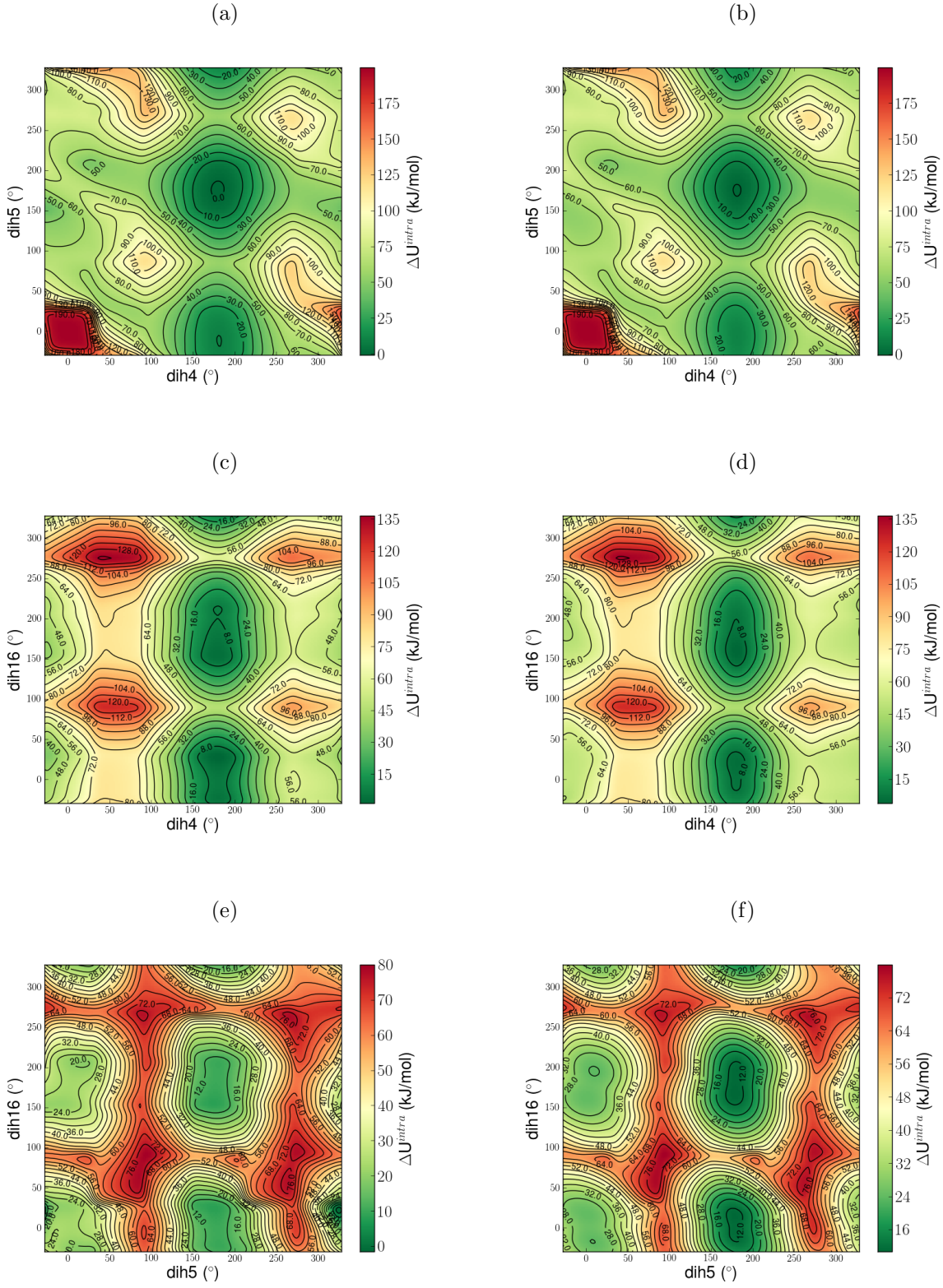


Figure 5.15: 2D heat maps for various conformations of molecule XXIX. (a) uniform & (b) adaptive LAMs for dih4 and dih5, (c) uniform & (d) adaptive LAMs for dih4 and dih16, and (e) uniform & (f) adaptive LAMs for dih5 and dih16. Calculations were performed at the PBE0/6-311+G(d,p) level of theory.

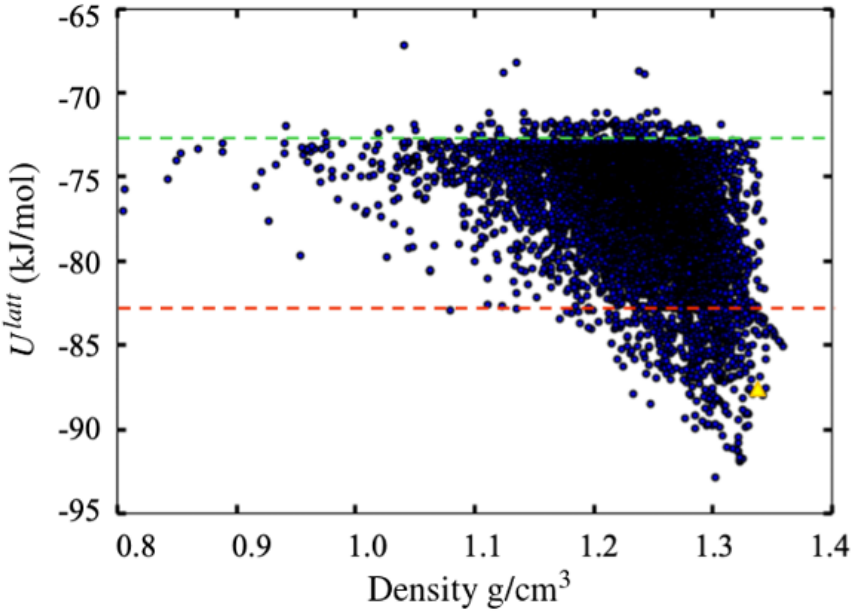


Figure 5.16: CrystalPredictor Z'=1 landscape for molecule XXIX after clustering. The yellow triangle indicates the locally minimised experimental structure (FASMEV), which has Z'=3 and was not found in the search. The red and green dashed lines indicate the 10 kJ/mol and 20 kJ/mol thresholds from the global minimum.

in the asymmetric unit cell (Z'=1) are generated using Sobol' sequencing [19]. All the initial structures are minimised with CrystalPredictor's energy model resulting in 16,856 structures within 20 kJ/mol from the global minimum in lattice energy, with a value of -92.860 kJ/mol. Clustering with CSD python API [122, 101] further reduced the number of structures to 6,980 structures, which make up the final global search landscape shown in Figure 5.16.

Global searches with up to four molecules in the asymmetric unit cell (Z'=2, Z'=3 and Z'=4) were performed with the same energy model and LAM database used in the Z'=1 search. The searches with more than one molecules in the unit cell were incomplete due to the shortage of computational resources and are not discussed in this thesis. The experimental structure of Target XXIX (CSD RefCode FASMEV) has Z'=3 and has not been found at the global search stage. Following the release of the experimental form, we have applied our workflow to asses the performance of our energy models in reproducing the experimental crystalline geometry. Additionally, we examine the energy of experimentally minimised form and its rank in the lattice and free energy landscape. FASMEV is locally minimised with CrystalPredictor using the same modelling choices as during the blind test. It is found to have a lattice energy value of -87.860 kJ/mol (see figure 5.16). The calculated lattice energy of the experimental form is approximately 5 kJ/mol above the Z'=1 CrystalPredictor global minimum and lies within the 20 kJ/mol energy threshold. The lattice energy of FASMEV is very close to the CrystalPredictor lattice energy minimum, indicating that if this structure had been generated, it would likely have been refined in the next stage. As far as geometry is concerned, the structure obtained

after local minimisation of FASMEV is a good match for the experimental structure with an $RMSD_{20}$ of 0.194 Å. This value is well below the matching cutoff value of 0.600 Å that was used in the 6th blind test [11].

5.3.2 Refinement stage I - Lattice energy

Before we proceed to refining the global search landscape, we locally minimise a similar compound to assess the accuracy of the energy model in reproducing the experimental geometry and stability at 0 K. For this purpose, we performed a polymorphic stability study on Anthranilic Acid. Target XXIX and Anthranilic Acid are fairly similar, with the former being an ester while the latter is a carboxylic acid. The difference of Target XXIX is the presence of a terminal methyl (-CH_3) group attached to one of the oxygen atoms. Anthranilic Acid has three polymorphs, form I (AMBACO07), form II (AMBACO10) and form III (AMBACO08) all of which have been minimised. CrystalOptimizer was used instead of CSO-FM as the latter had not been implemented at the time of this study. The three conformational DoFs that were chosen to be treated as flexible are chosen to match the flexible DoFs in the global search stage. Quantum mechanical calculations were performed at the PBE0/6-311+G(d,p) level of theory. Two parameter sets were used to model the repulsion-dispersion interactions. The parameter set DB2020_PBE0_6-311+G(d,p) that was developed at the PBE0/6-311+G(d,p) functional/basis set [40] and FIT potential. The geometry reproduction is shown in Table 5.6 and is satisfactory for all the forms using both sets of repulsion-dispersion interactions. Experimental studies on Anthranilic Acid polymorphs [129, 130, 131] suggest that form I is the most stable polymorph at 0 K, in agreement with the energy ranking suggested by the chosen level of theory. This study on Anthranilic Acid polymorphs provides justification to adopt the aforementioned computational parameters in the subsequent refinement stages.

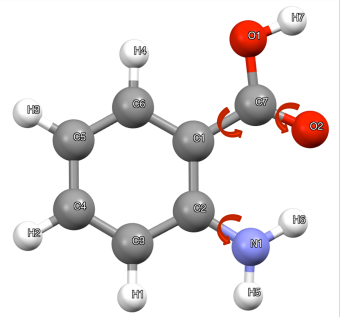


Figure 5.17: Molecular connectivity diagram of Anthranilic Acid taken from CSD RefCode: AMBACO08.

Having established the energy model to be utilised, we proceed to refining the 100 lowest energy structures of the Z'=1 CrystalPredictor landscape of molecule XXIX using CSO-FM. Quantum mechanical calculations are performed at the PBE0/6-311+ level of theory. Multipoles with a combined rank of up to four ($l = 4$) are derived from gas phase densities using GDMA [31]. The electrostatic interactions are modelled using a real space cutoff 10 Å and

Table 5.6: Geometry and energy of the three Anthranilic Acid polymorphs locally minimised with CrystalOptimizer using two different parameter sets for the repulsion-dispersion interactions.

Form	RefCode	FIT		DB2020_PBE0_6-311+G(d,p)	
		$RMSD_{15}$(Å)	U^{latt} (kJ/mol)	$RMSD_{15}$(Å)	U^{latt} (kJ/mol)
I	AMBACO07	0.184	-100.965	0.235	-90.384
II	AMBACO10	0.323	-95.052	0.399	-83.486
III	AMBACO08	0.216	-101.155	0.405	-88.127

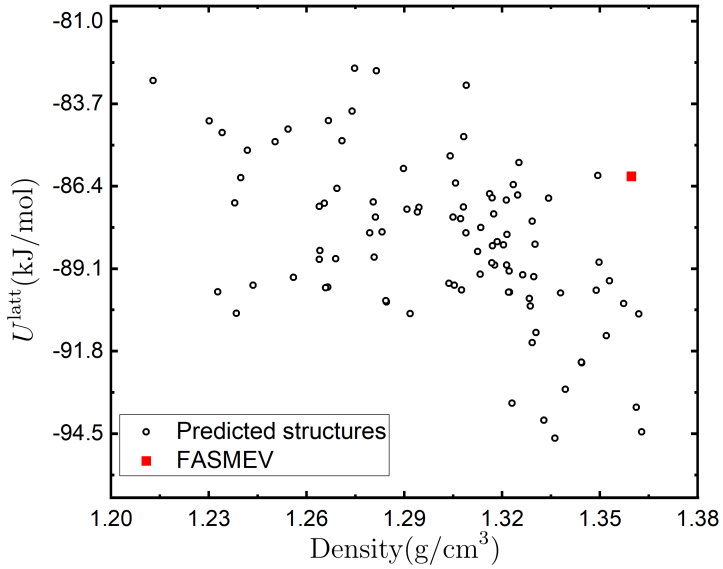


Figure 5.18: CSO-FM landscape for molecule XXIX containing 100 CSO-FM minimised structures. The req square indicates the locally minimised experimental structure (FASMEV)

an accuracy of Ewald summation equal to 10^{-6}. Repulsion-dispersion parameters are derived at the aforementioned level of theory and calculated with a cutoff of 15 Å. An "all torsions" minimisation is carried out treating a total of 17 torsions as the flexible degrees of freedom. The landscape was refined utilising LAMs during the minimisation, although upon convergence a single point isolated-molecule gas phase calculation is performed to cancel the impact of LAMs on the final list from refinement stage I.

The experimental crystal structure of Molecule XXXI (FASMEV) is locally minimised with CSO-FM utilising the same energy model with the one described in this section. The $RMSD_{20}$ between the experimental structure (FASMEV) and its locally-minimised equivalent is 0.193 Å, indicating a close match. The minimised structure is found to lie -8.579 kJ/mol above the lowest-energy structure in the Z'=1 landscape. We can conclude that the models used in the global search and local lattice energy minimisation are suitable for Target XXIX and that, had extensive Z'> 1 searches been completed, the experimental structure would have been very likely included in the submitted list due to its low rank in and CSO-FM landscape (See Figure 5.18).

5.3.3 Refinement stage II - Free energy

We employ the CrystalDynamics algorithm to evaluate the vibrational free energy of the 100 lowest energy structures obtained in the lattice energy refinement stage. 79% of structures emerge as stable in the free energy landscape as they display positive phonon frequencies along all sampled $\boldsymbol{k}$ points. As mentioned earlier, the experimental structure of Target XXIX (FASMEV) has Z'=3 and was not found in the global search stage. We have evaluated harmonic phonon frequencies for FASMEV to compare the free energy of the experimental form with the free landscape of the Z'=1 predicted structures. Prior to the free energy evaluation the experimental crystal structure is locally minimised with CSO-FM treating all torsions as flexible DoFs. FASMEV, locally minimised with CSO-FM emerges as a stable structure with positive phonons for all sampled $\boldsymbol{k}$ points.

The ZPE landscapes of Target XXIX is shown in Figure 5.19a alongside the calculated ZPE value of the locally minimised FASMEV structure. All computed ZPE values are distributed within 1 kJ/mol, in agreement with distribution reported by Nyman and Day [51]. Minor reranking can be observed with the inclusion of ZPE to lattice energy landscape to use free energy at 0 K as the relative stability criterion (see Figure 5.19b). The rest of free energy landscapes for Target XXIX between 100 K and 400 K are shown in Figures 5.19b-5.19f (including FASMEV). At 400 K, FASMEV is found to have an free energy of 286.954 kJ/mol, which places the experimental structure -6.542 kJ/mol above the global minimum in the Z'=1 landscape. We can conclude that if the the experimental structure had been generated, it would have been ranked low in energy in the Z'=1 landscape.

The relative differences in entropic contribution ($T\Delta S$) are evaluated and are shown in Figures 5.19b-5.20d. We can observe that in general vibrational contributions tend to stabilise the less dense structures as these display higher entropic contributions. This finding has also been reported in the literature [17, 123]. Despite the fact that the experimental form has not been generated and the free energies have been calculated *a posteriori*, our findings suggest that CrystalDynamics algorithm can be used for refining CSP landscapes of compounds exhibiting molecular flexibility and identify structures with high entropic contributions.

Similarly to the CSP studies on the other molecules presented in this chapter, we assess the dependence of temperature on relative stability. The rank of the structures that emerges as stable in the free energy landscape at various temperatures is shown as a parity plot in Figure 5.21 relative to their rank at 0 K. In addition, we evaluate the Spearman rank correlation coefficient to assess the effect of temperature on the ranks of structures (see Table 5.7). As evidenced by both the parity plot and the calculated Spearman rank correlation coefficients the impact of temperature on the ranks increases with increasing temperature. In comparison to Benzoic acid, which has a smaller size compared to Target XXIX and and exhibits just one flexible DoF, we can observe an increasing scatter of the data on the parity plot. This increased scatter can be attributed to the larger size and flexibility of Target XXIX.

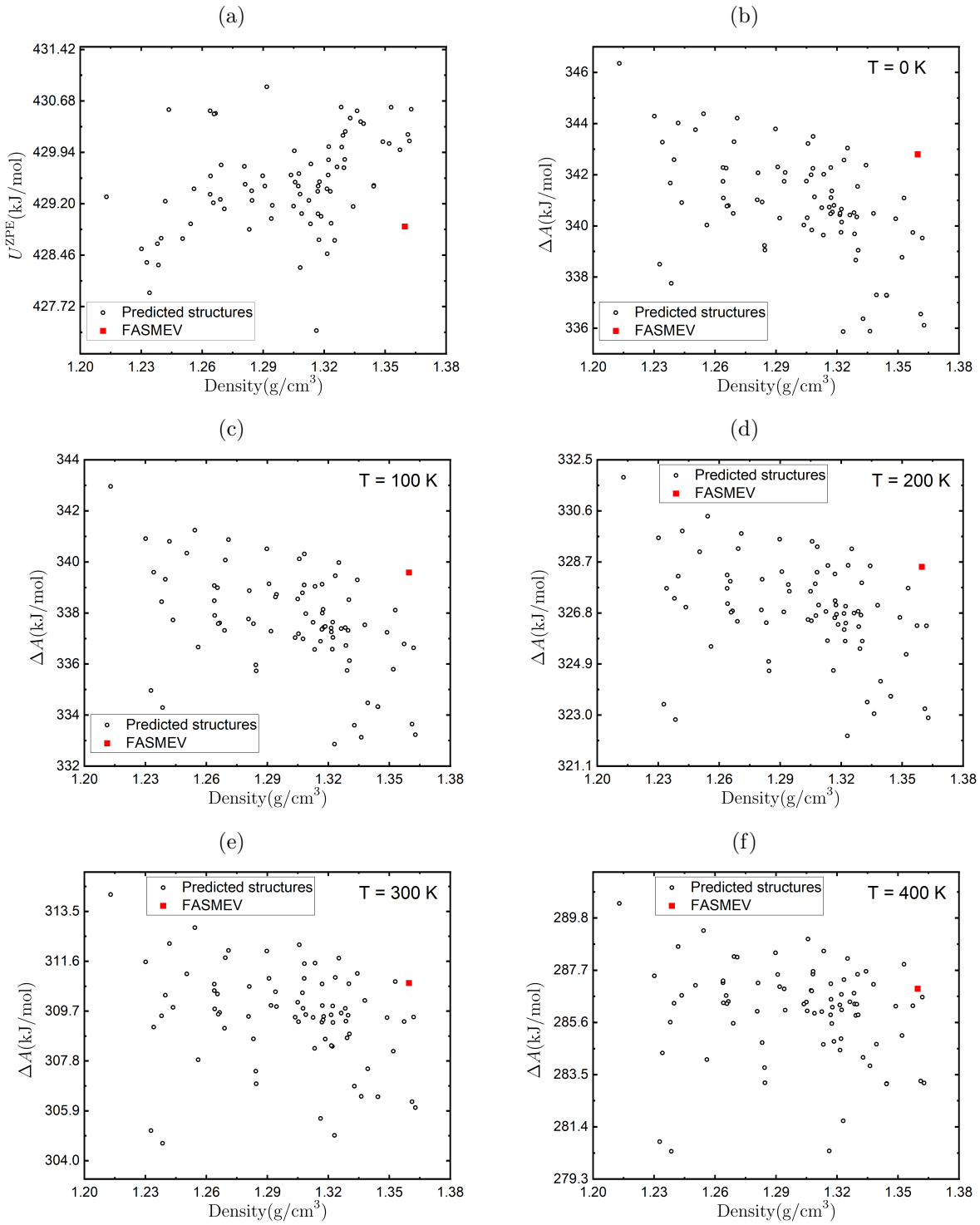


Figure 5.19: (a) ZPE energy landscape of XXIX. Free energy landscape at (b) 0 K, (c) 100 K, (d) 200 K, (e) 300 K and (d) 400 K. The red square represents the ZPE and free energy at various temperatures of the experimental form (FASMEV), which was not found in the CSP study.

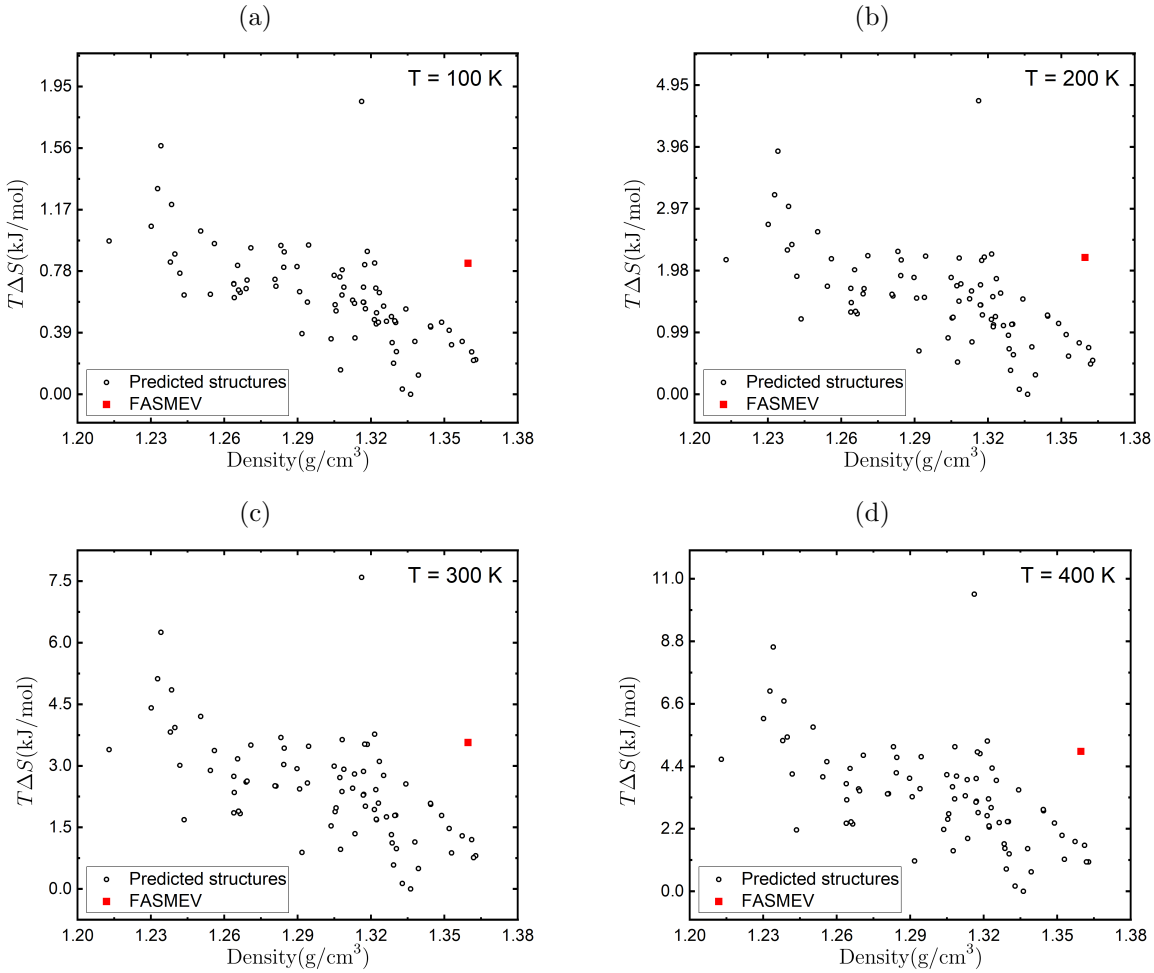


Figure 5.20: Calculated relative entropic contributions to free energy for the landscape of Target XXIX at (a) 100 K, (b) 200 K, (c) 300 K and (d) 400 K. The red square represents the relative entropic contributions of the experimental form (FASMEV), which was not found in the CSP study

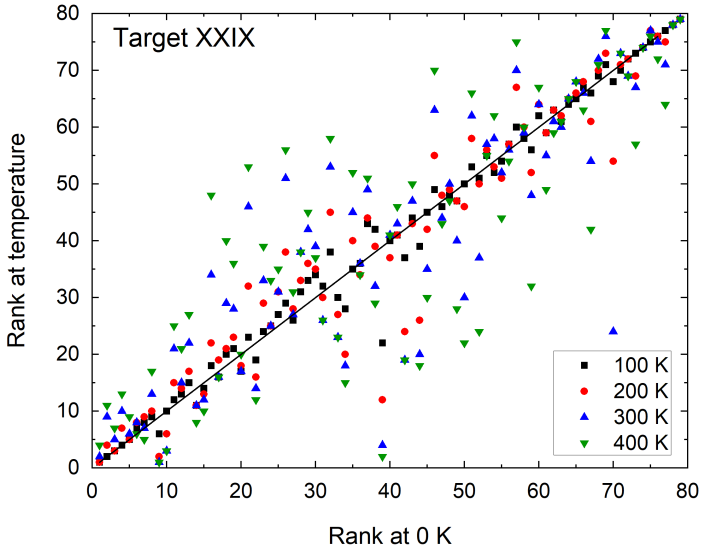


Figure 5.21: Rank of the 79 structures of Target XXIX free energy landscape at various temperatures relative to their rank at 0 K.

Table 5.7: Spearman rank correlation coefficient at various temperatures relative to 0 K for Target XXIX.

	100 K	200 K	300 K	400 K
$R_S(T)$	0.992	0.960	0.875	0.771
$R_S(T) - R_S(0)$	0.008	0.040	0.125	0.229

5.4 Conclusions

In this chapter, we have conducted CSP studies for TCNE, Benzoic Acid and Target XXIX from the 7th CCDC blind test, using our multistage approach that involves candidate generation and two refinement stages. The global search stage has been carried out with CrystalPredictor. The experimental forms of TCNE and Benzoic acid were identified as low energy structures, with satisfactory reproduction in the geometry. In the case of Target XXIX, the experimental structure (FASMEV) has three molecules in the asymmetric unit cell (Z'=3) and was not generated in the global search. The local minimisation of the experimental form of Target XXIX placed the structure low in the energy landscape suggesting that it would have been likely to be included in subsequent refinement stages.

The first refinement stage involved lattice energy minimisation using either CrystalOptimizer or CSO-FM. The energy model used in the lattice energy mini..misation stage accounts for molecular flexibility. In addition, the electrostatic interactions are modelled using multipole moments instead of point charges. This more accurate energy model results in an improved ranking in comparison to CrystalPredictor as well as a better geometry reproduction of the experimental forms. For TCNE, the cubic form was identified as the global minimum and

the monoclinic as the 6th lowest energy structure. The experimental structure of Benzoic acid matched the 3rd lowest energy structure of the CSO-FM landscape. The local minimisation of FASMEV placed the form within 10 kJ/mol from the global minimum in the Z'=1 lattice energy landscape.

The second refinement stage involved evaluation of the vibrational free energy using CrystalDynamics. Our free energy workflow was applied to 100 lowest energy structures generated during the candidate generation stage. For TCNE, the cubic form was also predicted as the global minimum in the free energy landscape at 0 K. The enantiotropic transition from the cubic to the monoclinic form was found to occur at 280 K. Free energy calculations has been performed for Benzoic acid and Target XXIX. This is the first time that CrystalDynamics algorithm has been utilised to evaluate the free energy of crystals comprised of compounds exhibiting molecular flexibility. For Benzoic acid, evaluation of Helmholtz free energies did not alter the rank of the experimental form, which remained the 3rd lowest energy across the whole temperature range of 0-400 K. The calculated free energies for FASMEV indicated that the experimental form is ranked among the lowest energy structures in the Z'=1 landscape.

Finally, the stability of crystal structures across different temperatures is assessed in terms of free energy. The vibrational free energies, ZPE, entropy and thermal energy have been computed by evaluating phonons frequencies for all sampled points in the reciprocal space. Most of the examined structures have positive phonon frequencies and emerge as stable highlighting the applicability of our method to CSP generated landscapes. Overall, the application of our energy workflow to three molecules with progressively increasing size and flexibility demonstrates that the inclusion of free energy calculation in CSP is feasible for flexible molecules. In the next chapter, we assess the performance and limitations of our workflow by applying it to a benchmark set of molecules with larger sizes and more flexible functional groups.

Chapter 6

On the use of lattice dynamics to investigate polymorphic stability

In the current chapter, a polymorphic stability study is conducted by applying our free energy workflow to crystal structures obtained from the CSD. Given a set of polymorphic structures that exhibit enantiotropic transitions experimentally, we investigate the effect of ZPE on relative stability at 0 K. Furthermore, we assess how the calculated relative stability is affected by temperature using free energy as the stability metric. In turn, we examine whether the harmonic approximation of lattice dynamics theory can accurately predict the experimentally reported polymorphic stability and enantiotropic transitions temperatures. To validate our workflow, we propose a benchmark set consisting of six flexible molecules each possessing well-characterised enantiotropic transition temperatures. We compare our results with experimental findings as well as free energy calculations performed with the QHA and MBAR.

6.1 Benchmark set of enantiotropically related polymorphs

To understand the concepts and findings presented in this chapter, we first introduce some fundamental principles and prerequisites. At a given temperature and pressure, the stability relationship between two polymorphs is governed by their Gibbs free energy, with the lowest free energy structure considered to be the most stable. The work term, PV, is often neglected, and the Helmholtz free energy is used as an approximation of the Gibbs free energy for determining stability rankings. In the harmonic approximation of lattice dynamics theory, the volume of the unit cell is kept fixed to its lattice energy minimum value, neglecting the impact of thermal expansion at elevated temperatures. As a result, the difference in PV between polymorphs is a function of the differences in the static volumes of each optimised unit cell. For molecular crystals under ambient conditions, the contribution of the PV term is typically on the order of J/mol, while the Gibbs free energies are on the order of kJ/mol, justifying the common practice to use Helmholtz free energy as the stability criterion. Similarly to the previous chapter, we employ Helmholtz free energy as the metric for assessing the relative stability of polymorphs.

The stability relation between two polymorphs can be either monotropic or enantiotropic. In cases where the free energy curves of the two polymorphs do not intersect across all temperatures up to the melting point, the relation between the forms is referred to as monotropic. In a

monotropic relation, one polymorph is thermodynamically more stable than the other between 0 K and the melting point. On the other hand, an intersection between the free energy curves of the two polymorphs indicates a change in their stability relation via an enantiotropic transition (see Figures 6.2c - 6.2f). All the polymorphic transitions discussed in this section are temperature-mediated, meaning that the transitions occur due to changes in temperature. We remind the reader that throughout this chapter, the polymorph that is experimentally determined to be the most stable at low temperatures is referred to as the low-temperature form. In contrast, the polymorph that is more stable at elevated temperatures is referred to as the high-temperature form.

We have selected a set of six enantiotropically related polymorphs, comprised of flexible molecules, to benchmark our free energy workflow against enantiotropic transition temperatures reported in experimental studies. The selected molecules contain varying numbers and types of flexible groups, allowing us to thoroughly test and validate the applicability of our method across a diverse range of crystal structures. Most of the compounds included in the set are chosen from the work of Dybeck *et al.* [63]. In that work, the free energies of polymorphic pairs were calculated using the QHA and MBAR along a pseudo-supercritical path (PSCP), with interactions quantified using the OPLS force field [132]. Although Dybeck *et al.* [63] considered 12 polymorphic pairs (24 crystal structures), not all of them are included in our benchmark set. Since the application of our methodology to flexible molecules is novel, we chose to restrict the selected compounds to those containing only Carbon (C), Hydrogen (H), Oxygen (O), and Nitrogen (N) atoms. Chlorpropamide, Tolbutamide, Resorcinol and Diiodobenzene are excluded because they contain Sulphur, (S) Deuterium (D), or Iodine (I), atoms. Furthermore, Cyclopentane and Succinonitrile, are discarded because the structure of one of their polymorphs is disordered. We also excluded Aripiprazole, which is the largest molecule with 57 atoms, due to its size.

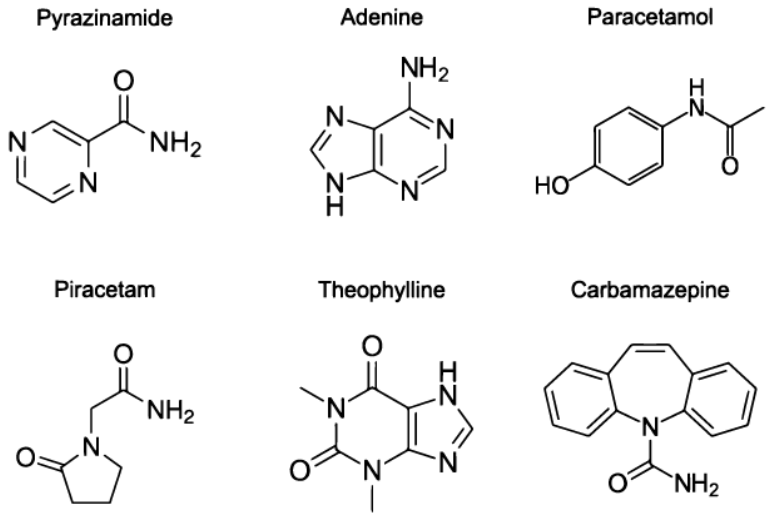


Figure 6.1: Molecular connectivity diagrams of all compounds included in the benchmark set.

Table 6.1: Compound names, CSD RefCodes, names of polymorphs, Z/Z' numbers and enantiotropic transition temperatures of all compounds used in the benchmark set.

Compound	RefCode	Form	Z/Z'	Stable T	Stable T Ref
Pyrazinamide	PYRZIN16	IV (delta)	2/1	0-298 K	[133]
	PYRZIN20	III (gamma)	2/1	> 428 K	
Adenine	KOBFUD	I	8/2	0-523 K	[134]
	KOBFUD01	II	16/1	> 523 K	
Paracetamol	HXACAN	II (ortho)	8/1	0-108 K	[135]
	HXACAN01	I (mono)	4/1	> 108 K	
Piracetam	BISMEV02	III	4/1	0-393 K	[136]
	BISMEV03	I	4/1	> 393 K	
Theophylline	BAPLOT02	IV	8/2	293-483 K	[137]
	BAPLOT06	II	4/1	483-541K	
	BAPLOT04	I	4/1	> 541 K	
Carbamazepine	CBMZPN02	III	4/1	0-435 K	[138]
	CBMZPN11	I	8/4	> 435 K	

The molecules included in our set are Pyrazinamide ($C_5H_5N_3O$), Adenine ($C_5H_5N_5$), Paracetamol ($C_8H_9NO_2$), Piracetam ($C_6H_{10}N_2O_2$), Theophylline ($C_7H_8N_4O_2$) and Carbamazepine ($C_{15}H_{12}N_2O$). The chemical connectivity diagrams of the selected polymorphs are illustrated in Figure 6.1. Theophylline is the only compound that is present in our set but was not included in the molecule set in the work of Dybeck *et al* [63]. Theophylline is observed in one monohydrate form and four anhydrous forms (I-IV), out of which form III is metastable [139, 140]. The enantiotropic transition temperatures of Theophylline are taken from the work of Khamar *et al* [137]. In that work, the polymorphic stability relation of the stable anhydrous forms of Theophylline was examined using DSC and hot stage microscopy. The monoclinic form IV was reported to be the most stable polymorph at room temperature. Form II and IV are enantiotropically related, with form II found to be the most stable form between 483 K and 541 K. The high-temperature form I is considered to be the most stable between 541 K and the melting point via an enantiotropic transition from form II to form I. The melting points of form I and form II were reported to be 549 K and 546 K, respectively [140]. In summary, the benchmark set employed in this study consists of five polymorphic pairs and one triplet corresponding to fifteen crystal structures. The transition temperatures of the enantiotropically related forms can be found in Table 6.1 alongside other details, such as CSD RefCodes, names of forms, and Z/Z' numbers. Note that the low-temperature form is shown on the top row and the high-temperature form on the bottom row for each compound.

6.2 Selection of an energy model

The application of our free energy workflow involves several user-specified settings that can significantly impact the obtained lattice energy and free energy values. These settings are related either to the execution of the algorithm or the utilised energy model. We refer to the former type of user-defined settings as algorithmic parameters. These algorithmic parameters include energy cutoff values, steps sizes used in finite differences during minimisation, among others. In this section, we initially present the algorithmic parameters consistently used throughout the calculations shown in this chapter. Subsequently, we assess the performance of various energy models to determine which combination of functionals and basis sets can accurately reproduce the experimental stability of Paracetamol polymorphs at 0 K.

6.2.1 Algorithmic parameter values

A number of algorithmic parameters are employed throughout the application of our free energy workflow. The lattice energy minimisation of the experimental forms is performed with CSO-FM and the free energy evaluation of the minimised experimental crystal structures with CrystalDynamics. In both algorithms, the FIT potential is utilised for the quantification of the repulsion-dispersion interactions, using a cutoff of $r_c^{\text{dir}} = 10$ Å.

Table 6.2: Parameters values used in CSO-FM and CrystalDynamics for the lattice energy minimisation and free energy evaluation of all experimental structures of benchmark set.

Algorithmic parameter	Symbol	Value
repulsion-dispersion cutoff	r_c^{dir}	10 Å
spline fraction	ζ_s	0.1
combined multipole rank	l	4
real electrostatic cutoff	r_c^{real}	10 Å
reciprocal electrostatic cutoff	k_c^{rec}	0.50 Å^{-1}
Ewald convergence parameter	α	0.18
torsions FD step size	-	10^{-3} rads
number of points for the GLQ	-	4

In CSO-FM, a spline fraction of $\zeta_s = 0.1$ is utilised for all intermolecular contributions to lattice energy. Electrostatic interactions are modelled with multipole moments up to hexadecapole, with a combined multipole rank equal to four ($l = 4$). A cutoff of $r_c^{\text{real}} = 10$ Å is used consistently in the two algorithms for the evaluation of the electrostatic interactions in the real space. All CSO-FM lattice energy minimisations are carried out by treating all torsions as the flexible degrees of freedom employing finite difference step size of 10^{-3} rads for torsions. The default values in the CSO-RM/CSO-FM input files are used for the rest of the algorithmic parameters, such as LAMs validity region and other options related to the minimisation procedure [40]. In

the CrystalDynamics algorithm, the same real space cutoff for the electrostatics and repulsion-dispersion interactions cutoff are utilised. The reciprocal space cutoff is chosen as $k_c^{\text{rec}} = 0.50$ Å^{-1} with $\alpha = -0.18$. Four points per direction are used for the Gauss-Legendre quadrature. The free energies are calculated with a temperature step of 10 K. All the algorithmic parameters values used in CSO-FM and CrystalDynamics are summarised in Table 6.2 aiming to facilitate the reproducibility of calculations shown in this chapter. In the next section, we examine the most suitable energy model to be employed in the lattice energy minimisation and free energy evaluation stages.

6.2.2 Choosing an energy model: Paracetamol

The reproduction of the experimental relative stability of enantiotropically related forms, using vibrational free energy as the stability metric, requires two criteria to be met. First, the computed free energy of the low-temperature form should be lower than that of the high-temperature form at 0 K. Second, an enantiotropic transition can only be observed when the free energy of the high-temperature form has sufficiently low vibrational contributions relative to the low-temperature form, leading to a change in relative stability above a certain temperature. In cases where the first criterion is not met but the second criterion is satisfied, our method is unable to capture any changes in polymorphic stability. Therefore, it is crucial that the energy model utilised in the free energy calculation accurately predicts the experimental polymorphic stability at 0 K.

In this section, we examine how the choice of the energy model impacts the computed lattice energies and free energies of Paracetamol polymorphs. Paracetamol was specifically chosen due to its moderate size compared to the other molecules in the set. The experimental forms of Paracetamol with CSD RefCodes HXACAN01 (form I) and HXACAN (form II) are used as starting points for minimisation using CSO-FM. Our objective is to identify a combination of functional/basis set that correctly predicts Paracetamol form II as the most stable polymorph at 0 K, in agreement with experimental findings [134]. To achieve this, we apply our free energy workflow to Paracetamol polymorphs using four different levels of theory: B3LYP/6-31G(d,p), PBE/6-31G(d,p), PBE0/6-311+G(d,p) and M06/6-31G(d,p). We find that the CSO-FM minimisation of Paracetamol experimental crystal structures at the M06/6-31G(d,p) level of theory could not converge, with the specific cause of this failure not further investigated. This leaves three levels of theory. The lattice energies calculated for each level of theory are shown in Table 6.3. For all combinations of functional/basis sets, form I is incorrectly predicted to be the more stable than form II using lattice energy as the stability metric. Notably, the smallest lattice energy difference between the two forms is obtained at the PBE0/6-311+G(d,p) level of theory. We can conclude that utilising lattice energy as the ranking criterion of Paracetamol polymorphs at 0 K the experimentally minimised forms is insufficient.

Continuing our analysis, we move on to calculate the free energy of the CSO-FM minimised structures using the CrystalDynamics algorithm. The vibrational contributions to free energy

Table 6.3: Lattice energies and Helmholtz free energies at 0 K of the two paracetamol polymorphs using different levels of theory.

Functional/basis set	ΔU^{latt} (kJ/mol)		ΔA(0 K) (kJ/mol)	
	form I	form II	form I	form II
B3LYP/6-31G(d,p)	-115.385	-112.056	unstable	unstable
PBE/6-31G(d,p)	-110.797	-108.177	303.913	306.920
PBE0/6-311+G(d,p)	-126.928	-125.799	303.106	302.506

include ZPE, thermal energy and entropy. At 0 K, both the thermal energy and the entropy are equal to zero. The free energy at 0 K, A(0 K), is thus given as the sum of lattice energy and ZPE. The computed values of A(0 K) using different combinations of functionals/basis sets are shown in Table 6.3. Employing the B3LYP/6-31G(d,p) level of theory fails to produce minimised structures with positive phonon frequencies. Conversely, the minimised structures emerges as stable for the other two levels of theory. Calculations are performed using PBE/6-31G(d,p) functional/basis set suggest that form I is more stable, with a free energy difference of -3.007 kJ/mol, in disagreement to the experimental stability. For the case of the PBE0/6-311+G(d,p) level of theory, the calculated A(0 K) for two polymorphs indicate that form II is more stable. In summary, the inclusion of ZPE to lattice energy correctly identifies form II having lower free energy at the PBE0/6-311+G(d,p) level of theory. The aforementioned level of theory is consistently utilised in the application of our free energy workflow to the rest of the molecules in the benchmark set.

6.3 Lattice energy minimisation of 13 polymoprhs

So far, we have introduced the compounds within the benchmark set, outlined the algorithmic parameters, and identified the energy model to be utilised in our free energy workflow. In turn, we shift our focus to optimising the unit cells of the experimental crystal structures at 0 K using the CSO-FM lattice energy minimisation algorithm. All lattice energy minimisations are carried out by treating all torsions as the flexible DoFs. Quantum mechanical calculations are performed at the PBE0/6-311+G(d,p) level of theory and the FIT potential is used to quantify the repulsion-dispersion interactions. The algorithmic parameters used in all calculations are described in Section 6.2.1.

We initially examine the ability of our energy model to replicate the crystalline geometry of the experimental forms. The differences in the crystalline packing and molecular conformation are calculated using RMSD_{15} and RMSD_{1}, respectively (see Table 6.4). In Chapter 5, we used a threshold of RMSD_{15} = 0.600 Å, below which we consider two crystal structures to be a close match. With the exception of form I of Adenine, all computed RMSD_{15} values are less than 0.300 Å, indicating a high degree of similarity between the experimental and minimised

structures. Form I of Adenine has an $RMSD_{15}$ = 0.399 Å, which is still within the matching threshold. This slightly higher value can be attributed to the presence of a hydrogen bonded motif in the crystal structures of Adenine's form I [134]. For form I, the inclusion of PCM to mimic the inductive environment might have increased the similarity between the forms, but this was not attempted to maintain consistency in the energy model applied to all compounds. The all-torsion minimisation scheme is suitable to replicate the molecular conformation of the experimental forms, as revealed by the calculated $RMSD_1$ values. Overall, both the crystal geometries and molecular conformations of the experimental and minimised forms exhibit a satisfactory degree of similarity for all compounds in the benchmark set.

Table 6.4: Geometry comparison between the experimental structure (see CSD RefCode) and locally minimised experimental structure with CSO-FM for all 13 structures in the benchmark set. Lattice energies are calculated at the PBE0/6-311+ level of theory using FIT repulsion-dispersion parameters.

Compound	RefCode	Form	$RMSD_{15}$ (Å)	$RMSD_1$ (Å)	ΔU^{latt} (kJ/mol)
Pyrazinamide	PYRZIN16	IV	0.228	0.014	-98.978
	PYRZIN20	III	0.248	0.026	-100.293
Adenine	KOBFUD	I	0.399	0.013	-153.408
	KOBFUD01	II	0.134	0.022	-143.942
Paracetamol	HXACAN	II	0.282	0.061	-125.799
	HXACAN01	I	0.316	0.029	-126.928
Piracetam	BISMEV02	III	0.131	0.040	-123.404
	BISMEV03	I	0.281	0.112	-121.138
Theophylline	BAPLOT02	IV	0.134	0.020	-127.542
	BAPLOT06	II	0.175	0.022	-131.603
	BAPLOT04	I	0.173	0.042	-122.358
Carbamazepine	CBMZPN02	III	0.261	0.078	-135.273
	CBMZPN11	I	0.191	0.030	-132.714

We turn our focus on assessing whether the lattice energies computed with CSO-FM are in agreement with the experimental stability at 0 K. The lattice energies of the experimental structures, minimised with CSO-FM are shown in Table 6.4. We remind the reader, that the polymorph with CSD RefCode shown on the top row of each compound correspond to the low-temperature form. The form with the lowest lattice energy is considered to be the most stable. The calculated lattice energies correctly indicate the experimental polymorphic stability at 0 K for Adenine, Piracetam and Carbamazepine. The biggest energy difference is calculated for form I of Adenine. As discussed above, Adenine forms a hydrogen-bonded motif, however induction effects are not taken into account in our energy models, impacting the calculated lattice

energies. For Pyrazinamide and Paracetamol, the lattice energies do not reproduce the stability at 0 K, with an energy difference between the two forms of 1.315 kJ/mol and 1.140 kJ/mol, respectively. As already mentioned, Theophylline has three polymorphs, with experimental studies suggesting that form IV is the most stable form at low temperatures. In disagreement with the experimental stability, the calculated stability relation is form II > form IV > form I, with form II being the most stable form based on lattice energy. Overall, the lattice energy is a suitable metric for predicting the correct ranking for half of the compounds in the benchmark set. In the next section, we employ CrystalDynamics to examine whether including vibrational free energies contributions can accurately capture the experimental stability at 0 K and elevated temperatures.

6.4 Free energy evaluation for the 13 polymorphs

In the previous section, we investigated the performance of the CSO-FM lattice energy minimisation in reproducing the experimental polymorphic stability using lattice energy as the stability metric. Here, we evaluate the vibrational free energy of all compounds in the benchmark set using the CrystalDynamics algorithm. The three main objectives in this section can be summarised as follows. First, we assess the effect of ZPE on the relative stability at 0 K. Second, we examine how the calculated relative stability is affected by temperature and whether the free energy shows the correct trend towards the stabilisation of the high-temperature form at elevated temperatures. Lastly, we investigate the ability of the HA of lattice dynamics to predict an enantiotropic transition temperatures in agreement with experimental findings.

From a computational standpoint, the identification of enantiotropic transition is challenging as accurate energy values are required to reproduce the experimental stability across a wide range of temperatures. The uncertainty in predicting enantiotropic transition temperatures has been studied by Abramov *et al.* [141]. They found that the error associated with transition temperatures calculated through virtual polymorph screening experiments ranges from 50 K to 270 K. The free energies in virtual polymorph screening [142, 143] are evaluated with QHA or PSCP and electrostatics are modelled via point charges. Despite the differences in the utilised free energy calculation methods and energy models, we expect a discrepancy between the experimental and calculated enantiotropic transition temperatures.

6.4.1 Methodology

The polymorphic stability is expressed as the free energy difference, $\Delta\Delta A(T) = \Delta A_i(T) - \Delta A_j(T)$, between the Helmholtz free energies of forms i and j. Throughout all calculations shown in this chapter, form j is consistently selected to be as the low-temperature form. In other words, all $\Delta\Delta A$ values are always calculated relative to the form that is experimentally more stable at low temperatures. Therefor, a positive free energy difference ($\Delta\Delta A > 0$) indicates that form j is more stable than form i as it has lower free energy ($\Delta A_i > \Delta A_j$). Conversely, $\Delta\Delta A < 0$ indicates that the high-temperature form is more stable. Under our convention, the

correct reproduction of experimental stability requires $\Delta\Delta A$ to be positive at 0 K. The free energy difference as a function of temperature for each form is shown in Figure 6.2. The values of $\Delta\Delta A$ calculated at 0 K and 400 K are also given in Table 6.5. In addition, we also evaluate the difference in entropy between the forms, $T\Delta S$, at 400 K. We adopt the same convention as described above and $T\Delta S$ is always calculated relative to the low-temperature form. As a result, positive $T\Delta S$ values indicate that the high-temperature form has higher entropy. In the remainder of this section, we present our findings for each molecule in the benchmark set.

6.4.2 Results

Pyrazinamide: Pyrazinamide Form IV, also known as the delta polymorph, is the most stable form between 0 K to 298 K. Conversely, form III, also referred to as the gamma form, has been found to be more stable above 428 K [133]. Our free energy workflow is applied to both experimental crystal structures of Pyrazinamide, with CSD RefCodes PYRZIN20 (form III) and PYRZIN16 (form IV). Free energy differences calculated with CrystalDynamics are shown in Figure 6.2a. The calculated $\Delta\Delta A$ between 0 K and 400 K suggests that form III is more stable than form IV across the entire temperature range indicating a monotropic relation. This monotropic relation is in disagreement with experimental studies [133], where form IV was found to be more stable at low temperatures. At 0 K, the free energy difference at 0 K is around 1.0 kJ/mol. It's worth noting that if the free energy ranking had been correctly identified at 0 K, we could have captured the enantiotropic transition as our calculations correctly suggest the stabilisation of the high temperature form III at elevated temperatures.

Adenine: Adenine is one of the four nucleotide bases found in DNA and is known to exist in two forms. Experimental findings indicate that form I is the low-temperature polymorph, while form II becomes more stable above 523 K [134]. Our free energy workflow is applied to the experimental polymorphs I (KOBFUD) and II (KOBFUD01) of Adenine. The free energy differences calculated with CrystalDynamics are shown in Figure 6.2b. At 0 K, $\Delta\Delta A$ is positive indicating that form I is more stable than form II, by more than 8 kJ/mol. The calculated $\Delta\Delta A$ relative to form I decreases with increasing temperature, correctly identifying the stabilisation of form II at elevated temperatures. Note that although the free energy curves of Adenine are displayed until 400 K, we have evaluated the vibrational contributions of both forms up to 2000 K to determine the temperature at which the relative stability between forms changes. The calculated enantiotropic transition occurs at 1090 K, overestimating the experimental value by 527 K. This overestimation is attributed to the large free energy difference at 0 K. As mentioned in Section 6.3, the effect of the presence of hydrogen bonded motifs in the crystal structure of Adenine Form I may not taken into account with sufficient detail in our energy model, which neglects polarisability.

Paracetamol: We proceed with applying our free energy workflow to the two Paracetamol polymorphs. The orthorhombic polymorph II (HXACAN) has been found to be the most stable

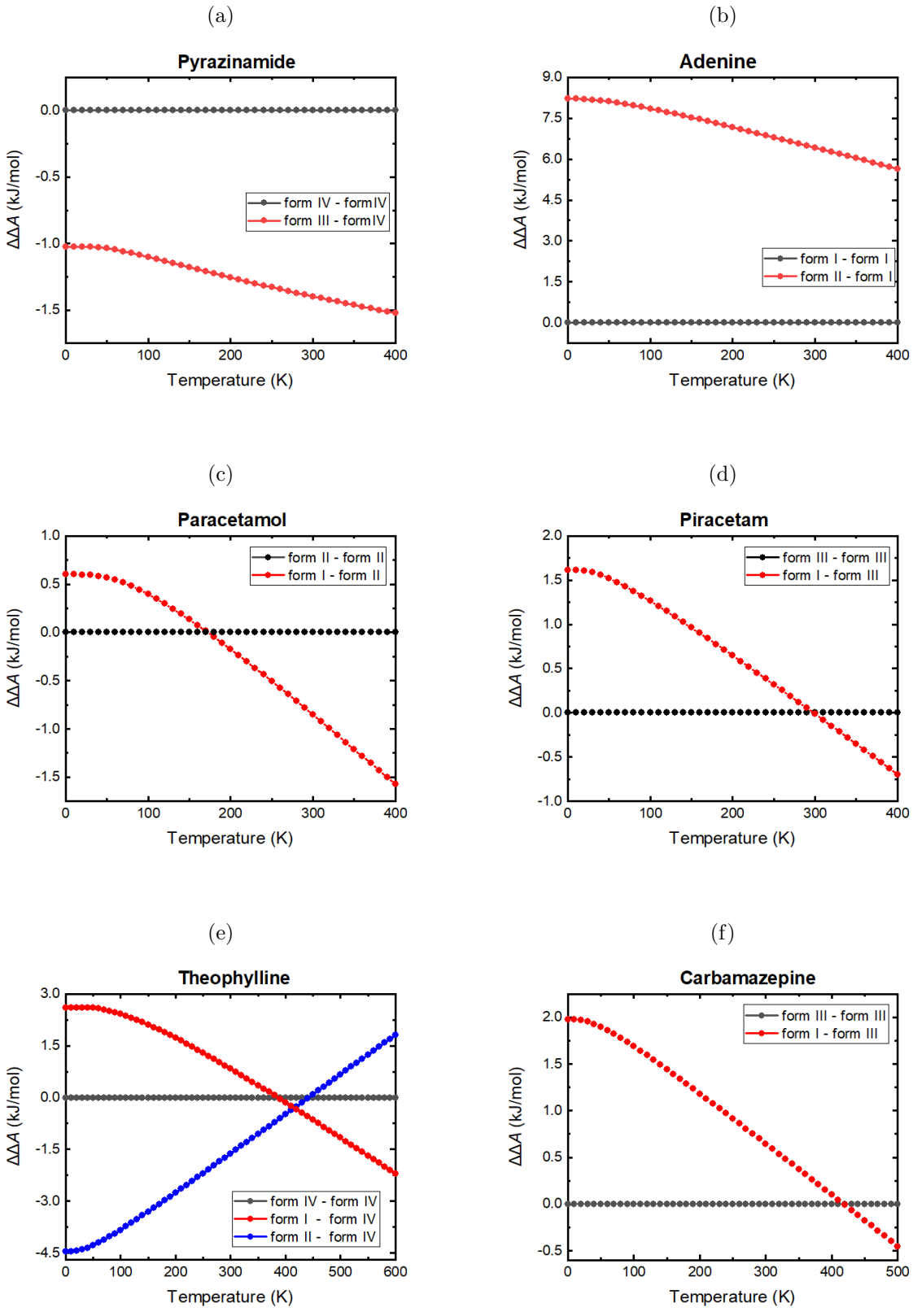


Figure 6.2: Free energy difference relative to the low-temperature form ($\Delta\Delta A$) vs temperature for (a) Pyrazinamide, (b) Adenine, (c) Paracetamol, (d) Piracetam, (e) Theophylline and (f) Carbamazepine.

form between 0 K and 108 K. At the latter temperature, the monoclinic polymorph I (HXACAN01) becomes the most stable form through an enantiotropic transition [135]. The calculated free energy differences relative to form II are shown in Figure 6.2c. As discussed on Section 6.2.2, the lattice energy model cannot predict the correct ranking at 0 K. However, the correct stability of Paracetamol's polymorphs is obtained by calculating the free energy at 0 K via the inclusion of ZPE. The inclusion of ZPE in the lattice energy yields a positive $\Delta\Delta A$ at 0 K, indicating that form II is the most stable polymorph at low temperatures in agreement with experimental findings. The calculations show that $\Delta\Delta A$ decreases with increasing temperature and an enantiotropic transition from form II to form I occurs at approximately 180 K. Overall, the stability relation between the polymorphs of Paracetamol is reproduced well and the calculated transition temperature is overestimated by around 72 K.

Piracetam: The next molecule in the set is Piracetam, for which form I (BISMEV03) and form III (BISMEV02) are examined. Form III is the low-temperature form and it undergoes an enantiotropic transition to form I at 393 K [136]. The CrystalDynamics free energies of the locally minimised forms of Piracetam also suggest that the stability relationship is enantiotropic (see Figure 6.2d). At 0 K, the correct relative stability is identified with a $\Delta\Delta A$ value around 1.6 kJ/mol. The free energy difference between the forms drops as the temperature increases yielding an enantiotropic transition temperature at 300 K. The calculated stability for Piracetam is in reasonable agreement with the experimental value of 393 K.

Theophylline: Theophylline is the only compound in the benchmark set with three polymorphs. Namely, these polymorphs are form I (BAPLOT04), form II (BAPLOT06), and form IV (BAPLOT02). Experimental studies [137] indicate that at lower temperatures, form IV is the most stable polymorph, while form II becomes more stable between 483 K and 541 K. At temperatures above 541 K, form I has been found to be the most stable form. We evaluate the free energies of the minimised experimental structures using CrystalDynamics. The relative stability of Theophylline polymorphs is shown in Figure 6.2e. At 0 K, form II has the lowest $\Delta\Delta A$ and the predicted order of stability is form II > form IV > form I. This calculated relative stability contradicts the experimental stability (form IV > form II > form I), as form II is predicted to be more stable instead of form IV. With increasing temperature, the $\Delta\Delta A$ values between forms I and IV is decreases, signifying a tendency towards the stabilisation of form I. At 390 K, the relative stability between forms IV and I undergoes a change, with form II remaining the most stable polymorph (form II > form I > form IV). This change in relative stability is important as it allows the prediction of the enantiotropic transition from form II to form I at 420 K. The experimental enantiotropic transition temperature is underestimated by approximately 120 K. In summary, although our method does not accurately reproduce the relative polymorphic stability at low temperatures, it correctly identifies the stability of form II at moderate temperatures and form I at higher temperatures.

Carbamazepine: The largest molecule included in the benchmark set is Carbamazepine, for which we examined forms I (CBMZPN11) and III (CBMZPN02). Experimental findings indicate that form III is the most stable between 0 K and 435 K, while form I becomes more stable above 435 K [138]. The calculated vibrational free energies predict the correct polymorphic stability at both low and high temperature as well as the stabilisation of form I with increasing temperature (see Figure 6.2f). The predicted enantiotropic transition occurs around 420 K, which is in excellent agreement with the experimentally reported value.

Table 6.5: Calculated $\Delta\Delta A$ at 0 K and 400 K and $T\Delta S$ at 400 K of all compounds in the benchmark set. The differences are calculated relative to the form that is known experimentally to be more stable at low temperatures. All reported energies are in kJ/mol.

Compound	Forms	$\Delta\Delta A$(0 K)	$\Delta\Delta A$(400 K)	$T\Delta S$(400 K)
Pyrazinamide	III - IV	-1.022	-1.525	0.483
Adenine	II - I	8.213	5.637	3.269
Paracetamol	I - II	0.600	-1.574	2.925
Piracetam	I - III	1.609	-0.694	2.754
Theophylline	II - IV	2.599	0.140	-4.011
	I - IV	-4.462	0.489	4.585
Carbamazepine	I - III	1.979	0.101	2.197

Overall discussion The seven experimentally-determined enantiotropic transitions are the result of a combination of five polymorphic pairs and one polymorphic triplet. In summary, the vibrational free energies computed using the HA method have proven capable of qualitatively identifying experimental enantiotropic transitions in five out of seven cases. The two compounds for which the temperature-dependence of relative stability is not reproduced qualitatively are: (i) Pyrazinamide, where the high-temperature is wrongly identified as the lowest free energy form across the whole temperature range and (ii) for Theophylline, where the predicted stability relation is incorrect at low temperatures although a better stability ranking is found at higher temperatures. With the exception of Theophylline, the free energies of the compounds of the benchmark set have been previously calculated using the QHA and MBAR and the OPLS [132] point charge force field in the work of Dybeck *et al.* [63]. In that study, in five out of twelve compounds the high-temperature was falsely identified as the most stable form at 0 K. For the remaining seven compounds, an enantiotropic transition was identified in just two cases. Although the QHA is expected to be more accurate than the HA, the vibrational free energies calculated within our workflow provide better quantitative agreement. This improved agreement can perhaps be attributed to the more accurate description of lattice energy contributions by our force field. Specifically, intramolecular energies and charge densities are calculated through gas

phase QM calculations, and multipole moments are derived from the obtained charge densities. Conversely, in the OPLS force field, intramolecular energies are described using harmonic potentials, and electrostatics are modelled with point charges, employing transferable parameters for the evaluation of all interactions.

Besides the performance of the utilized energy models, another important consideration is the number of points sampled to generate the phonon density of states (pDOS), which provides a direct link to free energy (see 4.3). In general, DFT-d methods are expected to yield more accurate lattice energies and geometries, outperforming our force field, which uses distributed multipole analysis (DMA) for describing electrostatic interactions and the Buckingham potential to model repulsion-dispersion interactions [11]. However, vibrational free energies calculated using DFT-d methods [144] typically sample a limited number of points to generate the pDOS due to the high computational cost. Our method, on the other hand, can sample $\boldsymbol{k}$-points on the order of hundreds to thousands, as the analytical expressions allow for a more dense sampling of the reciprocal space. We conclude that, despite errors associated with calculated lattice energies, when polymorphic stability is reproduced at 0 K, our free energy workflow is capable of predicting enantiotropic transition temperatures with quantitative accuracy. Apart from the utilised energy model, the assumptions of the harmonic approximation (HA) (see Section 2.3.1) limit its applicability at elevated temperatures. At lower temperatures, the potential well is expected to be parabolic, ensuring the validity of the HA. However, with increasing temperature, the effects of thermal expansion and anharmonicity become more pronounced. These effects are neglected in the HA of lattice dynamics theory, leading to decreased accuracy in the calculated free energies as temperature rises.

We further analyse the entropic contributions computed for each polymorphic form. As can be seen in Table 6.2, $T\Delta S$ (400 K) values, calculated relative to the low-temperature forms, are always positive. These positive $T\Delta S$ (400 K) values indicate that the entropy of the high-temperature forms is greater than that of their low-temperature counterparts for all examined compounds in the benchmark set. A similar conclusion was drawn in the work of Dybeck *et al.*, where the high-temperature form of all twelve compounds consistently displayed higher entropy under elevated temperature conditions, with free energies computed using the QHA and MBAR. In addition, entropy has the largest effect on the stability relation as the differences between $\Delta\Delta A$ (0 K) and $\Delta\Delta A$ (400 K) are roughly equal to $T\Delta S$ (400 K). Overall, we can conclude that the HA in conjunction with the utilised energy model (see Chapter 4) is capable of qualitatively and, in some instances, even quantitatively predicting the enantiotropic transition between polymorphs.

6.5 Conclusions

In this chapter, we have investigated the ability of our free energy workflow to reproduce the experimental relative stability of polymorphic pairs and predict the enantiotropic transition temperatures. To achieve this, we have introduced a benchmark set consisting of six compounds,

with varying size and flexibility, all featuring well-characterised experimental enantiotropic transition temperatures. Among the four examined energy models, PBE0/6-311+G(d,p) was found to correctly identify the experimental stability of Paracetamol polymorphs and was used throughout the application of our workflow to the rest of the compounds in the set. The lattice energy minimisation was carried out with CSO-FM, treating all torsions as the flexible degrees of freedom. The reproduction of the geometry was satisfactory for all the compounds in the set, as evidenced by the calculated RSMD, revealing a close match between the experimental and minimised forms. Using lattice energy as the stability metric, the relative stability of only half of the low-temperature forms was predicted in agreement with experimental data.

Free energy evaluations were performed with CrystalDynamics. All minimised experimental forms are found to display positive phonon frequencies, confirming their stability in the free energy landscape. Adding ZPE to the lattice energy correctly reproduces the experimental relative stability of four out of six polymorphic systems at 0 K. The transitions of Paracetamol, Piracetam, Adenine and Carbamazepine polymorphs as well as the transition between forms II and I of Theophylline are predicted with variable quantitative agreement. In the case of Piracetam, the experimental stability is not reproduced at 0 K. The incorrectly calculated stability relation at 0 K, combined with the predicted tendency towards the stabilisation of the high-temperature form, prevents any crossing between the free energy curves of two forms. In the case of Adenine, the large overestimation of the transition temperature can be attributed to polarisability arising from the presence of hydrogen bonded motifs in the crystal structure of form I, which is not accounted for in our force field. In the case of Theophylline, although the relative stability at low temperatures is not accurately predicted, the enantiotropic transition from form I to form II is correctly identified.

Remarkably, the vibrational free energies calculated for all molecule in the benchmark set suggest a tendency towards the stabilisation of the high-temperature form as revealed by the decrease of $\Delta\Delta A$ with increasing temperature. The evaluation of the entropy differences shows that all high-temperature forms exhibit higher entropy than their low-temperature counterparts. This observation is in agreement with free energy calculations performed with the QHA and MBAR [63]. Overall, the number of atoms and types of flexible groups do not have an effect on performance of our algorithm. We can conclude on the basis of this small set that the HA in conjunction with our energy model is able to predict the relative stability in agreement with experiments provided the energy models is sufficiently accurate *i.e.* predicts the correct stability order at 0 K and accounts for key intermolecular interactions.

Chapter 7

Conclusions

7.1 Key contributions

This thesis contributes to the development of an accurate and efficient method for evaluating the Helmholtz free energy of molecular crystals, particularly suitable for refining CSP generated landscapes. Building upon Vasileiadis's work [14], this method is extended to enable the calculation of free energy of molecules exhibiting conformational flexibility by developing a workflow that yields consistent and reliable results. The CSP studies, conducted for three molecules of varying size and flexibility, suggest that the evaluation of the free energy can be incorporated as a final refinement stage in our CSP multistage approach. As evidenced from the free energy evaluations of the benchmark set of polymorphic molecules, the computed vibrational free energy can be utilised to reproduce the relative stability and enantiotropic transition temperatures in good agreement with experimental studies.

In Chapter 2, we reviewed existing CSP methods that have employed free energies as part of their ranking procedure. Molecular simulation methods can provide a direct link to properties, such a solubility, but are restricted by the force field accuracy. As far as vibrational free energies are concerned, atomistic DFT-d methods can provide accurate results at a higher computational cost, while rigid-body force field methods achieve better computational performance by neglecting the impact of intramolecular vibrations. The gap in the literature highlighted the need of the development of an accurate yet efficient free energy calculation method that takes into account molecular flexibility.

The derivation of key equations necessary to apply the harmonic approximation of lattice dynamics to models of lattice energy used in our CSP approach has been described in Chapter 3. Our CSP tailor-made force field decomposes the lattice energy interactions into intramolecular, repulsion-dispersion and electrostatic contributions. The analytical derivation of the second derivatives of the electrostatic interactions follows this partitioning of the energy. The analytical derivatives for the electrostatic expressions increases robustness, while maintaining a low computational cost and enables the application of our lattice dynamics approach to a plethora of structures generated in CSP studies.

The CrystalDynamics algorithm was presented in Chapter 4. The implementation is based on the formalism described in Chapter 3. We employ CrystalDynamics to evaluate the Helmholtz

free energy of molecular crystals that have been previously minimised with CrystalOptimizer (rigid molecules) or CSO-FM (flexible molecules). The successful application of our workflow has been demonstarted for the two polymorphs of TCNE. Our method is able to reproduce the relative stability as a function of temperature and identify the enantiotropic transition temperature in agreement with experimental and computational studies in the literature. In addition to free energies, entropy heat capacity and thermal energy have been computed in agreement with calculations performed with DFT-d.

The method has been successfully applied to three molecules of varying size and flexibility. For all examined molecules, we observed a satisfactory agreement between the geometry of the experimental and predicted structures. Our workflow allows for capturing the effect of entropic contributions on the relative stability of predicted structures. We can conclude that the proposed approach can be used within CSP studies to rank potential crystal structures using vibrational free energy as the stability metric.

Our free energy workflow is applied to a set of 13 experimental crystal structures with well-documented enantiotropic transition temperatures. All structures, minimised with CSO-FM, displayed positive phonon frequencies. In most cases the relative experimental stability of polymorphic molecules was reproduced well both at low and high temperatures. For all examined molecules, the tendency towards the stabilisation of the polymorph that is more stable at high temperatures was predicted by our free energy calculations. When the utilised energy model reproduces the experimental stability at 0 K, our method is able to qualitatively identify the polymorph with larger entropic contributions and thus is capable of predicting reversible changes in stability of polymorphs as a function of temperature.

7.2 Direction for future studies

In this section we provide suggestions for future investigations and advances of the work shown in this thesis. An number of advances are possible in the implementation of CrystalDynamics algorithm. A parallel implementation of the sampling of the dynamical matrix could speed up the calculations significantly. In addition, the implementation of the free energy calculation inside the CSO-FM minimisation could minimise the lost numerical accuracy of CrystalDynamics to read the output geometry file of the lattice energy minimisation algorithm.

In Chapter 6, we performed a comparison of the use of different functionals. In future work, this comparison can be extended to assess the effect of repulsion-dispersion parameters on the calculated stability. FIT parameters have been derived from experimental and this results in an implicit double counting of, leading to quantitatively large errors in the calculated free energies. As an example, parameters derived at a specific level of theory can be used instead of FIT parameters.

As highlighted for Adenine, another aspect that becomes important in the case of crystals with hydrogen bonds is polarisability. Polarisability can be incorporated in our models by employing calculations that include a polarisable continuous model (PCM) instead of being

performed *in vacuo*. As an example, the free of molecules in the benchmark set, shown on Chapter 6, can be computed using repulsion-dispersion parameters derived at a specific level of theory and PCM settings.

The lattice dynamics expressions described in Chapter 4 can be used to perform calculation using the QHA. In QHA, multipole HA calculation are conducted for different unit cell volumes allowing for the minimisation of free energy with respect to volume and potentially provide a better estimate of the Gibbs free energy as the volume is not kept fixed. The volumes of the unit cells can be expanded by scaling the values of unit cell lengths. The expansion can be isotropic or anisotropic based on the shape of the unit cell.

Another consideration for future work is solubility. The evaluation of solubility requires the utilisation of a reference state that is consistent between the solid and liquid phases. The importance of an *in silico* prediction of solubility has been highlighted in a very recent publication in Nature [54].

Bibliography

[1] David H Bowskill, Isaac J Sugden, Stefanos Konstantinopoulos, Claire S Adjiman, and Constantinos C Pantelides. Crystal structure prediction methods for organic molecules: State of the art. *Annual Review of Chemical and Biomolecular Engineering*, 12(1):593–623, 2021.

[2] B Schatschneider, J Liang, S Jezowski, and A Tkatchenko. Phase transition between cubic and monoclinic polymorphs of the tetracyanoethylene crystal: the role of temperature and kinetics. *CrystEngComm*, 14(14):4656–4663, 2012.

[3] J Maddox. Crystals from first principles. *Nature*, 335(6187):201–201, 1988.

[4] J Bernstein. *Polymorphism in Molecular Crystals 2e*, volume 30. International Union of Crystal, 2020.

[5] S Chen, I A Guzei, and L Yu. New polymorphs of ROY and new record for coexisting polymorphs of solved structures. *Journal of the American Chemical Society*, 127(27):9881–9885, 2005.

[6] GJO Beran, IJ Sugden, C Greenwell, DH Bowskill, CC Pantelides, and CS Adjiman. How many more polymorphs of ROY remain undiscovered? *Chemical science*, 13(5):1288–1297, 2022.

[7] CC Pantelides, CS Adjiman, and AV Kazantsev. General computational algorithms for *ab initio* crystal structure prediction for organic molecules. *Prediction and Calculation of Crystal Structures: Methods and Applications*, pages 25–58, 2014.

[8] SL Price, DE Braun, and SM Reutzel-Edens. Can computed crystal energy landscapes help understand pharmaceutical solids? *Chemical Communications*, 52(44):7065–7077, 2016.

[9] J Bauer, S Spanton, R Henry, J Quick, W Dziki, W Porter, and J Morris. Ritonavir: an extraordinary example of conformational polymorphism. *Pharmaceutical research*, 18:859–866, 2001.

[10] DA Bardwell, CS Adjiman, YA Arnautova, E Bartashevich, SXM Boerrigter, D E Braun, AJ Cruz-Cabeza, GM Day, RG Della Valle, GR Desiraju, et al. Towards crystal structure

prediction of complex organic compounds–a report on the fifth blind test. *Acta Crystallographica Section B: Structural Science*, 67(6):535–551, 2011.

[11] AM Reilly, RI Cooper, CS Adjiman, S Bhattacharya, AD Boese, JG Brandenburg, PJ Bygrave, R Bylsma, JE Campbell, R Car, et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 72(4):439–459, 2016.

[12] SL Price. Is zeroth order crystal structure prediction (csp_0) coming to maturity? what should we aim for in an ideal crystal structure prediction code? *Faraday discussions*, 211:9–30, 2018.

[13] GJO Beran. Frontiers of molecular crystal structure prediction for pharmaceuticals and functional organic materials. *Chemical Science*, 14(46):13290–13312, 2023.

[14] M Vasileiadis. *Calculation of the free energy of crystalline solids*. PhD thesis, Imperial College London London, UK, 2013.

[15] MT Dove. *Introduction to lattice dynamics*. Cambridge university press, 1993.

[16] PW Atkins, J De Paula, and J Keeler. *Atkins' physical chemistry*. Oxford university press, 2023.

[17] GM Day. Current approaches to predicting molecular organic crystal structures. *Crystallography Reviews*, 17(1):3–52, 2011.

[18] AV Kazantsev, PG Karamertzanis, CS Adjiman, and CC Pantelides. Efficient handling of molecular flexibility in lattice energy minimization of organic crystals. *Journal of Chemical Theory and Computation*, 7(6):1998–2016, 2011.

[19] IM Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4):784–802, 1967.

[20] PG Karamertzanis. *Prediction of crystal structure of molecular solids*. PhD thesis, Imperial College London (University of London), 2004.

[21] IJ Sugden, CS Adjiman, and CC Pantelides. Accurate and efficient representation of intramolecular energy in *ab initio* generation of crystal structures. ii. Smoothed intramolecular potentials. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 75(3):423–433, 2019.

[22] M Habgood, IJ Sugden, AV Kazantsev, CS Adjiman, and CC Pantelides. Efficient handling of molecular flexibility in *ab initio* generation of crystal structures. *Journal of chemical theory and computation*, 11(4):1957–1969, 2015.

[23] DH Case, JE Campbell, PJ Bygrave, and GM Day. Convergence properties of crystal structure prediction by quasi-random sampling. *Journal of chemical theory and computation*, 12(2):910–924, 2016.

[24] P Zhang, G PF Wood, J Ma, M Yang, Y Liu, G Sun, YA Jiang, B C Hancock, and S Wen. Harnessing cloud architecture for crystal structure prediction calculations. *Crystal Growth & Design*, 18(11):6891–6900, 2018.

[25] M Yang, E Dybeck, G Sun, C Peng, B Samas, VM Burger, Q Zeng, Y Jin, MA Bellucci, Y Liu, et al. Prediction of the relative free energies of drug polymorphs above zero Kelvin. *Crystal Growth & Design*, 20(8):5211–5224, 2020.

[26] PJM Van Laarhoven and EHL Aarts. *Simulated annealing*. Springer, 1987.

[27] J Kendrick, FJJ Leusen, MA Neumann, and J van de Streek. Progress in crystal structure prediction. *Chemistry–A European Journal*, 17(38):10736–10744, 2011.

[28] F Curtis, X Li, T Rose, A Vazquez-Mayagoitia, S Bhattacharya, L M Ghiringhelli, and N Marom. Gator: a first-principles genetic algorithm for molecular crystal structure prediction. *Journal of chemical theory and computation*, 14(4):2246–2264, 2018.

[29] F Curtis, T Rose, and N Marom. Evolutionary niching in the gator genetic algorithm for molecular crystal structure prediction. *Faraday discussions*, 211:61–77, 2018.

[30] MA Neumann. Tailor-made force fields for crystal-structure prediction. *The Journal of Physical Chemistry B*, 112(32):9810–9829, 2008.

[31] AJ Stone. Distributed multipole analysis: Stability for large basis sets. *Journal of Chemical Theory and Computation*, 1(6):1128–1132, 2005.

[32] J Hafner. Ab-initio simulations of materials using VASP: Density-functional theory and beyond. *Journal of computational chemistry*, 29(13):2044–2078, 2008.

[33] R Dovesi, R Orlando, A Erba, CM Zicovich-Wilson, B Civalleri, S Casassa, L Maschio, M Ferrabone, M De La Pierre, P D'Arco, et al. CRYSTAL14: a program for the *ab initio* investigation of crystalline solids. *International Journal of Quantum Chemistry*, 114(19):1287–1317, 2014.

[34] V Blum, R Gehrke, F Hanke, P Havu, V Havu, X Ren, K Reuter, and M Scheffler. *Ab initio* molecular simulations with numeric atom-centered orbitals. *Computer Physics Communications*, 180(11):2175–2196, 2009.

[35] JP Perdew and K Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. In *AIP Conference Proceedings*, volume 577, pages 1–20. American Institute of Physics, 2001.

[36] A Tkatchenko and M Scheffler. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Physical review letters*, 102(7):073005, 2009.

[37] S Grimme, J Antony, S Ehrlich, and H Krieg. 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*, 132(15):154104, 2010.

[38] A Tkatchenko, R A DiStasio Jr, and M Car, Rand Scheffler. Accurate and efficient method for many-body van der Waals interactions. *Physical Review Letters*, 108(23):236402, 2012.

[39] SL Price, M Leslie, GWA Welch, M Habgood, L Price, PG Karamertzanis, and GM Day. Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Physical Chemistry Chemical Physics*, 12(30):8478–8490, 2010.

[40] DH Bowskill. *Reliable and Efficient Parameter Estimation Methodologies for Crystal Structure Prediction*. PhD thesis, Imperial College London London, UK, 2021.

[41] M Born and K Huang. *Dynamical theory of crystal lattices*. Clarendon press, 1954.

[42] JG Brandenburg, J Potticary, HA Sparkes, SL Price, and SR Hall. Thermal expansion of carbamazepine: Systematic crystallographic measurements challenge quantum chemical calculations. *The Journal of Physical Chemistry Letters*, 8(17):4319–4324, 2017.

[43] JL McKinley and GJO Beran. Identifying pragmatic quasi-harmonic electronic structure approaches for modeling molecular crystal thermal expansion. *Faraday Discussions*, 211:181–207, 2018.

[44] NS Abraham and MR Shirts. Thermal gradient approach for the quasi-harmonic approximation and its application to improved treatment of anisotropic expansion. *Journal of Chemical Theory and Computation*, 14(11):5904–5919, 2018.

[45] J Nyman and GM Day. Modelling temperature-dependent properties of polymorphic organic molecular crystals. *Physical Chemistry Chemical Physics*, 18(45):31132–31143, 2016.

[46] NS Abraham and MR Shirts. Adding anisotropy to the standard quasi-harmonic approximation still fails in several ways to capture organic crystal thermodynamics. *Crystal Growth & Design*, 19(12):6911–6924, 2019.

[47] J Hoja, H-Y Ko, MA Neumann, R Car, R A DiStasio Jr, and A Tkatchenko. Reliable and practical computational description of molecular crystal polymorphs. *Science Advances*, 5(1):eaau3338, 2019.

[48] J Hoja and A Tkatchenko. First-principles stability ranking of molecular crystal polymorphs with the DFT+MBD approach. *Faraday Discussions*, 211:253–274, 2018.

[49] B Monserrat, ND Drummond, and R J Needs. Anharmonic vibrational properties in periodic systems: energy, electron-phonon coupling, and stress. *Physical Review B*, 87(14):144302, 2013.

[50] J Nyman. *Computational predictions of structures, inclusion behaviour and properties of organic molecular crystals*. PhD thesis, University of Southampton, Southampton UK, 2017.

[51] J Nyman and GM Day. Static and lattice vibrational energy differences between polymorphs. *CrystEngComm*, 17(28):5154–5165, 2015.

[52] JG Brandenburg and S Grimme. Organic crystal polymorphism: a benchmark for dispersion-corrected mean-field electronic structure methods. *Acta Crystallographica Section B*, 72(4):502–513, 2016.

[53] AM Reilly and A Tkatchenko. Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals. *The Journal of Chemical Physics*, 139(2):024705, 2013.

[54] D Firaha, Y M Liu, J van de Streek, K Sasikumar, H Dietrich, J Helfferich, L Aerts, DE Braun, A Broo, AG DiPasquale, et al. Predicting crystal form stability under real-world conditions. *Nature*, 623(7986):324–328, 2023.

[55] J Hoja, AM Reilly, and A Tkatchenko. First-principles modeling of molecular crystals: structures and stabilities, temperature and pressure. *WIREs: Computational Molecular Science*, 7(1):e1294, 2017.

[56] JG Brandenburg and S Grimme. Dispersion corrected hartree–fock and density functional theory for organic crystal structure prediction. *Prediction and Calculation of Crystal Structures: Methods and Applications*, pages 1–23, 2014.

[57] D Frenkel and B Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier, 2001.

[58] T-Q Yu and ME Tuckerman. Temperature-accelerated method for exploring polymorphism in molecular crystals based on free energy. *Physical Review Letters*, 107(1):015701, 2011.

[59] E Schneider, L Vogt, and ME Tuckerman. Exploring polymorphism of benzene and naphthalene with free energy based enhanced molecular dynamics. *Acta Crystallographica Section B*, 72(4):542–550, 2016.

[60] AM Reilly, S Habershon, CA Morrison, and DWH Rankin. Simulating thermal motion in crystalline phase-I ammonia. *The Journal of Chemical Physics*, 132(13):134511, 2010.

[61] M J Schnieders, J Baltrusaitis, Y Shi, G Chattree, L Zheng, W Yang, and P Ren. The structure, thermodynamics, and solubility of organic crystals from simulation with a polarizable force field. *Journal of Chemical Theory and Computation*, 8(5):1721–1736, 2012.

[62] EC Dybeck, NP Schieber, and MR Shirts. Effects of a more accurate polarizable hamiltonian on polymorph free energies computed efficiently by reweighting point-charge potentials. *Journal of Chemical Theory and Computation*, 12(8):3491–3505, 2016.

[63] EC Dybeck, NS Abraham, NP Schieber, and MR Shirts. Capturing entropic contributions to temperature-mediated polymorphic transformations through molecular modeling. *Crystal Growth & Design*, 17(4):1775–1787, 2017.

[64] D Shivakumar, J Williams, Y Wu, W Damm, J Shelley, and W Sherman. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *Journal of chemical theory and computation*, 6(5):1509–1519, 2010.

[65] J Wang, RM Wolf, JW Caldwell, PA Kollman, and DA Case. Development and testing of a general AMBER force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.

[66] SL Mayo, BD Olafson, and WA Goddard. DREIDING: a generic force field for molecular simulations. *Journal of Physical chemistry*, 94(26):8897–8909, 1990.

[67] JW Ponder, C Wu, P Ren, VS Pande, JD Chodera, MJ Schnieders, I Haque, DL Mobley, DS Lambrecht, RA DiStasio Jr, et al. Current status of the AMOEBA polarizable force field. *The journal of physical chemistry B*, 114(8):2549–2564, 2010.

[68] P M Piaggi and M Parrinello. Predicting polymorphism in molecular crystals using orientational entropy. *Proceedings of the National Academy of Sciences*, 115(41):10251–10256, 2018.

[69] A Laio and M Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[70] R Martoňák, A Laio, and M Parrinello. Predicting crystal structures: the Parrinello-Rahman method revisited. *Physical Review Letters*, 90(7):075503, 2003.

[71] P M Piaggi, O Valsson, and M Parrinello. Enhancing entropy and enthalpy fluctuations to drive crystallization in atomistic simulations. *Physical review letters*, 119(1):015701, 2017.

[72] M Salvalaglio, C Perego, F Giberti, M Mazzotti, and M Parrinello. Molecular-dynamics simulations of urea nucleation from aqueous solution. *Proceedings of the National Academy of Sciences*, 112(1):E6–E14, 2015.

[73] PG Karamertzanis, P Raiteri, M Parrinello, M Leslie, and SL Price. The thermal stability of lattice-energy minima of 5-fluorouracil: metadynamics as an aid to polymorph prediction. *The Journal of Physical Chemistry B*, 112(14):4298–4308, 2008.

[74] T Zykova-Timan, P Raiteri, and M Parrinello. Investigating the polymorphism in PR179: A combined crystal structure prediction and metadynamics study. *The Journal of Physical Chemistry B*, 112(42):13231–13237, 2008.

[75] IJ Sugden, N F Francia, T Jensen, CS Adjiman, and M Salvalaglio. Rationalising the difference in crystallisability of two sulflowers using efficient in silico methods. *CrystEngComm*, 24(39):6830–6838, 2022.

[76] D Frenkel and AJC Ladd. New monte carlo method to compute the free energy of arbitrary solids. application to the fcc and hcp phases of hard spheres. *The Journal of chemical physics*, 81(7):3188–3193, 1984.

[77] C Vega, E Sanz, JLF Abascal, and EG Noya. Determination of phase diagrams via computer simulation: methodology and applications to water, electrolytes and proteins. *Journal of Physics: Condensed Matter*, 20(15):153101, 2008.

[78] L Li, T Totton, and D Frenkel. Computational methodology for solubility prediction: Application to the sparingly soluble solutes. *The Journal of chemical physics*, 146(21), 2017.

[79] L Li, T Totton, and D Frenkel. Computational methodology for solubility prediction: Application to sparingly soluble organic/inorganic materials. *The Journal of chemical physics*, 149(5), 2018.

[80] Michael R Shirts and John D Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of chemical physics*, 129(12), 2008.

[81] R Taylor and PA Wood. A million crystal structures: the whole is greater than the sum of its parts. *Chemical Reviews*, 119(16):9427–9477, 2019.

[82] AJ Cruz-Cabeza, GM Day, and W Jones. Structure prediction, disorder and dynamics in a DMSO solvate of carbamazepine. *Physical Chemistry Chemical Physics*, 13(28):12808–12816, 2011.

[83] AE Gray, GM Day, M Leslie, and SL Price. Dynamics in crystals of rigid organic molecules: contrasting the phonon frequencies calculated by molecular dynamics with harmonic lattice dynamics for imidazole and 5-azauracil. *Molecular Physics*, 102(9-10):1067–1083, 2004.

[84] M Habgood. Form II caffeine: a case study for confirming and predicting disorder in organic crystals. *Crystal Growth & Design*, 11(8):3600–3608, 2011.

[85] M Habgood, R Grau-Crespo, and S L Price. Substitutional and orientational disorder in organic crystals: a symmetry-adapted ensemble model. *Physical Chemistry Chemical Physics*, 13(20):9590–9600, 2011.

[86] GR Woollam, MA Neumann, T Wagner, and RJ Davey. The importance of configurational disorder in crystal structure prediction: the case of loratadine. *Faraday Discussions*, 211:209–234, 2018.

[87] DE Braun, JA McMahon, RM Bhardwaj, J Nyman, MA Neumann, J van de Streek, and SM Reutzel-Edens. Inconvenient truths about solid form landscapes revealed in the polymorphs and hydrates of gandotinib. *Crystal Growth & Design*, 19(5):2947–2962, 2019.

[88] MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheeseman, G Scalmani, V Barone, GA Petersson, H Nakatsuji, X Li, M Caricato, AV Marenich, J Bloino, BG Janesko, R Gomperts, B Mennucci, HP Hratchian, JV Ortiz, AF Izmaylov, JL Sonnenberg, D Williams-Young, F Ding, F Lipparini, F Egidi, J Goings, B Peng, A Petrone, T Henderson, D Ranasinghe, VG Zakrzewski, J Gao, N Rega, G Zheng, W Liang, M Hada, M Ehara, K Toyota, R Fukuda, J Hasegawa, M Ishida, T Nakajima, Y Honda, O Kitao, H Nakai, T Vreven, K Throssell, JA Montgomery, Jr, JE Peralta, F Ogliaro, MJ Bearpark, JJ Heyd, EN Brothers, KN Kudin, VN Staroverov, TA Keith, R Kobayashi, J Normand, K Raghavachari, A P Rendell, JC Burant, SS Iyengar, J Tomasi, M Cossi, JM Millam, M Klene, C Adamo, R Cammi, W Ochterski, RL Martin, K Morokuma, O Farkas, JB Foresman, and DJ Fox. Gaussian 16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.

[89] RA Buckingham. The classical equation of state of gaseous helium, neon and argon. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 168(933):264–283, 1938.

[90] DE Williams. Improved intermolecular force field for molecules containing H, C, N, and O atoms, with application to nucleoside and peptide crystals. *Journal of Computational Chemistry*, 22(11):1154–1166, 2001.

[91] DE t Williams and SR Cox. Nonbonded potentials for azahydrocarbons: the importance of the coulombic interaction. *Acta Crystallographica Section B: Structural Science*, 40(4):404–417, 1984.

[92] CM Breneman and KB Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials. the need for high sampling density in formamide conformational analysis. *Journal of Computational Chemistry*, 11(3):361–373, 1990.

[93] A Stone. *The theory of intermolecular forces*. oUP oxford, 2013.

[94] SL Price, AJ Stone, and M A. Explicit formulae for the electrostatic energy, forces and torques between a pair of molecules of arbitrary symmetry. *Molecular Physics*, 52(4):987–1001, 1984.

[95] BRA Nijboer and FW De Wette. On the calculation of lattice sums. *Physica*, 23(1-5):309–321, 1957.

[96] GF Signorini, R Righini, and V Schettino. Lattice dynamics of the orthorhombic phase of kclo4: Ewald's method in molecular coordinates. *Chemical physics*, 154(2):245–261, 1991.

[97] I Sugden, CS Adjiman, and CC Pantelides. Accurate and efficient representation of intramolecular energy in *ab initio* generation of crystal structures. i. adaptive local approximate models. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 72(6):864–874, 2016.

[98] PG Karamertzanis and CC Pantelides. Ab initio crystal structure prediction—i. rigid molecules. *Journal of computational chemistry*, 26(3):304–324, 2005.

[99] PG Karamertzanis and CC Pantelides. Ab initio crystal structure prediction. ii. flexible molecules. *Molecular Physics*, 105(2-3):273–291, 2007.

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

[101] P Sanschagrin. Using the csd python api for interactive analytics and data mining of the cambridge structural database. *Acta Crystallogr., Sect. A: Found. Adv.*, 73:a67, 2017.

[102] RA Mayo, A Otero-de-la Roza, and ER Johnson. Development and assessment of an improved powder-diffraction-based method for molecular crystal structure similarity. *CrystEngComm*, 24(47):8326–8338, 2022.

[103] AV Kazantsev. *Molecular Flexibility in Crystal Structure Prediction*. PhD thesis, Imperial College London (University of London), 2011.

[104] T Van Mourik, PG Karamertzanis, and SL Price. Molecular conformations and relative stabilities can be as demanding of the electronic structure method as intermolecular calculations. *The Journal of Physical Chemistry A*, 110(1):8–12, 2006.

[105] WH Press. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

[106] M Leslie. Dl_multi—a molecular dynamics program to use distributed multipole electrostatic models to simulate the dynamics of organic crystals. *Molecular Physics*, 106(12-13):1567–1578, 2008.

[107] VA Barker, LS Blackford, J Dongarra, JD Croz, S Hammarling, M Marinova, J Waśniewski, and P Yalamov. *LAPACK95 users' guide*. SIAM, 2001.

[108] SJ Clark, MD Segall, CJ Pickard, PJ Hasnip, MIJ Probert, K Refson, and MC Payne. First principles methods using castep. *Zeitschrift für kristallographie-crystalline materials*, 220(5-6):567–570, 2005.

[109] M Vasileiadis, AV Kazantsev, PG Karamertzanis, CS Adjiman, and CC Pantelides. The polymorphs of ROY: application of a systematic crystal structure prediction technique. *Acta Crystallographica Section B: Structural Science*, 68(6):677–685, 2012.

[110] J Hauke and T Kossowski. Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae*, 30(2):87–93, 2011.

[111] PM Usov, H Jiang, H Chevreau, VK Peterson, CF Leong, and DM D'Alessandro. Guest–host complexes of tcnq and tcne with cu3 (1, 3, 5-benzenetricarboxylate) 2. *The Journal of Physical Chemistry C*, 121(47):26330–26339, 2017.

[112] DA Pejaković, C Kitamura, JS Miller, and AJ Epstein. Optical control of magnetic order in molecule-based magnet mn (tcne) x y (ch 2 cl 2). *Journal of applied physics*, 91(10):7176–7178, 2002.

[113] RG Little, D Pautler, and P Coppens. X-ray structure analysis of cubic tetracyanoethylene and the length of the C≡N bond. Application of a double-atom refinement method. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, 27(8):1493–1499, 1971.

[114] SL Chaplot, A Mierzejewski, and GS Pawley. The monoclinic structure of tetracyanoethylene (tcne), c6n4, at 5,150 and 295 k; powder diffraction analysis. *Acta Crystallographica Section C: Crystal Structure Communications*, 40(4):663–666, 1984.

[115] R Mukhopadhyay, SK Deb, Amitabh Das, and SL Chaplot. Phase transitions at high pressure in tetracyanoethylene. *Solid state communications*, 149(43-44):1914–1918, 2009.

[116] J Murgich and S Pissanetzky. Temperature dependence of the 14n nuclear quadrupole resonance and the observation of the phase transition in cubic tetracyanoethylene. *The Journal of Chemical Physics*, 62(1):92–93, 1975.

[117] R Mukhopadhyay, SL Chaplot, and KR Rao. Phase transition in tetracyanoethylene. *physica status solidi (a)*, 92(2):467–474, 1985.

[118] SL Chaplot, A Mierzejewski, GS Pawley, J Lefebvre, and T Luty. Phonon dispersion of the external and low-frequency internal vibrations in monoclinic tetracyanoethylene at 5k. *Journal of Physics C: Solid State Physics*, 16(4):625, 1983.

[119] SL Chaplot. Free energy and the relative stability of the phases of solid tetracyanoethylene with pressure and temperature. *Physical Review B*, 36(16):8471, 1987.

[120] S Baroni, S De Gironcoli, A Dal Corso, and P Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Reviews of modern Physics*, 73(2):515, 2001.

[121] JA Chisholm and S Motherwell. Compack: a program for identifying crystal structure similarity using distances. *Journal of applied crystallography*, 38(1):228–231, 2005.

[122] CF Macrae, I Sovago, SJ Cottrell, PTA Galek, P McCabe, E Pidcock, M Platings, G P Shields, JS Stevens, M Towler, et al. Mercury 4.0: From visualization to analysis, design and prediction. *Journal of applied crystallography*, 53(1):226–235, 2020.

[123] JD Dunitz and A Gavezzotti. Attractions and repulsions in molecular crystals: what can be learned from the crystal structures of condensed ring aromatic hydrocarbons? *Accounts of chemical research*, 32(8):677–684, 1999.

[124] P Sacchi, SM Reutzel-Edens, and AJ Cruz-Cabeza. The unexpected discovery of the ninth polymorph of tolfenamic acid. *CrystEngComm*, 23(20):3636–3647, 2021.

[125] D-K Bučar, GM Day, I Halasz, GGZ Zhang, JRG Sander, DG Reid, LR MacGillivray, M J Duer, and W Jones. The curious case of (caffeine)·(benzoic acid): how heteronuclear seeding allowed the formation of an elusive cocrystal. *Chemical Science*, 4(12):4417–4425, 2013.

[126] M Polito, E D'Oria, L Maini, PG Karamertzanis, F Grepioni, D Braga, and SL Price. The crystal structures of chloro and methyl ortho-benzoic acids and their co-crystal: rationalizing similarities and differences. *CrystEngComm*, 10(12):1848–1854, 2008.

[127] A Berzins, A Kons, K Sarsuns, S Belyakov, and A Actins. On the rationalization of formation of solvates: Experimental and computational study of solid forms of several nitrobenzoic acid derivatives. *Crystal Growth & Design*, 20(9):5767–5784, 2020.

[128] U Westhaus, T Dröge, and R Sass. Detherm®—a thermophysical property database. *Fluid phase equilibria*, 158:429–435, 1999.

[129] F El-Kabbany, MM Abdel-Kader, S Taha, and A El-Shawarby. A study of the phase transformation in anthranilic acid single crystals. *Thermochimica acta*, 188(1):87–93, 1991.

[130] GE Hardy, WC Kaska, BP Chandra, and JI Zink. Triboluminescence-structure relationships in polymorphs of hexaphenylcarbodiphosphorane and anthranilic acid, molecular crystals, and salts. *Journal of the American Chemical Society*, 103(5):1074–1079, 1981.

[131] WH Ojala and MC Etter. Polymorphism in anthranilic acid: a reexamination of the phase transitions. *Journal of the American Chemical Society*, 114(26):10288–10293, 1992.

[132] GA Kaminski, RA Friesner, J Tirado-Rives, and WL Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28):6474–6487, 2001.

[133] S Cherukuvada, R Thakuria, and A Nangia. Pyrazinamide polymorphs: relative stability and vibrational spectroscopy. *Crystal Growth & Design*, 10(9):3931–3941, 2010.

[134] T Stolar, S Lukin, J Pozar, M Rubcic, G M Day, I Biljan, D S Jung, G Horvat, K Uzarevic, E Mestrovic, et al. Solid-state chemistry and polymorphism of the nucleobase adenine. *Crystal Growth & Design*, 16(6):3262–3270, 2016.

[135] M Sacchetti. Thermodynamic analysis of dsc data for acetaminophen polymorphs. *Journal of thermal analysis and calorimetry*, 63(2):345–350, 2000.

[136] A Maher, Å C Rasmuson, DM Croker, and BK Hodnett. Solubility of the metastable polymorph of piracetam (form ii) in a range of solvents. *Journal of Chemical & Engineering Data*, 57(12):3525–3531, 2012.

[137] D Khamar, IJ Bradshaw, GA Hutcheon, and L Seton. Solid state transformations mediated by a kinetically stable form. *Crystal growth & design*, 12(1):109–118, 2012.

[138] AL Grzesiak, M Lang, K Kim, and AJ Matzger. Comparison of the four anhydrous polymorphs of carbamazepine and the crystal structure of form i. *Journal of pharmaceutical sciences*, 92(11):2260–2271, 2003.

[139] D Khamar, I J Bradshaw, GA Hutcheon, and L Seton. Solid state transformations mediated by a kinetically stable form. *Crystal growth & design*, 12(1):109–118, 2012.

[140] UJ Griesser, M Szelagiewicz, UCh Hofmeier, C Pitt, and S Cianferani. Vapor pressure and heat of sublimation of crystal polymorphs. *Journal of thermal analysis and calorimetry*, 57:45–60, 1999.

[141] Y Abramov, G Sun, Y Zhou, M Yang, Q Zeng, and Z Shen. Solid-form transition temperature prediction from a virtual polymorph screening: A reality check. *Crystal Growth & Design*, 19(12):7132–7137, 2019.

[142] YA Abramov. Current computational approaches to support pharmaceutical solid form selection. *Organic Process Research & Development*, 17(3):472–485, 2013.

[143] G Sun, Y Jin, S Li, Z Yang, B Shi, C Chang, and Y A Abramov. Virtual coformer screening by crystal structure predictions: Crucial role of crystallinity in pharmaceutical cocrystallization. *The Journal of Physical Chemistry Letters*, 11(20):8832–8838, 2020.

[144] J Hoja, AM Reilly, and A Tkatchenko. First-principles modeling of molecular crystals: structures and stabilities, temperature and pressure. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 7(1):e1294, 2017.