Modelling

Molecular Flexibility for

Crystal Structure Prediction

Ogaga Glory Uzoh

14-01-2015

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Engineering

to

University College London

Department of Chemistry

University College London

20 Gordon Street

London

WC1H 0AJ

United Kingdom

Declaration

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

_____

Ogaga Glory Uzoh

January 2015

Abstract

In the crystal packing of molecules wherein a single bond links aromatic groups, a change in the torsion angle can optimise close packing of the molecule. The improved intermolecular interactions, U_{inter}, outweigh the conformational energy penalty, ΔE_{intra}, to give a more stable lattice energy, $E_{latt} = U_{inter} + \Delta E_{intra}$. This thesis uses this lattice energy model hierarchically in a new Crystal Structure Prediction (CSP) algorithm, *CrystalPredictor* version 1.6, which varies the low-barrier torsion angles at the start of generating hypothetical crystal structures.

The crystal structure of 1-benzyl-1*H*-tetrazole was successfully predicted in an informal 'blind test' when given the chemical diagram and the number of molecules in the asymmetric unit cell. Then, the concept of whether specific molecular fragments favour polymorphism (i.e. polymorphophore) was investigated by analysing the crystal energy landscapes of the monomorphic fenamic acid and the polymorphic derivative tolfenamic acid. The CSP results show that the polymorphophore promotes but does not guarantee polymorphism and that the substituents on the polymorphophore fragment decide the relative energies of the crystal structures.

Molecular Dynamics (MD) cannot use this lattice energy model because many *ab initio* calculations of ΔE_{intra} on a single molecule are expensive. However, the examination of the physical origin of the torsional barrier in fenamates aided the derivation of an analytical model for ΔE_{intra}. This thesis develops codes for fitting analytical intramolecular force fields to *ab initio* conformational profiles of fenamates. An intramolecular *exp*-6 atom-atom term (for the non-bonded repulsion-dispersion contributions) plus a *cosine* term (that represents the changes to the Molecular Orbitals) accurately model the *ab initio* conformational energy surfaces of fenamic and tolfenamic acids. This thesis provides a first step in extending ΔE_{intra} data generated from CSP studies to help MD on condensed phases of pharmaceutical-like organic molecules.

Dedication

I dedicate this thesis to my wife, Jodi Uzoh, my mother, Felicia Uzoh, and my father, Emmanuel Uzoh. For their endless love, support, and encouragement.

Acknowledgments

Foremost, I would like to thank my supervisor, Professor Sally Price, for her support and direction throughout my time at UCL. Her encouragement and wise words at the start of my doctorate program were vital.

My sincere gratitude to Louise Price for helping me to understand the practical aspects of the computational methods of Crystal Structure Prediction (CSP) and for the interesting and insightful discussions about research and life. My colleagues in Room G18 and friends at UCL made my time there fun, inspiring, and exciting. I would like to thank Rex Palmer for setting the informal blind test of Chapter 3 and for the subsequent collaborative paper.

Many thanks to the Centre for Doctoral Training in Molecular Modelling and Materials Science (CDTM3S) at UCL for the rigorous training they gave me during the first year, and the yearly industrial day invitation to present my research. The opportunity to work with sixth form students in the third year, sponsored by CDTM3S, was one of the most rewarding experiences I had at UCL.

My industrial sponsor, Cambridge Crystallographic Data Centre (CCDC), has been instrumental at every research stage. The encouragement received after presenting my work at the annual CCDC student day was invaluable. My industrial supervisors at CCDC—Aurora J. Cruz-Cabeza (a year and a half) and Peter Galek (two and a half years)—have been instrumental in collaborative research and publications. I appreciate their insight, energy, time, and encouragement during our many meetings.

Finally, this work would not be possible without the financial support from the CCDC and the Engineering and Physical Science Research Council (EPSRC) under the CDTM3S grant EP/G036675/1.

List of Publications

1. **Uzoh, O. G**.; Cruz-Cabeza, A. J.; Price, S. L. Is the Fenamate Group a Polymorphophore? Contrasting the Crystal Energy Landscapes of Fenamic and Tolfenamic Acids. *Cryst. Growth Des.* 2012, *12* (8), 4230-4239.
2. Spencer, J.; Patel, H.; Deadman, J. J.; Palmer, R. A.; Male, L.; Coles, S. J.; **Uzoh, O. G**.; Price, S. L. The Unexpected but Predictable Tetrazole Packing in Flexible 1-Benzyl-*1H*-Tetrazole. *CrystEngComm* 2012, *14* (20), 6441-6446.
   I carried out all the computational work and prepared the manuscript.
3. **Uzoh, O. G**.; Galek, P. T. A; Price, S. L. Towards more Accurate Force Fields for pharmaceutical molecules – Analysing the Conformational Energy Profiles of Fenamates. 2014, (*Submitted*).

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CCDC	Cambridge Crystallographic Data Centre
CSP	Crystal Structure Prediction
API	Active Pharmaceutical Ingredient
DFT	Density Functional Theory
HF/SCF	Hatree Fork/Self Consistent Field
PBE	Perdew Burke and Ernzerhof exchange-correlation functional
B3LYP	Becke, three-parameter, Lee-Yang-Parr
MP-*n*	Møller-Plessett n^{th} order of pertubabtion
PCM	Polarizable Continuum Model
DMA	Distributed Multipole Analysis
GDMA	Gaussian Distributed Multipole Analysis
BT	1-Benzyl-*1H*-Tetrazole
FA	Fenamic Acid
TA	Tolfenamic Acid
PA	Phenylamine (Aniline)
n-MA	*N*-Methylaniline
MA	Mefenamic Acid

List of Symbols

ξ	Flexible torsion angle
N_p	Number of data points
N_k	Number of fitted coefficients
N_a	Number of atoms
$\mathbf{X}_{ij}$	Regressors: $i = 1,\ldots, N_p$; $j = 1,\ldots, N_k$
$\mathbf{Y}_i$	Regressand: $i = 1,\ldots, N_p$
E_{latt}	Lattice energy
ΔE_{intra}	Intramolecular energy relative to the most stable conformation
U_{inter}	Intermolecular energy

Chapter 1. Introduction

"*May the Force be with you*"

— General Dodonna, *Star Wars*: *A New Hope*

1.1. Background

Polymorphism, the ability of a molecule to crystallise or pack in more than one structural form[1], is a phenomenon that, though common, is not completely understood. The phenomenon, which has been the subject of long and extensive research, is of crucial importance in many branches of the chemical industry. Most of all in the pharmaceutical industry, because polymorphs can show significant differences in properties such as solubility and dissolution rate[2], which in turn affect the bioavailability of drug formulations, as well as their melting point, physicochemical stability and mechanical properties[3].

Researchers recognise that the polymorphic differences among molecules can be exploited to take advantage of better compactibility or powder flow rate. In addition, a thermodynamically metastable form may be chosen to improve drug administration, as such a form may have a higher dissolution rate than a more stable polymorph[3]. A third reason for the increasing industrial interest in *polymorphism* is the need to protect intellectual property, as different patents can protect the different polymorphs of an active pharmaceutical ingredient (API).

Polymorphism is common, with about one-third of all APIs confirmed to be polymorphic[4], yet not so long ago, the concept of *polymorphism* was either treated as an annoyance or completely ignored by academics. The dismissive attitude changed after the sudden emergence of a new polymorph of the drug ritonavir, a protease inhibitor for the HIV virus. Production had to stop as Abbott Laboratories (the manufacturer) was forced to embark upon an expensive reformulation of the drug. The more stable polymorph of the API was discovered after capsules suddenly started failing dissolution tests two years after the drug had been launched on the market[5]. The ritonavir incident taught the pharmaceutical industry that the best insurance against such surprises is to conduct an exhaustive experimental search before going to market[6], a view that has since led to regulations requiring extensive screening and process control for polymorphs in APIs and inactive pharmaceutical ingredients [7].

The unexpected emergence of the new ritonavir polymorph, despite proactive attempts to discover polymorphs before the drug was marketed, forcibly reminded us that metastable polymorphs can appear to be kinetically stable[1]. If Abbot Laboratories had computed the energies of possible polymorphs for ritonavir before marketing the drug, the company may have been able to prevent a multi-million dollar problem. As Graeme Day[6] noted, 'computational methods of predicting crystal structures could play a large part in this research, both in identifying all possible polymorphic forms and judging their relative stabilities – an examination and evaluation of all possible crystal structures of a given compound is a step towards the understanding and perhaps, control of polymorphism'. Given the ubiquity and importance of polymorphs, it is clear that researchers need a computer tool that is capable of postulating possible crystal structures of organic molecules before synthesis, especially since a "standard" experiment that guarantees the finding of all crystalline forms is impossible[8,9].

Crystal Structure Prediction (CSP, Figure 1.1) is the ability to predict, using computational methods, how a molecule will crystallise (i.e. unit cell, space group and all atomic positions), given only its chemical diagram[10-12]. CSP has been a long-standing goal in computational materials chemistry for its potential to predict possible polymorphs. CSP is vital to the pharmaceutical industry because a successful search method would smooth the process of manufacturing and patenting crystalline products[13].

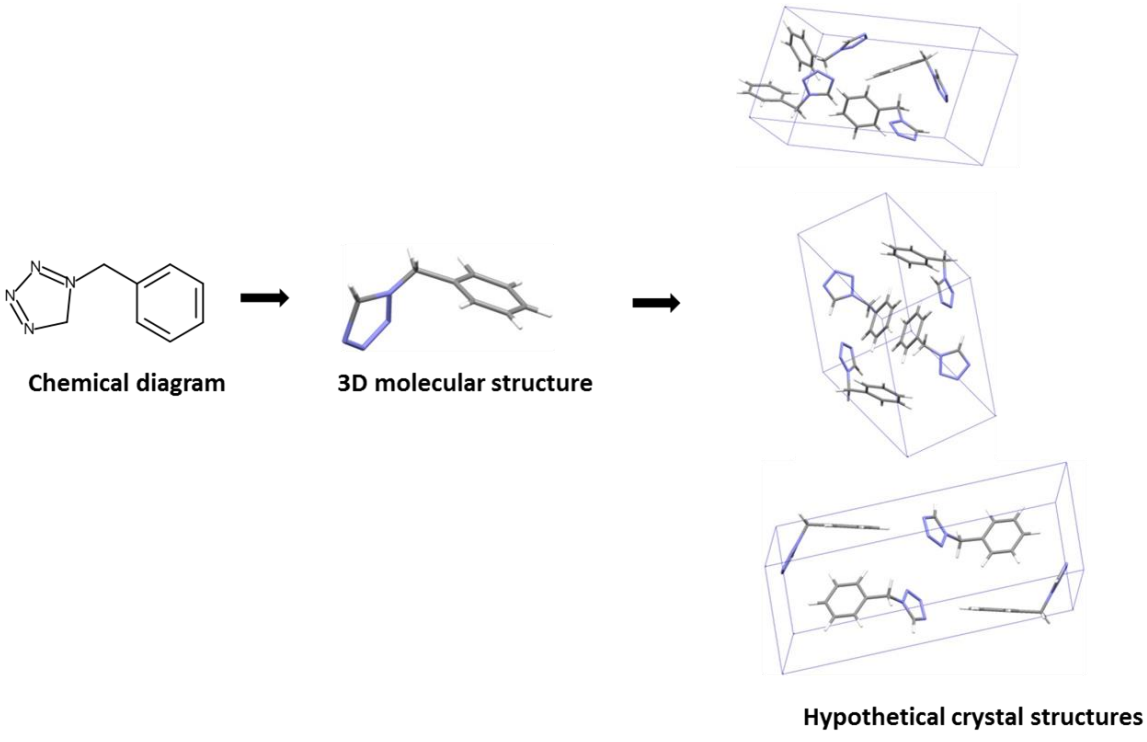


Figure 1.1: Diagrammatic illustration of crystal structure prediction (CSP) methodology.

The international blind tests[14], organised by the Cambridge Crystallographic Data Centre (CCDC), is a collaborative effort that charts the progress and state of the art in predicting the crystal structures of small organic molecules. Since the first blind test in 1999, there have been five more challenges issued by the CCDC, and Figure 1.2 shows how complex the target molecule can be. The sixth and most recent blind test[15] was issued on September 1, 2014. Scientific insight can be gained from successful CSP studies, as these methods reflect the factors that control the crystallisation process[16].

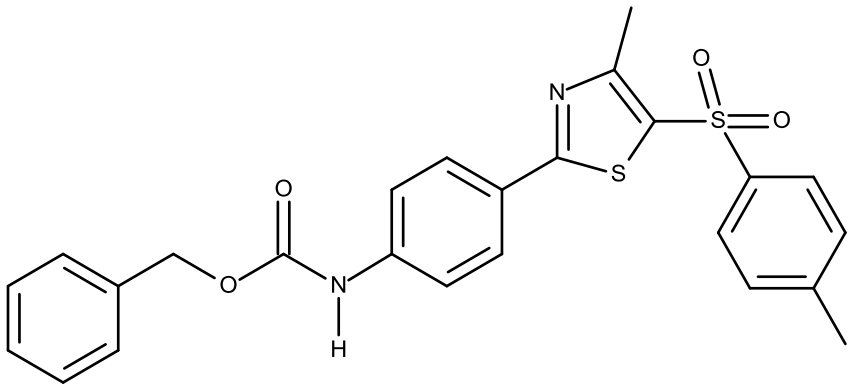
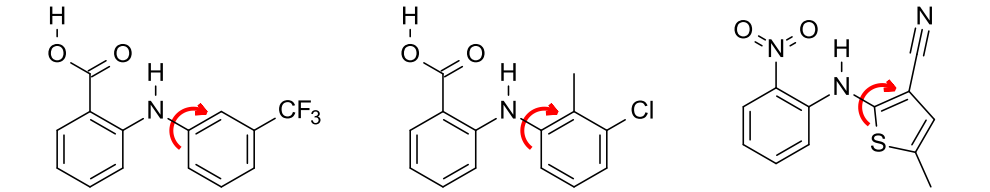


Figure 1.2: A chemical diagram of a target molecule XX from the fifth blind test[17].

Most CSP methodologies, as we shall see in section 2.8, assume that the experimentally observed crystal structures are the most stable thermodynamic structures on the crystal energy surface. In other words, the temperature effects are ignored because of poor thermal results obtained in the past for organic molecules. Therefore, a better method is needed for computing the thermal effects of organic molecules. This thesis takes the first step towards accurate thermal effects calculations and Molecular Dynamics simulations by deriving new force fields that accurately model the intramolecular energies. These discoveries on thermal effects will advance polymorph prediction.

Conformational flexibility is an important property to consider when generating hypothetical crystal structures for a flexible molecule. The modelling of conformation flexibility is the central theme of this thesis because in many organic molecules the flexibility leads to *polymorphism*. For example, flufenamic acid[18] (FFA, Figure 1.3), a non-steroidal anti-inflammatory drug (NSAID), has nine known polymorphs. Another NSAID, tolfenamic acid[19] (TA, Figure 1.3), has five polymorphs, while ROY (Figure 1.3), named after its red, orange and yellow crystals (Figure 1.4), has ten known polymorphs[20,21].



FFA[18,22] TA[19] ROY[20]

Figure 1.3: The chemical diagram of tolfenamic (TA) and flufenamic (FFA) acids.

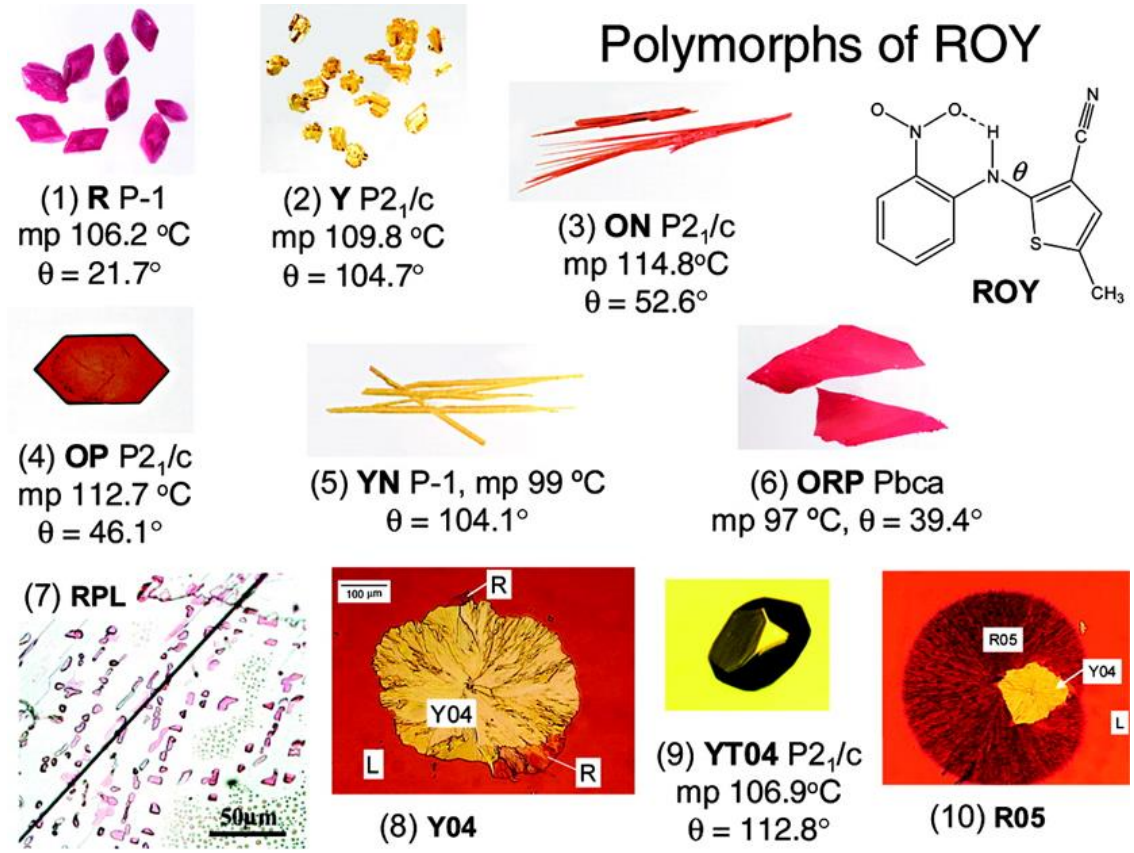


Figure 1.4: Polymorphs of ROY numbered in the order of discovery. The polymorphs have different colours, melting points, and molecular conformations (most pronounced in the torsion angle θ). (Source: Ref. [20])

Just like FFA and TA, the wide range of values for the torsion angle is also present in ROY (Figure 1.5). Therefore, the origin of polymorphism in these compounds is partly due to the *conformational flexibility* of the molecules (see Figure 1.5 and the red curly arrows in Figure 1.3). FFA, TA, and ROY are also examples of organic compounds that are considered *polymorphophores*[19,23]. A ***polymorphophore*** is a structural element that favours the formation of polymorphic crystal forms when incorporated into a molecule, i.e. molecules with similar fragments are likely to be polymorphic. Consequently, *conformational flexibility*, *polymorphism*, and *polymorphophore* are linked.

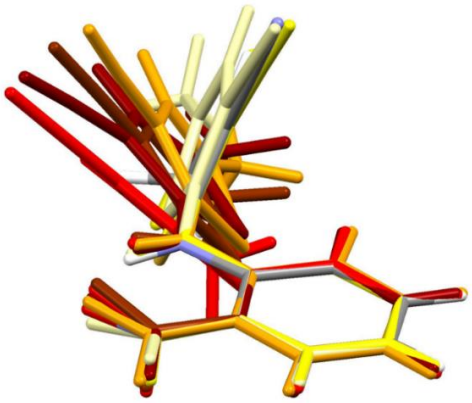


Figure 1.5: Overlay of the experimental conformers of ROY, viewed with the overlaid nitrobenzene fragment horizontal. (Source: Ref. [2])

1.2. Scope and Outline of Thesis

Chapter 2 examines the theoretical methods relevant to the work presented in this thesis. I will discuss these in several parts: intermolecular forces and their models, intramolecular forces, electronic structure methods, force field methods, least square estimation, CSP, and crystal structure comparisons. Chapters 3 to 6 contain the original work of this thesis. A CSP algorithm, *CrystalPredictor* version 1.6, is used in Chapters 3 and 4. This algorithm generates crystal structures while simultaneously varying the low-energy barrier torsion angles of the flexible molecule. Chapter 3 tests the ability of CSP to predict the crystal structure of the flexible 1-benzyl-*1H*-tetrazole molecule. This chapter is an informal 'blind test' challenge. Following this, Chapter 4 examines the *polymorphophore* concept by investigating two flexible fenamates that are similar but have different polymorphic behaviours; I use CSP and compare the two fenamates' crystal energy landscapes. Then, Chapter 5 investigates the origin of the barrier to rotation of fenamate molecules. Chapter 6 develops new analytical force field models that fit the intramolecular energy of flexible fenamate molecules as the conformation changes. This chapter investigates three analytical forms: the cosine series, rescaling repulsion and combined physical models. Finally, Chapter 7 offers conclusions and suggests areas for further research and development.

Chapter 2. Theoretical Background

"*A people without the knowledge of their past history, origin and culture is like a tree without roots*"

— Marcus Garvey

2.1. Overview

This chapter discusses the theoretical methods that are relevant to the work completed in this thesis. Central to this research is the evaluation of the lattice energy, E_{latt},—the sum of the intermolecular energy, U_{inter}, and the intramolecular energy, ΔE_{intra},—of a flexible organic molecule. Thus, section 2.2 sets out the physical basis for the lattice energy model by introducing the origin of intermolecular forces. Section 2.3 then discusses the models of intermolecular contributions of organic molecules. Section 2.4 provides a brief discussion of intramolecular forces because Chapters 5 and 6 investigate the origin of, and derive an analytical model for, the intramolecular energies of flexible molecules. Next, section 2.5 discusses some force field methods by way of laying a strong foundation for the new analytical force fields derived in Chapter 6. After this, section 2.6 discusses the fitting routines coded into the new analytical intramolecular force fields for flexible organic molecules. Section 2.7 highlights the theoretical background of the electronic structure methods used to calculate the intramolecular energy of isolated molecules (Chapters 3 to 5) and the electrostatic contribution to intermolecular energy (Chapters 3 and 4). Section 2.8 further discusses Crystal Structure Prediction (CSP), which was introduced in Chapter 1. Finally, section 2.9 discusses various tools for crystal structure comparison.

2.2. Intermolecular Forces

This section discusses the importance, relevance, origin, and categories of intermolecular forces. The intermolecular force is fundamental in the evaluation of lattice energy, which is central to CSP studies (section 2.8) in Chapters 3 and 4. Intermolecular forces are present everywhere in nature, and control properties of gases, liquids, solids, and complexes. In science, the quantitative study of intermolecular forces can reconcile diverse phenomena at the atomic and increasingly electronic level[24]. In pharmaceutical industries, the use of intermolecular forces as a computational tool for predicting new polymorphs is necessary because of the experimental problems in establishing all polymorphs that could be manufactured in a controlled way. This is essential in

preventing another ritonavir disaster (section 1.1), where a new thermodynamically stable polymorph started precipitating out of the anti-HIV drug[25].

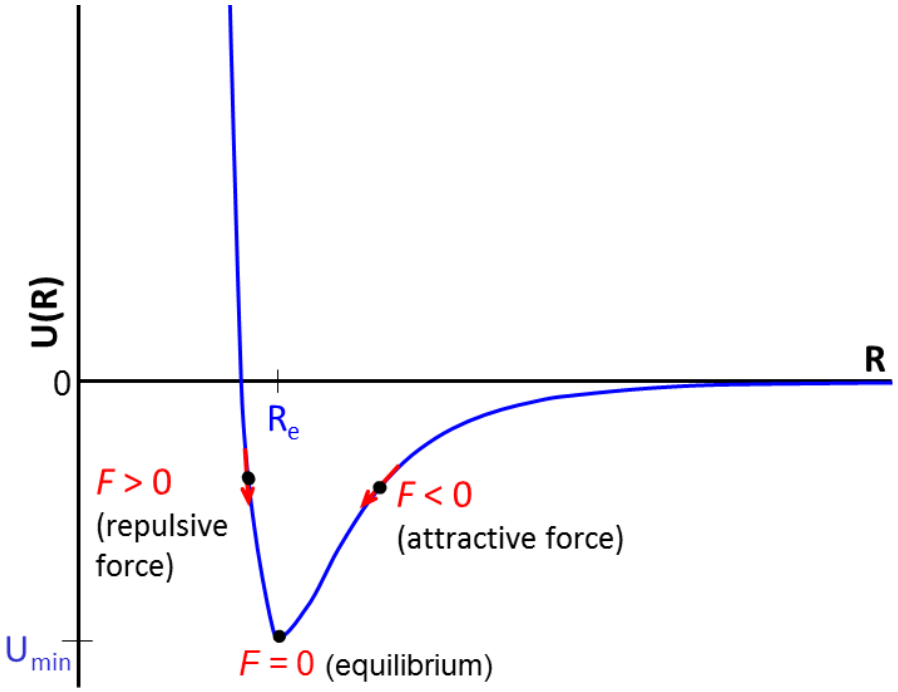


Figure 2.1: Typical intermolecular force potential energy functions for spherical atoms.

The *intermolecular pair potential*, U, is the interaction energy of a pair of molecules as a function of their separation, R, and orientation, Ω. This definition applies to rigid molecules. Intermolecular forces are repulsive at short-range and attractive at long-range. Liquids and solids exist because molecules attract one another when they are far apart[26]. The fact that water has a definite density, meaning it cannot be easily compressed to a smaller volume, shows that at short-range the forces between the molecules become repulsive[26]. The relationship between this energy of interaction, U, as a function of distance, R, takes the form shown in Figure 2.1. Figure 2.1 has an attractive region at long-range, where the force is negative, and a steep repulsive region at short-range to account for the low compressibility of condensed materials.

2.2.1. Pairwise Additivity

Due to the difficulty and computational expense of modelling intermolecular interactions, this thesis makes some approximations to the sum of the intermolecular energies. One such approximation is the pairwise additivity, which assumes that the energy of a system is the sum of all two-body interactions[26]. For instance, for three molecules A, B, and C, the pairwise approximation to the total energy is $U_{AB}+U_{BC}+U_{AC}$.

2.2.2. Physical Origin of Intermolecular Force

The significant forces between molecules have an electromagnetic origin[27]. Strong and weak nuclear forces do not contribute to the intermolecular force because they are significant over a range of the order of 10^{-4} nm; this is extremely small compared to molecular dimensions, which are typically 0.5 nm. The gravitational force, which is long-range, does not contribute to the intermolecular force because the gravitational potential energy of two atoms 0.4 nm apart is only 7 x 10^{-52} J, some thirty orders of magnitude smaller than the typical intermolecular force[28]. Consequently, the intermolecular forces have an electromagnetic origin.

The work reported in this thesis will focus on closed-shell ground state molecules. *Resonance* interactions do not occur between ordinary closed-shell molecules in their ground state. *Magnetic* interactions are several orders of magnitude smaller and are not significant to the intermolecular forces. Therefore, these two contributions will not be considered any further.

Intermolecular forces are of quantum mechanical origin and can in principle be determined by solving Schrödinger's equation and using the Born-Oppenheimer approximation (i.e. nuclei are in fixed positions). The physical origin of the intermolecular force from perturbation theory* is as follows:

$$H' = \sum_{a \in A} \sum_{b \in B} \frac{e_a e_b}{4\pi\varepsilon_o r_{ab}} \qquad 2.1$$

where e_a is the charge on particle a, one of the particles of molecule A, and r_{ab} is the distance between it and the particle b in molecule B. The H' is the perturbation on $H_A + H_B$ and consists of electrostatic interactions between the particles of a (electron and nuclei) and those of b.

The intermolecular force, F, and intermolecular energy, U, are related:

$$F = -\frac{\partial U}{\partial R} \qquad 2.2$$

where R is the distance between the spherical molecules.

When considering the intermolecular potentials, it is helpful to separate various contributions. These contributions of intermolecular energy can be separated into two

*Perturbation theory finds approximate solution to problem by starting from the exact solution of a related problem. It is can be used for describing intermolecular forces because they are relatively weak.

main types: short-range and long-range interactions. Israelachvili[29] argued against making this distinction due to the danger of counting twice, which can happen in theoretical analysis when forces with the same physical origin are both short-ranged and long-ranged.

2.2.3. Long-range Interactions

If the interacting molecules are far enough apart i.e. long-range, then the perturbation theory becomes simpler because the overlap between the wavefunctions of the two molecules can be neglected. This is called the *polarization approximation*, which was first formulated by London[30] in 1936[31]. Three main types of long-range interactions exist: *electrostatic*, *induction*, and *dispersion*. These long-range contributions vary by R^{-n}.

Electrostatic

The *electrostatic* force is a classical Coulombic interaction, which arises from the interaction between the static charge distributions of two molecules. This force can be attractive or repulsive but is strictly pairwise additive and dependent on orientation. Therefore, the electrostatic force plays an important role in determining the structure adopted by the solid state. If $\rho^{A}(\mathbf{r})$ is the charge density for molecule A, then equation 2.1 becomes

$$\int \frac{\rho^{A}(\mathbf{r})\rho^{B}(\mathbf{r})}{4\pi\varepsilon_{o}\left|\mathbf{r}-\mathbf{r}'\right|}\mathrm{d}^{3}\,\mathbf{r}\,\mathrm{d}^{3}\,\mathbf{r}' = \left\langle 0^{A}0^{B}\left|H'\right|0^{A}0^{B}\right\rangle = W'_{00} \qquad 2.3$$

The first-order energy described the electrostatic interaction. The electrostatic expression in equation 2.3 is the exact classical Coulombic interaction energy of two molecular charge distributions at long-range.

Induction

The *induction* effects, a second-order energy, arise from the distortion of a particular molecule in the electric field of all its neighbours and are always attractive. These effects are non-additive because the fields of several neighbouring molecules may reinforce or nullify each other[26]. The induction energy of molecules *A* and *B* are considered separately. The induction contribution of molecule *A* in state *m* and non-degenerate ground state *0*, is

$$U_{ind}^{A} = \sum_{m\neq 0}\frac{\left\langle 0^{A}0^{B}\left|H'\right|m^{A}0^{B}\right\rangle\left\langle m^{A}0^{B}\left|H'\right|0^{A}0^{B}\right\rangle}{W_{m}^{A}-W_{0}^{A}} \qquad 2.4$$

The induction energies are always attractive because distortion only occurs when there is a lowering of energy. Because the energies are non-additive, they are difficult to implement computationally[32]. The inclusion of the explicit induction to the

intermolecular energy can significantly reorder the relative energies of a molecule[33], especially for large molecules like carbamazepine[32]. The distributed polarizability can be used to model the induction energy, as Gareth Welch[34] did. However, this approach is computationally demanding because a large basis set and high-quality wavefunctions are often required[32] for convergence criteria to be met. In this thesis, the induction energy is not modelled explicitly. Instead, it uses PCM (Polarizable Continuum Model, section 2.8.3) to model the polarization effect that arises from the electrostatic interactions in a solvent medium[29].

Dispersion

Dispersion is a universal force of quantum mechanical origin. We know that because electrons are in continuous motion, their electron density oscillates continuously, producing instantaneous electric dipoles that change as the electron density changes. The instantaneous dipole in one molecule induces another electric dipole in the other molecule. There is an instantaneous correlation of fluctuation of charge density. This is why this contribution is universal and requires a description of electron *correlation.* Dispersion is an *attractive force* and is approximately pairwise additive. The perturbation theory expression for dispersion contribution is as follows:

$$U_{disp} = \sum_{m\neq 0, n\neq 0} \frac{\langle 00|H'|mn\rangle\langle mn|H'|00\rangle}{W_m^A + W_m^B - W_0^A - W_0^B} \qquad 2.5$$

London[30] provided the most widely understood treatment of dispersion forces, where atoms can be visualised as a set of harmonic oscillators. In this treatment, the coupling of the oscillators leads to the lowering of the energy. The dispersion energy depends on the square of the coupling constant. The leading term in the dispersion energy is proportional to R^{-6} because the coupling for a pair of interacting dipoles is proportional to R^{-3}.

2.2.4. Short-range Interactions

Exchange-Repulsion

When electron clouds of two molecules approach each other and become sufficiently close, they overlap. At short-range, the most important contributions are *exchange* and *repulsion*; these two are often taken together, despite having opposite signs, and are called *exchange-repulsion*[31]. The *exchange* contribution is the attractive energy resulting from electrons of one molecule becoming free to move over both molecules. This increases the uncertainty of the electron's position, allowing the momentum and energy to decrease. *Repulsion* contribution arises from the slight overlap of electrons that are trying to occupy the same region of space. This overlap, which is prohibited by the Pauli Exclusion

Principle, leads to redistribution of electrons and increase of the repulsion between the nuclei. The exponential repulsion domination leads to an overall *repulsive effect* and is approximately pairwise.

2.2.5. Penetration Energy and Charge Transfer

The overlap at short-range leads to modification of the long-range terms. When charge densities overlap at short-range, the electron density will no longer shield the nuclei of one molecule, and therefore, the molecules will experience a greater attraction of electron density associated with other species. The energy difference from this increased attraction is the *charge penetration*; it is the modification of the electrostatic energy at short-range and so is a first-order effect. *Charge transfer* also occurs during redistribution of electrons at short-range. It is a part of the short-range induction energy and so is a second-order effect. The energy contributions from *charge transfer* and *penetration* are small and are often negligible[35]. Some of these effects may be absorbed in the empirically fitted models of the dispersion and repulsion terms (section 2.7).

In summary, for the organic molecules studied in this thesis, repulsion dominates at short-range, while electrostatic and dispersion contributions dominate the intermolecular energy at long-range. Therefore, the total intermolecular energy can be expressed as the summation of the electrostatic, U_{elec}, repulsion, U_{repul}, and dispersion, U_{disp}, contributions:

$$U_{inter} = U_{elec} + U_{repul} + U_{disp} = U_{elec} + U_{repul-disp} \qquad 2.6$$

2.3. Modelling the Intermolecular Forces for Organic Molecules

The potential model used in this thesis for evaluating the intermolecular energy, U_{inter}, is an intermediary between conventional force fields and periodic electronic structure calculations. The functional form of conventional force fields have limitations because the same charges are used for intermolecular and intramolecular interactions[2] (section 2.5.4). Periodic electronic structure calculations are very expensive and were not available at the start of my research in 2010. For example, the plane-wave DFT-D study of selected pharmaceutical crystal structures using Quantum Espresso was published in 2013[36]. The Generation, Ranking and Characterisation Engine (GRACE[37] program, section 2.8.4), which when coupled to another periodic DFT-D code (section 2.7.5) was successful in the fourth blind test[10], is still not available to researchers as of this writing.

The work presented in this thesis models the dispersion, repulsion, and electrostatic interactions (equation 2.6) explicitly. The dispersion and repulsion contributions were modelled using a transferrable, empirically fitted model. In contrast, the electrostatic

contribution was modelled using *ab initio* wavefunction calculation on an isolated molecule for each molecular conformation due to limited transferability.

2.3.1. Repulsion-Dispersion Potentials

The dispersion and repulsion terms are modelled with an empirical *exp*-6 atom–atom potential that is assumed to be transferrable[38,39]. The repulsion term is represented by the exponential-based term, and the dispersion is represented by the inverse sixth power term:

$$U_{repul-disp} = \sum_{i,k} A_{\iota\kappa} \exp(-B_{\iota\kappa} R_{ik}) - C_{\iota\kappa} R_{ik}^{-6} \qquad 2.7$$

where atom i in molecule 1 is of type ι, atom k in molecule 2 is of type κ and they are separated by interatomic distances R_{ik}. This potential (equation 2.7) was derived by fitting the parameters to known crystal structures, and it obeys the following combining rules:

$$A_{\iota\kappa} = (A_{\iota\iota} A_{\kappa\kappa})^{\frac{1}{2}}, B_{\iota\kappa} = \frac{1}{2}(B_{\iota\iota} + B_{\kappa\kappa}), C_{\iota\kappa} = (C_{\iota\iota} C_{\kappa\kappa})^{\frac{1}{2}} \qquad 2.8$$

Parameters for carbon, non-polar H and N atoms were derived from fitting to azahydrocarbon crystal structures[39], while the parameters for the oxygen atom were derived from fitting to oxohydrocarbon crystal structures[38]. The polar H atom parameters were derived from azabenzenes, nitrobenzenes, and other simple molecules[40]. To account for the smaller effective van der Waals radius of protons, the distinction between polar and non-polar H atoms is necessary[41]. In Chapters 3 and 4, the repulsion-dispersion contributions to the intermolecular energy will be calculated using the potential of equation 2.7 and the "FIT"[34] parameters. These FIT parameters were fitted using a model that had explicit electrostatic interactions. Once the repulsion and dispersion terms were generated, they were summed over all intermolecular atom–atom distances up to a defined cut-off. This is typically 15 Å for small organic molecules for fast calculations and 30 Å for large molecules[34]. It should be noted that some of the other intermolecular contributions not modelled explicitly by equation 2.7 and any electrostatic model might have been absorbed by the empirically fitted models of dispersion and repulsion terms because these other contributions will be present to some extent in the sample crystal used in deriving the parameters A, B and C.

In contrast to "FIT", the empirical intermolecular potential by Gavezzotti[42]—another *exp*-6 potential that will be used in Chapters 5 and 6 to investigate the origin of intramolecular energy and design new analytical force fields—does not use combining rules and has no atomic or site charges.

2.3.2. Electrostatics Potential

This section discusses the two methods used in this thesis for modelling the electrostatic contribution to the intermolecular energy (equation 2.6) in Chapters 3 and 4:

(1.) The Atomic Charge Model is a crude but inexpensive approach. It uses Coulomb's law through pairwise interaction of isotropic (spherical) point charge, q, placed on each atom separated by distance R:

$$U_{elec}(\mathrm{R}_{ik}) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_k}{\mathrm{R}_{ik}} \qquad 2.9$$

where ε_0 is the dielectric vacuum constant. It assumes that the charge density of the molecule can be described by superimposing spherical atomic electron densities. The atomic charge model cannot represent non-spherical features such as lone pairs and π electrons, or short-distance directional intermolecular interactions like the geometry of the H-bond[41]. As a result, extensions to the isotropic model have been proposed to define lone pair sites as dummy atoms having isotropic electrostatic interactions[43] or at satellite positions that are optimised to provide best fit to the electrostatic field[44].

Although atomic charges can be derived from the experimental X-ray diffraction data, these results are unreliable due to the many approximations and assumptions that must be made during their calculation[45]. On the other hand, atomic charges can be generated from an *ab initio* wavefunction calculation of the isolated molecule to estimate the electrostatic contribution[41]. This involves using the CHELPG scheme (an acronym for CHarges from ELectrostatic Potentials, Grid based method), which performs the least-squares fit of the atomic charges to reproduce the molecular electrostatic potential as evaluated from *ab initio* density on a regularly spaced grid[46]. Unfortunately, higher moments such as octupoles, hexadecapoles, and so on make a small contribution to the potential, when the CHELPG scheme is used. Indeed, Francl et al.[47] showed that the CHELPG procedure is *ill conditioned* for higher atomic multipoles, so that the charges are not *well determined*; however, they did show that Singular Value Decomposition (section 2.6.2) could improve the *condition number*†.

(2.) The Distributed Multipole Moments. The electrostatic potential around a molecule can be represented by an expansion of multipole moments. At long range, this

†The condition number is the ratio of the largest to smallest singular value in the singular value decomposition of a matrix. A matrix is ill conditioned if the condition number is too large.

electrostatic potential is exact since the charge distribution is approximately spherical as the distance approaches infinity. At shorter range, the more apparent anisotropic nature of charge density is not easily represented by the central multipole expansion[48]. A better approach is to use a multiple expansion about several sites in the molecule, which can be achieved with the distributed multipole expansion.

The distributed multipole expansion is an explicit method of generating higher-order atomic multipoles that involves modelling the atomic charge distribution using a series of distributed multipoles (charge, dipole, quadrupole, octupole, hexadecapole etc., as illustrated in Figure 2.2)[26]. Consequently, the derivation of the distributed moments is essential. There are two ways of achieving this: computational analysis of the charge density, and fitting to experimental data. Deriving the multipole moments experimentally is challenging and often the data are insufficient to deduce any particular distribution of multipoles across specific sites.[49] For example, the Stark effect gives the magnitude, but not the sign, of the components of the dipole moment along the inertial axes[26].

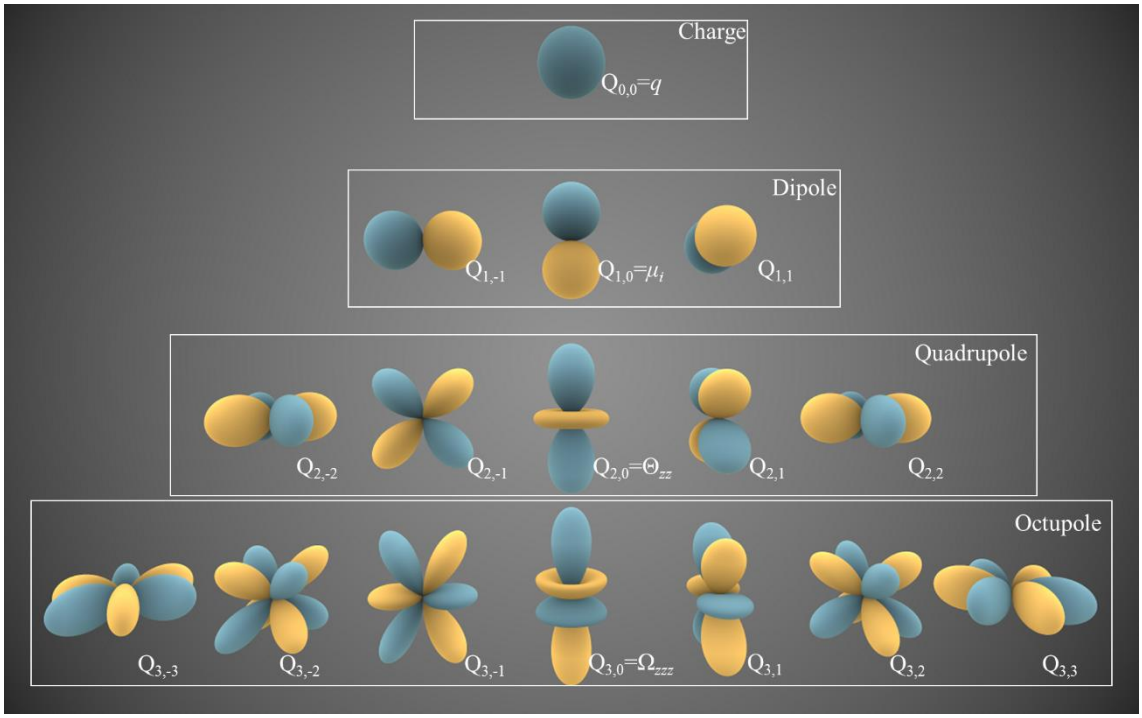


Figure 2.2: Charge distributions corresponding to pure multipoles. Blue portions represent regions where the function is positive and yellow portions represents where it is negative. $Q_{l,k}$ is the spherical harmonic multipole, where l is angular momentum quantum number and $k = l, l\text{-}1,\ldots, -l$ specifies the component of the angular momentum around the z-axis. (Source: http://en.wikipedia.org/wiki/Spherical_harmonics [Accessed 11/01/2015]).

Distributed Multipole Analysis (DMA) is a systematic way of determining the distributed moments[50]. This computational approach derives charge distribution from an

ab initio wavefunction[51] and describes it in terms of distributed multipoles (Figure 2.2) located at a number of site in the molecule using the Gaussian Distributed Multipole Analysis (GDMA)[52] program. The GDMA program carries out a DMA of the density matrix (.Fchk) file produced by GAUSSIAN[53] for the isolated molecule in the molecular axis system[34]. This thesis uses version 2[52] of GDMA because it handles diffuse functions more satisfactorily[51] than version 1[54].

In contrast to the atomic charge model, the distributed multipole moments have successfully predicted the directional (anisotropic) lone pair interactions, π-π stacking arrangements in aromatic rings and electrons and H-bond geometries in molecular organic crystal[55,56]. The long-range interactions, which include the charge–charge (R^{-1}), charge–dipole (R^{-2}) and dipole–dipole (R^{-3}) terms, are evaluated using the Ewald summation method[57], while the remaining long-range interactions are summed by direct method.

In summary, the electrostatic contribution is modelled by using either an atomic charge model or distributed multipole moments obtained from *ab initio* methods, while the repulsion and dispersion contribution are modelled using an empirical *exp*-6 potential.

2.4. Intramolecular Energy

The lattice energy of a crystal structure is the sum of the intermolecular and intramolecular energies. The focus of this section is the intramolecular part of the lattice energy. In Chapters 5 and 6, I investigate the origin and derive analytical models for the intramolecular energies of flexible molecules (Chapter 6). Therefore, this section concisely examines the intramolecular energy of flexible molecules.

The intramolecular energy of rigid molecules is negligible, so the early work of crystal structure prediction (section 1.1) focussed on the easier problem of rigid molecules. In this thesis, a ***flexible*** molecule has one or more internal degrees of freedom that change between phases, for example, torsion angles around single bonds. The ***intramolecular energy*** for a *flexible* molecule is the energy difference between the molecule in a specific conformation (e.g. in a crystal) and in the lowest energy of the isolated gas phase conformation. As evident from ROY in Chapter 1, polymorphic forms can exhibit very different torsion angles (Figure 1.5), i.e. can display *conformational polymorphism*.

Cruz-Cabeza and Bernstein[58] argued that, when referring to conformations in crystal structures, a distinction should be made between *conformational adjustment* and *conformational change*. ***Conformational adjustment*** arises when a flexible molecule adjusts to the crystal environment by slightly varying the conformation to minimise the

lattice energy of the crystal while ***conformational change*** involves a change of gas phase conformer (Figure 2.3). In the context of the polymorphism of flexible molecules, Cruz-Cabeza and Bernstein[58] referred to two polymorphs whose independent conformations are related by conformational change as conformational polymorphs. For example, crystal structures A and B in Figure 2.3 are conformational polymorphs.

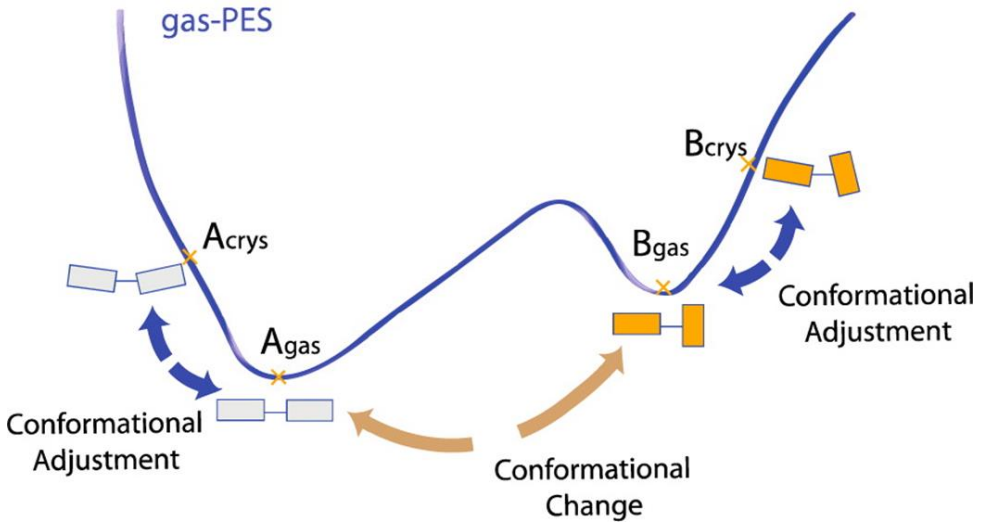


Figure 2.3: Schematic representation of conformational change and conformational adjustment. (Source: Ref. [58]).

The simplest approach to calculate the intramolecular energies is to use force field methods (section 2.5), which involve summing the contributions that arise from bond stretching, angle bending, and dihedral torsion angles. However, despite the speed and successes of force field methods in Molecular Dynamics and in modelling proteins[59], the molecular mechanics descriptions of the intramolecular energies and molecular geometries are not sufficiently accurate for evaluating the lattice energy of crystal structures (section 2.5.4).

Another method for evaluating intramolecular energy is electronic structure calculation (section 2.7). This approach avoids the problem encountered in force field methods but at the cost of higher computing time. The conformations of flexible molecules and their relative energies can be sensitive to the electron correlation method (because of the important dispersion contribution) and to the basis set that is used (since large basis sets are required to avoid intramolecular BSSE[60] during electronic structure calculations). The errors from intramolecular dispersion (section 2.7.5) can lead to the failure of quantum mechanical calculations for molecules that adopt a folded conformation[60,61].

One way of solving this problem is the dispersion-corrected (–D, section 2.7.5) method. Warnings of the intramolecular energy sensitivity toward the electronic structure level of

theory are relevant to the CSP studies of Chapter 3 and 4, where the focus is on other aspects of the computational model and not on the rigorous test of these errors. However, it is a common practise to use one electronic structure method for evaluating the intramolecular contributions toward the final lattice energies. A necessary compromise, then, is to conduct prior analysis of the effect that different levels of theory have on the intramolecular energy for a specific molecule.

The frequently used DFT level of theory, such as B3LYP/6-31G(d), may not give reliable results for molecular systems where the intramolecular dispersion energy is likely to be a major factor in determining the conformation[60]. Due to this limitation, I did not use the 6-31G(d) basis set in Chapters 3 and 4 when computing the final intramolecular and intermolecular energies of the crystal structures.

2.5. Force Fields (Atomistic Modelling)

Force field methods (also known molecular mechanics methods) ignore the molecule's electronic motions and calculate the energy of a system as a function of the nuclear positions. These methods are a cheap alternative to electronic structure methods (section 2.7). The basis of understanding the force field is the Born-Oppenheimer approximation, which assumes that the motions of the nuclei are independent of the motions of the electrons. In force field calculations, the arrangement of the electrons is assumed to be fixed.[62] Thus, traditional force field methods cannot predict properties based on electronic motions such as the breaking and formation of chemical bonds, polarization effects, and proton transfer.

The parameters determined for small molecules can be transferred to larger systems such as polymers[63] and can also be used for billions of atoms on supercomputers. Indeed, the ability to transfer force fields from one system to another—for systems that are too large for quantum mechanical calculations—is arguably the most valuable property of many force fields[63,64].

The downside to these classical force fields is that conformation and energy results are often prone to inaccuracies. These inaccuracies were observed when force fields methods were used in CSP to calculate the final lattice energies[65-67] (section 2.5.4).

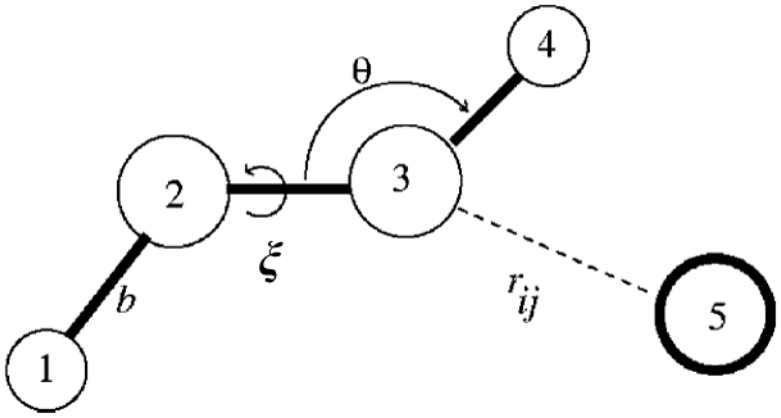


Figure 2.4: The schematic view of force field interactions. Covalent bonds are solid lines, while the dotted line indicates non-bonded interaction. (Source: Ref. [68]).

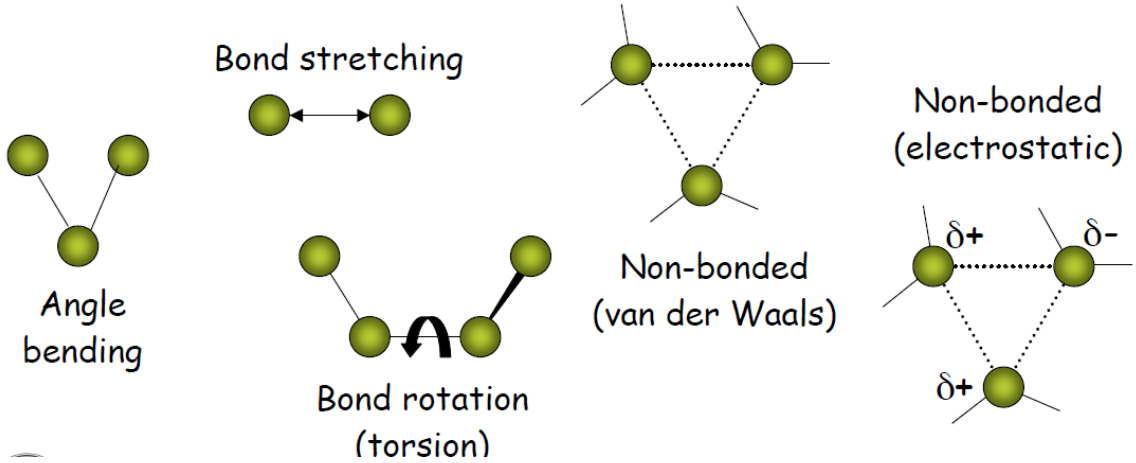


Figure 2.5: Schematic view of four key contributions to a molecular mechanics force field. These forces are bond stretching, bond stretching, angle bending, torsional terms and non-bonded interactions. (Source: Ref. [63]).

The application of force fields involves choosing an energy functional form followed by evaluating all unknown parameters. A typical force field method arises from four principal energy terms of intermolecular and intramolecular forces within the system[69,70]:

$$E_{MM} = \sum (E_b + E_\theta + E_\xi + E_{nb}) \qquad 2.10$$

where $\sum E_b$, $\sum E_\theta$, $\sum E_\xi$, and $\sum E_{nb}$ are the total energy of bond deformation, valence angle deformation, torsion deformation and non-bonded interactions (van der Waals and electrostatic) respectively. The deviation of bond, angle, and torsion from equilibrium describes the penalty of deformation (intramolecular), while non-bonded term describes the interaction between non-bonded parts of the system (intermolecular).

Hooke's law models bonds between atoms as a spring:

$$\sum E_b = \sum_{bonds} K_r (b - b_{eq})^2 \qquad 2.11$$

where K_r is the force constant and b_{eq} is the ideal bond length. A common mistake is to assume that the equilibrium value (e.g. b_{eq}) is the global minimum value; in actuality, b_{eq} is the value that b adopts when all force field terms equal zero. This Hooke's functional form is an example of applying the harmonic oscillator, in which the relationship between force and displacement is linear. A more accurate description of the relationship between force and displacement is non-linearity, i.e. anharmonic oscillator. Hence, Hooke's functional form is not as accurate as using cubic and higher terms (anharmonic terms).

Valence angles have a similar model to equation 2.11:

$$\sum E_\theta = \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \qquad 2.12$$

where K_θ is the strength of the spring holding the angle at ideal valence angle, θ_{eq}. Compared to a bond model, less energy is required to distort the valence angle away from the reference, and the force constants are smaller. The higher terms, such as cubic and quartic terms, improve the valence angle model. The 1-2 and 1-3 interactions (defined in Figure 2.4) are also known as hard degrees of freedom because large amounts of energy are required to deform bonds or angles from their equilibrium positions. However, *CrystalOptimizer*[71] (section 2.8.2)—an algorithm used in Chapters 3 and 4 to compute the lattice energy accurately—assumes that some 1-3 interactions are soft degrees of freedom. For example, hydroxyl (H–O–C) and amide (H–N–C) bond angles were treated as flexible in three model systems[72].

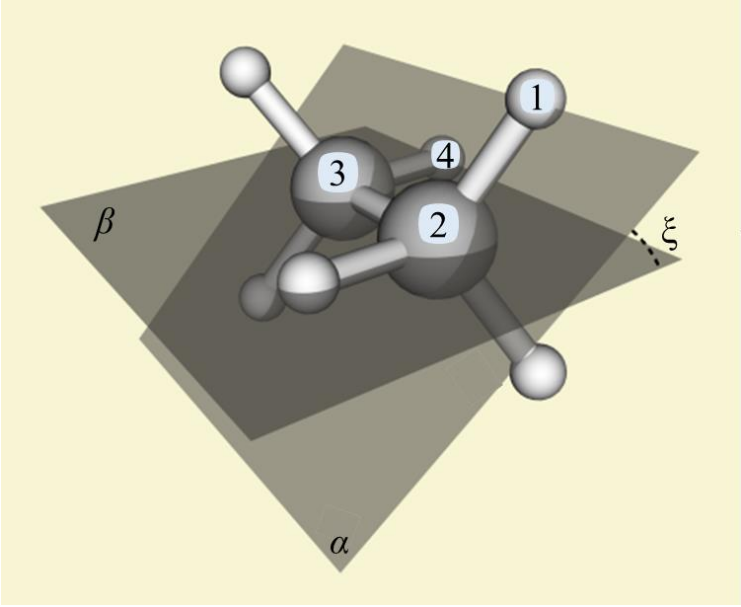


Figure 2.6: Schematic view of the torsion angle ξ using the ethane molecule. (Source: quantumwise.com/documents/manuals/vnl-2008.10_Manual/chap.molbuilder.html [Accessed 11/01/2015]).

The major change between crystal structures is their torsion angles. The torsion angle ξ in the ethane molecule shown in Figure 2.6 is defined as the intersection angle between the two planes α and β spanned by the respective atom chains 1–2–3 and 2–3–4. It is clear that the value of the torsion angle ξ between atoms 3 and 2 depends on the four atoms selected. Due to the periodic nature of torsion angles, the model is expressed as a periodic function:

$$\sum E_{\xi} = \sum_{torsions} \frac{V_n}{2}[1+\cos(\eta\xi - \gamma)] \qquad 2.13$$

where η is the periodicity and represents the number of minima as the bond rotates about 360°. The phase factor term, γ, is the torsion angle as the rotating bond passes the minimum or minima. V_n is the barrier height to rotation about torsion angle ξ. The value of V_n is larger for double bonds than for single bonds. The definition of V_n is ambiguous when more than one term is needed to model the torsional energy[64]. Even more misleading is that the 1-4 non-bonded (Figure 2.4) and higher interactions contribute to the torsional potential. The additional term with the correct qualitative behaviour of the barrier to rotation[63] is a better description of V_n.

The last term in equation 2.10 is the non-bonded interactions:

$$\sum E_{nb} = \sum_{i<j}^{nonbond} [\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} + \frac{q_i q_j}{\varepsilon r_{ij}}] \qquad 2.14$$

where the adjustable parameter σ_{ij} is the collision diameter and is defined as the separation for which the van der Waals energy is zero. The q_i and q_j terms are the partial or net atomic charges[63]. Equation 2.10 is a combination of the repulsion-dispersion (also known as the van der Waals term and is often fitted to a Lennard-Jones 6-12 potential[73] because of the computational ease) and an electrostatic term. It should be noted that intermolecular and intramolecular contributions have the same set of parameters. Many of the more recent developments in protein force fields include electronic polarizability or a solvation effect in an approximate way[68].

2.5.1. AMBER

The AMBER force fields are widely used for modelling proteins and nucleic acids[74]. AMBER, which is in continual development, has relied on the past knowledge of force fields such as ECEPP potentials[75] and Consistent Force Field (CFF[76]). The exclusion of H atoms from an earlier version of an AMBER force field was due to limited computer power. However, a compromise was adopted in later versions due to the importance of the H-bond[68].

The charges were derived by fitting partial charges to ESP (electrostatics potential) charges obtained via *ab initio* calculations (HF/STO-3G). An adapted version of van der Waals' term from amide data fitting by Lifson's group[77,78] and liquid state simulations[79] by Jorgensen was used. The first two terms (bonds and angles) of equation 2.30 were fitted from crystal structure data, while the torsional terms were extracted from either experimental data or *ab initio* calculations.

The fact that AMBER has limited parameters for organic molecules prevented this force field from being widely used in drug design and other studies of ligand–protein or ligand–DNA interactions[80]. Hence, a Generalised Amber Force Field (GAFF) was introduced by Wang et al.[80] to describe a wider variety of organic molecules, including most pharmaceutical molecules.

2.5.2. CHARMM

CHARMM[81] (Chemistry at HARvard using Molecular Mechanics) has some overlaps with AMBER, in part because the first versions of both force fields were developed around the same time. Like AMBER, H atoms were excluded in the earlier version but were later replaced with a compromise solution. Although *ab initio* calculations were used in fitting this model, a different level of theory (HF/6-31G) was employed. The interaction energies were scaled to match water dimerization energy predicted via a TIP3P model to achieve a balanced solute–water and water–water interaction. A rough continuum solvation model can be implemented by using the force field with distance-dependent dielectric constant. The second version of CHARMM, from 1985, is used today because of the respectable results obtained from solvated simulation.[68]

Subsequent developments were made in later versions of CHARMM[82,83] to improve the balance of interaction energies in solvent simulations. More recently, a force field for all atoms, often called the CHARMM general force field (CGenFF), was published[84]. This force field covers a wide range of chemical groups present in biomolecules and drug-like molecules and focusses on quality of result rather than on transferability.

2.5.3. Other Force Fields

There are more complex force fields than those discussed above. For example, cubic and quartic terms can be added to equations 2.10 and 2.11 to reflect anharmonicity and stretch-bend cross terms or the Lennard-Jones function can be replaced with a more flexible Buckingham (*exp*-6) or Morse potential[85]. Examples of such force fields are MM4[86] and Merck Molecular Force Field (MMFF)[87]. The cubic and quartic terms are also used in MM2[88].

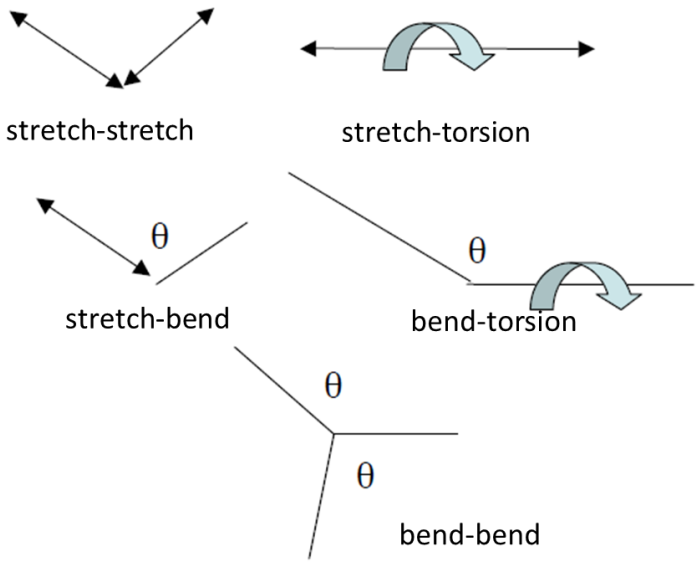


Figure 2.7: Schematic view of some of the common cross terms. (Source: Ref. [63]).

The cross terms in force fields represent coupling between internal coordinates. For example, an increase or decrease of valence angle can affect adjacent bonds. Stretch-stretch, stretch-bend, stretch-torsion, bend-torsion, and bend-bend (Figure 2.7) are examples of cross terms believed to be the most important in force fields; these cross terms in force fields are designed to predict vibrational spectra[63].

Another very popular force field is OPLS-AA (Optimized Parameters for Liquid Simulations–All Atoms), which specialises in simple liquids[85,89]. Although huge advances have been made in the understanding of biomolecular behaviour, there is still a demand for more accurate prediction of energies[90], even for proteins and nucleic acid polymers that have been studied extensively, e.g. AMOEBA[91], which uses atomic multipoles.

2.5.4. The Limits of Current Force Fields

A major limitation of current force fields is the use of atomic charges in the description of electrostatic contributions[92,93]. Another limitation is that the intermolecular and intramolecular forces have the same parameters. It is therefore not surprising that, in the blind tests of organic CSP, there have been no success based on the use of force fields for final lattice energy (sum of intermolecular and intermolecular energy) evaluations[11]. In specific cases, like aspirin, the intramolecular force field clearly has too strong a preference for the wrong conformation, so that the lowest-energy crystal structures have that conformation[94].

Indeed, some force fields' lattice energy minimizations were shown to change the conformation of the crystal structure so much that the crystal structure could not be

reproduced[65]. In other cases, the force fields were able to reproduce the crystal structures but failed to rank the energy of these crystal structures accurately. These failures have been traced to using the same charges and van der Waals interactions for the intermolecular and intramolecular forces[95].

Generally, force field potentials that are more accurate tend to have more terms because they usually yield better fit with experimental or *ab initio* data. However, additional parameters make the fitting procedure more complicated and time consuming with no scientific basis. Therefore, an ideal analytical model for the intramolecular energies should have terms that are few in number (Chapter 6), based on a consistent scientific origin (Chapter 5), and transferrable.

In summary, the traditional force field includes explicit periodic torsional potentials, and non-bonded terms, including the electrostatic term, that are applied to all atom pairs separated by three or more covalent bonds (1-4 interactions and above). Researchers built these traditional force fields on the reasonable assumption that, in biomolecules, the intramolecular non-bonded interactions are the same as intermolecular non-bonded interactions. This is clearly a limitation[2] because intramolecular atom-atom distances are often much shorter than intermolecular (van der Waals) distances. I exploited this limitation in Chapter 6 by decoupling the intermolecular and intramolecular contributions and derived new analytical intramolecular force fields.

2.6. The Fitting Routine

The Gavezzotti *exp*-6 and "FIT" potentials (section 2.3.1) and many force field methods were fitted to experimental or *ab initio* data. These potentials all depended on an assumed functional form, the ability to extrapolate, the quality of data, the range of sampling, a method of fitting (the focus of this section), and so on.

In Chapter 6, I developed codes (Appendix A) for fitting a new analytical intramolecular force field to *ab initio* conformational profiles of fenamates, which use one method of fitting. I assumed that the new analytical intramolecular force field was approximately linear. This section explains what a linear problem is, solves the problem, and outlines an alternative solution that uses a Numerical Algorithm Group (NAG) library that was implemented in the design of new analytical force fields (Appendix A).

2.6.1. Linear Least Square Method

A model is linear if given a random data of ($Y_i, X_{i,1}, \ldots, X_{i,p-1}$; $i = 1, \ldots, n$), the relationship between the response variable or observation Y_i and the explanatory or independent variable $X_{i,j}$ can be formulated as follows:

$$Y_i=\beta_0+\beta_1X_{i,1}+\beta_2X_{i,2}+\ldots+\beta_{p\text{-}1}X_{i,p\text{-}1}+\varepsilon_i,\ i=1,\ldots,n \qquad \textbf{2.15}$$

where ε_i is the error. $\beta_0, \beta_1, \beta_2, \ldots, \beta_{p\text{-}1}$ are the intercept and regression coefficients for the explanatory variables.

The linear regression model of equation **2.15** in matrix form is

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \beta_0+\beta_1X_{1,1}+\ldots+\beta_{p-1}X_{1,p-1} \\ \beta_0+\beta_1X_{2,2}+\ldots+\beta_{p-1}X_{2,p-1} \\ \vdots \\ \beta_0+\beta_1X_{n,1}+\ldots+\beta_{p-1}X_{n,p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad 2.16$$

or

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{1,1}\cdots X_{1,p-1} \\ 1 & X_{2,2}\cdots X_{2,p-1} \\ \vdots & \vdots \ \ddots \ \vdots \\ 1 & X_{n,1}\cdots X_{n,p-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad 2.17$$

or

$$\mathbf{Y}_{n\times1}=\mathbf{X}_{n\times p}\beta_{p\times1}+\boldsymbol{\varepsilon}_{n\times1} \qquad 2.18$$

where $\mathbf{Y}$ is $n\times1$, $\mathbf{X}$ is $n\times p$ (an $n\times p$ matrix has n rows and p columns), β is $p\times1$ and $\boldsymbol{\varepsilon}$ is $n\times1$ matrix.

We can solve this linear regression model by using the least squares method. The basic idea behind the least squares is minimising the sum of squares for the errors i.e. $\sum_i^n \varepsilon_i^2=\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ where T is the transpose. Rearranging equation **2.18** as a function of errors and evaluating least square problem gives

$$\begin{aligned} \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} &= (\mathbf{Y}-\mathbf{X}\beta)^T(\mathbf{Y}-\mathbf{X}\beta) \\ &= (\mathbf{Y}^T\mathbf{Y}-\mathbf{Y}^T\mathbf{X}\beta-(\mathbf{X}\beta)^T\mathbf{Y}+(\mathbf{X}\beta)^T(\mathbf{X}\beta)) \\ &= (\mathbf{Y}^T\mathbf{Y}-\mathbf{Y}^T\mathbf{X}\beta-\beta^T\mathbf{X}^T\mathbf{Y}+(\mathbf{X}\beta)^T(\mathbf{X}\beta)) \\ &= (\mathbf{Y}^T\mathbf{Y}-\mathbf{Y}^T\mathbf{X}\beta-(\beta^T\mathbf{X}^T\mathbf{Y})^T+(\mathbf{X}\beta)^T(\mathbf{X}\beta)) \\ &= (\mathbf{Y}^T\mathbf{Y}-\mathbf{Y}^T\mathbf{X}\beta-\mathbf{Y}^T\mathbf{X}\beta+\beta^T\mathbf{X}^T\mathbf{X}\beta) \\ &= (\mathbf{Y}^T\mathbf{Y}-2\mathbf{Y}^T\mathbf{X}\beta+\beta^T\mathbf{X}^T\mathbf{X}\beta) \end{aligned} \qquad 2.19$$

Then, differentiating the least square (equation 2.19) with respect to the coefficient gives

$$\begin{aligned} &\frac{\partial}{\partial\beta}(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon})=\frac{\partial}{\partial\beta}(\mathbf{Y}^T\mathbf{Y}-2\mathbf{Y}^T\mathbf{X}\beta+\beta^T\mathbf{X}^T\mathbf{X}\beta)=0 \\ &-2(\mathbf{Y}^T\mathbf{X})^T+2\mathbf{X}^T\mathbf{X}\beta=0 \\ &-2\mathbf{X}^T\mathbf{Y}+2\mathbf{X}^T\mathbf{X}\beta=0 \\ &-\mathbf{X}^T\mathbf{X}\beta=-\mathbf{X}^T\mathbf{Y} \end{aligned} \qquad 2.20$$

Then multiply both sides by $(\mathbf{X}^T\mathbf{X})^{-1}$:

$$\beta_{p\times 1} = (\mathbf{X}^T_{p\times n}\mathbf{X}_{p\times n})^{-1}\mathbf{X}^T_{p\times n}\mathbf{Y}_{n\times 1} \qquad 2.21$$

The expected/model value, $\hat{\mathbf{Y}}$, is

$$\hat{\mathbf{Y}}_{n\times 1} = \beta_{p\times 1}\mathbf{X}_{p\times n} \qquad 2.22$$

Equations 2.21 and 2.22 can be coded explicitly, which I did for cosine fits in my upgrade report[96].

2.6.2. Subroutine for Solving Least Squares

An alternative to coding equations 2.21 and 2.22 explicitly is to use the G02DDFE[97] NAG routine. This NAG routine computes the error of each coefficient and states whether a solution is unique or not, in addition to computing the coefficients. Here, I discuss some key concepts that will help in understanding the G02DDFE[97] NAG routine used in Chapter 6.

If **X** is an $n \times p$ matrix, the **rank** of **X** is the number of nonzero rows in the row echelon derived through the Gaussian elimination method. If matrix A has a rank as large as possible, then the matrix is a **full rank**; otherwise, the matrix is **rank deficient**. A matrix that is **full rank** has one unique solution. The G02DDFE[97] NAG routine can work out if any matrix is **full ranked** or **rank deficient**.

QR decomposition is the first step used for solving the least squares problem in the G02DDFE[97] NAG routine. **QR** decomposition transforms matrix **X** into a **QR** matrix that consists of an orthogonal matrix **Q** and an upper triangle matrix **R**:

$$\mathbf{X} = \mathbf{Q}\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix} = \begin{bmatrix}\mathbf{Q_1}\ \mathbf{Q_2}\end{bmatrix}\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix} = \mathbf{Q_1}\mathbf{R} \qquad 2.23$$

where $\mathbf{Q}_1$ is the first p columns of **Q** and $\mathbf{Q}_2$ is the last n-p columns of **Q**.

G02DDFE[97] subroutine calculates the regression parameters—α, β_{ik}, and γ_{ik} in Chapter 6—and Residual Sum of Squares (RSS) for a general linear regression model by minimising the sum squares residual. G02DEFE, a routine within G02DDFE, carries out **QR** decomposition of the matrix data, **X**. If **R** is of a full rank, i.e. the matrix has a rank as large as possible, then $\mathbf{R}^{-1}$ exists, and the solution of the regression model is as follows:

$$\hat{\beta}(\alpha, \beta_{ik}, \gamma_{ik}) = \mathbf{R}^{-1}\mathbf{Q}^T y \qquad 2.24$$

where **Q** and **R** are the decomposed matrices of data being rescaled and y is the *ab initio* data minus any remaining contributions not being fitted. If **R** is not full rank, Singular

Value Decomposition (SVD) of **X** (i.e. $\mathbf{X}=\mathbf{U}_1\mathbf{D}\mathbf{V}^T$) helps to extract just one solution of the linear regression model:

$$\hat{\beta}(\alpha, \beta_{ik}, \gamma_{ik}) = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}^T y \qquad 2.25$$

where **V**, **D** and **U** are respectively, the orthogonal, the non-negative diagonal element, and the orthogonal decomposed matrices. RSS and the standard error, σ_{fit}, are related:

$$\sigma_{fit} = \sqrt{\frac{RSS}{N_p - N_k}} \qquad 2.26$$

where N_p and N_k are the number of observations and regression coefficients for each of the independent variables plus the intercept.

Solving the least squares problem is a key feature of the algorithm in Chapter 6. The algorithm undertakes the challenge of deriving a new analytical model for intramolecular energies of flexible fenamate molecules. The remaining features of the algorithm are discussed in Chapter 6.

2.7. Electronic Structure Methods

The development of electronic structure methods and rapid advances in computing technology have transformed the role of quantitative numerical calculations in both theoretical and experimental studies of chemistry[98]. The first era, from the 1960s to 1980s, established the basic mathematical foundations and computational machinery of today's standard electronic structure methods. They included Hartree–Fock (HF) theory, Møller-Plesset (MP) perturbation theory, Density Functional Theory (DFT), and many others[98]. This thesis classifies HF, MP, and DFT as *ab initio* methods.

This thesis uses electronic structure methods to calculate the molecular properties of a single molecule (and, in principle, of dimers and crystals). The electronic structure methods seek to calculate the molecular properties from the first principles, i.e. solving Schrödinger's equation without using empirical data[26]. However, approximations are always necessary. DFT is a modern arsenal of quantum chemical methods[99], and to get a better appreciation of this method, it will be useful to look at more traditional wavefunction-based approaches. One of the simplest of these approaches is the Hartree Fock (HF) theory.

2.7.1. Hartree Fock Theory

The HF theory was developed to solve the electronic Schrödinger equation that results from the time-independent Schrödinger equation after invoking the Born-Oppenheimer approximation[100]. There are two good reasons for solving the Schrödinger equation. First, the solution gives the *potential energy experienced by the nuclei*, i.e. the *potential energy surface*. Second, the electronic wavefunction has lots of useful information about molecular properties such as dipole and multipole moments, polarizability, etc.

The HF method assumes that the exact *N*-body wavefunction of the system can be approximated by a single Slater determinant of *N* spin-orbitals. By employing the *variational theorem*, which states that the energy is always an upper bound to the true energy, one can derive a set of *N*-coupled equations for the *N* spin-orbitals. The solution to this equation yields the energy of the system and the wavefunction.

The HF method is called the "Self-Consistent Field (SCF) method" because the HF equations are solved using an iterative process. The SCF process starts with an initial guess for the orbitals, and successive iterations are performed with new orbitals until the *self-consistent* condition is achieved, i.e. until the solution no longer changes.

2.7.1. Molecular Orbital Theory

With the new Quantum Mechanics (QM) of Heisenberg and Schrödinger came two general theories: the Valence Bond (VB) and the Molecular Orbital (MO) theories. These two theories, developed at around the same time, quickly became rivals[101]. See ref[102] for a review of the classical valence bond approach. The focus of this section is the MO theory.

The ***MO theory***, based on the HF theory, theorises that each electron's motion can be described as a single-particle function (orbital) that does not depend on the instantaneous motions of the other electrons. MOs are calculated using the Linear Combination of Atomic Orbitals (LCAO), which is a quantum superposition of Atomic Orbitals (AOs). MO is a theory of bonding that can give insight into the forces involved in making and breaking chemical bonds, which are often the focus of organic chemistry.

The early complexities of Quantum Mechanics led to the visualisation of MOs that are easier to understand than numerical orbital coefficients[103]. Using SCF calculations, we can represent MOs using isosurfaces that correspond to the electron density. The two isosurfaces of particular importance are usually the Highest Occupied Molecular Orbital (HOMO) and the Lowest Unoccupied Molecular Orbital (LUMO). These are known as the frontier-orbitals[104]. It is not clear why the frontier-orbital theory is so successful[26], but

this theory is used extensively in computational chemistry to understand many phenomena. Frontier orbitals will be used in Chapter 5 of this thesis to understand the origin of barriers to rotation for flexible molecules.

HOMO is the isosurface from which it is energetically easiest to remove electrons from (e.g. electron density donation to form a bond), as well as the most relevant MO in this thesis. The only requirement of the empty MOs is that they form an orthogonal set with the filled MOs. These unoccupied MOs are not optimised during the SCF process because the energy of the molecule does not depend on them. As a result, these unoccupied MOs, including LUMOs, are less reliable than occupied MOs such as the HOMOs. Thus, Chapter 5 will focus on the HOMO isosurfaces of flexible molecules.

2.7.2. Correlation Effects

The major limitation of the HF method is that it treats each electron as moving under the influence of the average effect of all other electrons; the method does not take into account the electrons' instantaneous repulsion[26] (*correlation* effect). One of the most important effects of electron *correlation* is the *dispersion* interaction. In fact, the *dispersion* interaction (section 2.2) is a *correlation* effect[26,100]. The *correlation* energy is the difference between HF energy in a complete basis (HF limit) and the exact energy excluding the relativistic effects, i.e. the electrons are propagating at speeds much slower than light.

The HF method is now used merely as a starting point for more elaborate 'post-HF' *ab initio* quantum chemical approaches[99] because the poor performance of the HF method offsets its efficiency and simplicity. *Correlation* effects modify the *electrostatic* interaction between the molecules due to changes in electron distribution, and they tend to reduce charge separations. Electron *correlation* can be handled in quantum chemical theory calculations using Møller-Plesset Perturbation Theory (MPPT), Density Functional Theory (DFT)[105], etc.

2.7.3. Møller-Plesset Perturbation Theory

In perturbation theory, we start from a known solution and add a perturbation Hamiltonian (equation 2.1). Møller and Plesset[106] used perturbation theory to determine the c*orrelation* energy (small perturbation) after solving the HF problem (the known solution). *Correlation* corrections appear at the second order and beyond. The zeroth-order energy is the sum of the occupied orbital (e.g. HOMO). The first-order energy is the HF energy[26]. MPn is the n^{th}-order energy ($n \geq 2$). In this thesis, I did not perform

geometry optimisations using MP2 methods for molecule(s) with 20 or more atoms due to the computational expense.

Although MP2 methods are used as the standard, they have limitations. For example, van Mourik[61] noted that the MP2 description of intramolecular interactions with a π-electron system for flexible molecules is affected by a large intramolecular basis set superposition error (BSSE). This large BSSE of MP2 was responsible for the wrong order of stability for Phe-Gly-Phe tripetide[107].

BSSE is relevant in this thesis because many of the MP2 calculations were carried out on flexible molecules that contained at least one aromatic ring (Chapters 3 to 5). The origin of this error lies in the possibility that the unused basis functions of the second unit *B* in the associated complex *A*···*B* may augment the basis set of the first unit *A*, thereby lowering unit *A*'s energy compared to a calculation of this unit alone[26,108]. This lowering can be attributed to the variation principle that arises from the mutual enlargement of the basis set of each monomer by the presence of the orbitals of its partner molecule[109]. Rather than using an infinite basis set, an impossible task, the method for correcting this error is to use a counterpoise (CP).

2.7.4. Density Functional Theory

Density Functional Theory (DFT) is a cheaper alternative to MPPT. Indeed, P. J. Stephens et al.[110] suggested in 1994 that in the future DFT might be preferred over the MP2 method due to the lower computational demands of DFT. That future is now, for DFT is currently the method of choice for first-principles electronic structure calculations in condensed phase and complex molecular environments[111,112]. Over the last 30 years, physicists working on the electronic structure of solids, surfaces, defects, etc. have used DFT. More recently, DFT has become popular with theoretical and computational chemists[113].

DFT is a theory of electronic structure, exact in principle, that is based on the electron density distribution instead of the many-electron wave function[111]. DFT is based on a theorem, proposed by Hohenberg and Kohn[114], which states that knowledge of the electron density of a system in the ground state is enough to determine the energy: 'the energy is a functional of the density'[26,99,105]. A generalized DFT expression is

$$E_{DFT}[\rho] = E_{Ne}[\rho] + T[\rho] + E_{ee}[\rho] \qquad 2.27$$

where $T[\rho]$ is the kinetic energy associated with the given electron density, $E_{ee}[\rho]$ is the electron-electron interaction energy, and $E_{Ne}[\rho]$ is the energy of interaction between the electrons and 'external' fields. The E_{Ne} normally includes the interaction between the

electrons and the nuclei, but it can also include other fields imposed on the molecule. Each term depends on the electron density, ρ.

The exact form of $T[\rho]$ and $E_{ee}[\rho]$ are not known and no fundamental theory for determining them exists at present. The usual route for solving this problem is to use the method proposed by Kohn and Sham[115] in 1965 that involves *non-interacting* electrons. For this system, the electron density, $\rho(\mathbf{r})$, and the kinetic energy, $T_S[\rho]$ (S denotes that the kinetic energy is obtained from a Slater determinant), are exact, while the Coulomb (classical) part $J[\rho]$ of the electron repulsion energy can be calculated. The energy of the system now takes the form

$$E_{DFT}[\rho] = E_{Ne}[\rho] + T_S[\rho] + J[\rho] + E_{XC}[\rho] \qquad 2.28$$

where E_{XC}, the exchange-correlation functional, is:

$$E_{XC}[\rho]=(T[\rho] - T_S[\rho])+(E_{ee}[\rho]- J[\rho]) \qquad 2.29$$

The exchange and correlation energy E_{XC} is the functional that contains everything that is not known. Local-Density Approximation (LDA) and Generalized Gradient Approximation (GGA) are used to approximate the E_{XC}. LDA, the simpler approximation, assumes that the density corresponds to that of a homogeneous electron cloud. It proved to be an improvement over HF but is too inaccurate to be useful for chemistry[26]. GGA, an improvement on LDA, incorporates dependency on electron density and its gradient. In practise, *hybrid* (HF-DFT) functionals are used, which incorporate a portion of the 'exact exchange' from HF theory[105] with exchange and correlation from other sources (*ab initio* or empirical).

One of the most widely used *hybrid* functionals used is B3LYP, which incorporates Becke's three exchange-functionals[116] (B3) with Lee, Yang and Parr correlation functional LYP[117]. PBE0[118], another widely used functional, was used in this thesis (Chapters 3 to 6). These DFT methods are attractive because their results are comparable to those obtained using MP perturbation methods but come at a much smaller computational cost[105].

2.7.5. The DFT-D Approach

DFT methods have drawbacks. For one, they are non-convergent in that no one knows how to improve the results in a systematic way[99,119]. The success of a particular functional in one setting does not guarantee the performance in a different setting. These drawbacks led Orio[99] to stress the importance of enhancing the credibility of DFT either by including some form of validation or by estimating the error range through careful comparison between the calculated and measured observable.

Another drawback is that the density functional cannot describe long-range electron *correlations* responsible for dispersion forces[112,120]. To overcome this defect, researchers have introduced corrections. See refs[120,121] for a review of the impressive ongoing attempts to improve the description of dispersion forces of DFT methods. From a practical viewpoint, empirically damped $-C_6 \cdot R^{-6}$ corrections to the standard density functional approach provide robustness[122] and speed. We can solve the difficult *dispersion* interactions classically. This well-tested approach is DFT-D, where -D stands for *dispersion* correction. The total energy is

$$E_{DFT\text{-}D} = E_{DFT} + E_{disp} \qquad 2.30$$

where E_{DFT} is the self-consistent Kohn-Sham energy (equation 2.27) as obtained from the chosen density functional, and E_{disp} is an empirical *dispersion* correction. Although DFT-D was not used in this thesis, it is nevertheless relevant because this discussion highlights the importance of the *dispersion* contribution (i.e. correlation effect) to the overall energy of the system.

2.8. Crystal Structure Prediction (CSP)

CSP assumes that the experimentally determined polymorphs correspond to local minima in the crystal energy surface, E_{latt}. To put this approximation in context, it is important to note that the prediction of the relative thermodynamic stability of polymorphs at a given temperature T and pressure p requires the minimisation of the Gibbs free energy (G) with respect to the unit cell dimensions and the positions of all atoms in the unit cell:

$$\mathbf{min}\ G = \mathbf{min}[E + pV - TS] \qquad 2.31$$

where E is the internal energy, V is volume and S is the entropy[123]. The variation in the free energy arises from the differences in packing energy, crystal density, and entropy. Therefore, all three terms associated with the Gibbs free energy (equation 2.33) should be evaluated when determining the relative stability of polymorphs. It is often the case that only the lattice energy component of the first term (**min** E) is evaluated for two reasons:

- The $T\Delta S$ term is often neglected because the difference in lattice vibrational energies between polymorphs is usually very small[25,124]. At present, the thermal, entropic and zero-point contributions to the Gibbs free energy cannot be readily and accurate evaluated[56].

- The $p\Delta V$ term could be calculated but is often neglected because of its small ΔV value. The compressibility of solids being low, they contribute very little to Gibbs free energy at normal pressure. The $p\Delta V$ term dominates at high pressure, i.e. 10kbar[125].

Most of the computational approaches to CSP focus on minimising the lattice energy. The lattice energy, E_{latt}, is the internal energy at 0 K and 0 Pa, ignoring the zero-point energy. The thermodynamic stability is not always the decisive factor in the packing patterns of the molecules; kinetics also plays an important role because it strongly influences crystallisation conditions and solvent environment. Nonetheless, the experimental structure is expected to appear within a narrow energy range of the global lattice energy minimum.[56]

E_{latt} can be calculated in several different ways depending upon the relative importance of accuracy, ease of implementation, and computational cost in a given situation[56]. Ideally, the lattice energy of a crystal should be calculated quantum mechanically, but such calculations are expensive. Thus, most approaches use force field methods and/or a cheap method to generate and evaluating the lattice energy in a hierarchal manner. The most successful CSP approaches rely on some form of the computationally demanding but accurate electronic structure calculations either on an isolated molecule (to evaluate the intramolecular contribution combined with the intermolecular contribution that has been empirical fitted) or on a crystal (using periodic calculations).

Thus, the key to successful CSP methods is not just the generation of all possible crystal structures but also the accurate ranking of these generated crystal structures. This ranking requires balancing the intermolecular, U_{inter}, and intramolecular, ΔE_{intra}, energy contributions:

$$E_{latt} = U_{inter}(\mathbf{X}; \boldsymbol{\theta}) + \Delta E_{intra}(\boldsymbol{\theta}) \qquad 2.32$$

where U_{inter} is the sum of the dominant electrostatics, U_{elec}, and repulsion-dispersion, $U_{repul\text{-}disp}$, contributions and ΔE_{intra} is the energy required to distort the molecule from its gas phase conformation[71]. $\mathbf{X}$ denotes degrees of freedom that define the intermolecular contacts, while vector $\boldsymbol{\theta}$ denotes the intramolecular degrees of freedom, i.e. bond lengths, bond angles, and torsion angles. For rigid molecules, there is no $\boldsymbol{\theta}$; hence, the $\Delta E_{intra}(\boldsymbol{\theta})$ term vanishes. For flexible molecules, $\boldsymbol{\theta}$ and hence $\Delta E_{intra}(\boldsymbol{\theta})$ must be accounted for from the beginning.

2.8.1. CrystalPredictor

Chapters 3 and 4 use version 1.6 of *CrystalPredictor*[126], which is different from rigid or multiple rigid searches in that it incorporates conformation flexibility during the initial search stage of crystal structures. *CrystalPredictor* is not a black box; it requires careful analysis, clear thinking, and learning from earlier CSP studies. *CrystalPredictor* performs global lattice energy minimisation in order to predict the crystal structures of organic molecules whose conformation is likely to be significantly affected by the packing forces.

CrystalPredictor can find many plausible crystal structures in a multi-dimensional space defined by the conformational degrees of freedom and lattice variables, i.e. the positions of the centre(s) of mass of the molecule(s) in the unit cell, the orientation of the molecules and the unit-cell dimensions and angles[127]. Within *CrystalPredictor* are two important algorithms: the Sobol'[128] sequence and the E04UFF NAG routine[129]. The Sobol' sequence generates the initial structures; a characteristic of this approach is that each Sobol' corresponds to a distinct combination of every decision variable and thus ensures a uniform and efficient sampling. This prevents the problematic gaps and clusters that random sampling can exhibit[11]. For more detail on how Sobol' works, see ref. [128]. The E04UFF NAG routine[129] carries out an extensive search for local minima of the lattice energy surface using deterministic low-discrepancy sequences to ensure an optimal coverage of the search space.

CrystalPredictor models the lattice energy as the sum of the intramolecular and intermolecular energy (equation 2.32). This lattice energy evaluation is crude, but its computational efficiency is necessary because CSP studies often require hundreds of thousands of minimisations to sample the entire crystal energy surface. The algorithm computes the intramolecular energy as a function of flexible torsion angles using an interpolation method on a set of quantum mechanically pre-computed intramolecular energies by modelling the molecule as a set of rigid fragments connected by flexible torsion angles. The electrostatic contribution, U_{elec}, is evaluated by interpolating on pre-computed atomic charges, while an empirically fitted potential, FIT[34], computes the repulsive-dispersive interactions, $U_{repul\text{-}disp}$. The sum of these two contributions is the intermolecular energy, $U_{inter}=U_{elec}+U_{repul\text{-}disp}$.

In short, *CrystalPredictor* is a minimisation algorithm:

$$\mathbf{min}\ E_{latt} = \mathbf{min}[U_{inter}(\mathbf{X};\boldsymbol{\theta}) + \Delta E_{intra}(\boldsymbol{\theta})] \qquad 2.33$$

that can generate crystal structures in the 59 most common space groups and with more than 1 molecule in the asymmetric unit[126], Z′>1. However, Chapters 3 and 4 limit the search of hypothetical crystal structures to Z′=1 due to the huge computing time required to generate Z′>1 crystal structures.

2.8.2. CrystalOptimizer

CrystalOptimizer[71,130] refines crystal structures generated by *CrystalPredictor*[126]. To account for the molecular flexibility accurately, the *CrystalOptimizer*[71,130] algorithm is required. This algorithm solves the lattice energy problem (equation 2.33) as a two-level optimization problem:

$$\mathbf{min}\ E_{latt} = \mathbf{min}\ [\Delta E_{intra}(\boldsymbol{\theta}) + \mathbf{min}\ U_{inter}(\mathbf{X}; \boldsymbol{\theta})] \qquad 2.34$$

The outer problem varies and ultimately optimizes the flexible conformational degrees of freedom; the remaining degrees of freedom are adjusted to minimize the intramolecular energy[127]. A quasi-Newton scheme, which ensures rapid convergence with many optimization variable ($\mathbf{X}$; $\boldsymbol{\theta}$), solves the outer minimisation problem[71]. The inner minimization problem optimizes the lattice energy for the fixed molecular conformation determined by the outer problem. The DMACRYS[34] (section 2.6) program performs this rigid body minimization. *CrystalOptimizer* achieved a significant increase in the accuracy of lattice energy calculations by[127]

1. allowing additional flexibility during lattice energy minimization (for example considering more torsion angles and/or bond angles as flexible degrees of freedom),
2. accurately computing the intramolecular energy throughout optimization, and
3. modelling the electrostatic interactions using distributed multipoles (up to hexadecapole) obtained using the GDMA package[52].

Multiple *ab initio* calculations, a very time-consuming process, were often required. For this reason, only the most promising structures from *CrystalPredictor* were used as initial points for optimization. Sometimes, an intermediate step such as single-point *ab initio* calculations (see Chapters 3 and 4) can improve the ranking of the lattice energy so that only a handful of crystal structures (energy cut-off or a set number of unique crystal structures minima) are optimized with *CrystalOptimizer*.

CrystalOptimizer has one of the major features—Local Approximate Models (LAMs)––that models the conformational energy and the multipole representation of the charge density; these are updated with explicit quantum mechanical calculations only when necessary and not at every minimisation step. The conformational energies, their first and

second gradients, and the distributed multipoles are also stored in the LAMs database. Re-using expensive electronic structure calculations (if the same conformational ranges are encountered while minimising different hypothetical or experimentally determined crystal structures) reduces the computing time without sacrificing accuracy[71,72,130].

2.8.3. Polarizable Continuum Model (PCM)

Accurately modelling the balance between intramolecular and intermolecular energy is important (e.g. using *CrystalOptimizer*), but just as important is modelling the effect of the environment in distorting the conformer, especially in cases where there is a considerable difference in molecular conformations. Cooper et al.[131] noticed that the total calculated crystal energies, which combine atom-atom model potential calculations of intermolecular interactions with DFT intramolecular energies, do not effectively distinguish the real (known) crystal structures from the rest of the low-energy computer-generated alternatives when the molecular electrostatic models are derived from isolated molecule calculations. They found that introducing a simple model for the bulk crystalline environment when calculating the molecular energy and electron density distribution led to important changes in relative total crystal energies and made possible the correct distinction between the observed crystal structures and the set of computer-generated possibilities. This finding highlights the importance in CSP calculations of the polarization of the molecular charge distribution, especially for polar flexible molecules[131].

In Chapters 3 and 4, PCM[132] is used to account for the effect of organic environment after generating and accurately computing the lattice energies using *CrystalPredictor*, that is, an intermediate step that improves the electrostatic contribution to the intermolecular energy and *CrystalOptimizer*. I will model the polarizable continuum environment with dielectric constants that is typical of molecular organic crystals i.e. $\varepsilon = 3$[133]. The PCM calculations only affect the final energy. The geometry is not affected because there is no molecular geometry minimisation of the crystal structure during the PCM calculation.

2.8.4. Other Approaches to Lattice Energy Minimisation

There are other well-known approaches to lattice energy minimisations for flexible molecules. One computationally efficient approach is the Rigid *CrystalPredictor*-Molecular Mechanics Method (RCM)[17] by Graeme Day, a variant on the method used in Chapters 3 and 4. The first step performs constrained geometry optimization on a set number of distinct conformations obtained from conformational analysis (the flexible

torsion angles are fixed at the expected values from conformational analysis, and all other degrees of freedom are optimized). From this first step, we can obtain the geometry, intramolecular energy, and atomic charges. The second step generates crystal structures in 21 of the most common space groups by using a rigid *CrystalPredictor*[134] search for each distinct conformation[17]. Essentially, the molecular geometries are treated as rigid during the process of generating crystal structures. The U_{inter} is calculated from the atomic charges and an empirical *exp*-6 repulsion–dispersion potential, W99[135]. The third step refines the most stable crystal structures obtained from the rigid-molecule searches. For this, a two-stage procedure proposed by Day[136] and co-workers is used. The first stage minimises the crystal structures by allowing the molecular geometry to relax, using a molecular mechanics description of energies associated with changes to the flexible torsion angles. This method trusts the molecular mechanics force field to provide the correct molecular geometries, but discards the energy, which is not of sufficient accuracy for the final ranking of crystal structures. Then, the resulting crystal structures are re-optimized without further changes to the molecular conformations. DMACRYS[34] is used to model the electrostatic interactions, U_{elec}, using the atomic multipole model. The remaining intermolecular terms, U_{repul} and U_{disp}, are obtained using the empirical W99 potential[135]. The intramolecular energy, ΔE_{intra}, and the atomic multipole moments for each conformation are derived from a single-point DFT calculation that incorporates the PCM with a constant dielectric value of three to reflect the polarization effects in the crystalline environment.

Another approach is GRACE, one of the most accurate minimisation methods. This approach, presented by Neumann and Perrin was the first case wherein DFT-D methods (sections 2.7.5) were applied to the CSP of molecular organic crystals. First, during the crystal structure generation procedure, a Tailor-Made Force Field (TMFF)[13] developed for fitting an atom-atom force field (non-bonded and bonding terms) specifically for each molecule is used. Then, GRACE uses a hybrid method that combines the sum of the DFT component and an empirical van der Waals correction, –D (sections 2.7.4 and 2.7.5) to re-optimise the lowest-energy crystal structures. During this re-optimisation stage, all nuclear positions and electron density are optimised simultaneously, which naturally accounts for the molecular flexibility and other charge density reorganisation effects such as polarization. This approach reproduces experimental structures very accurately and ranks the structures favourably amongst other hypothetically generated structures in many (but not all) cases. However, these periodic DFT-D calculations are very expensive[137].

2.8.5. The Crystal Energy Landscapes

The CSP methods[6,11] discussed above, and many others, sample a wide range of possible crystal structures and determine the lattice energy in a hierarchal manner, thus generating a crystal energy landscape[14]. A crystal energy landscape is a plot of the lattice energy against the density or the packing coefficient, and it tells us how well the crystal packs. These landscapes often include the experimental structures, as well as other low-lying energy crystal structures that are worth comparing with the known polymorphs in order to determine the types of intermolecular interactions and conformations that could be feasible for a given molecule.

In a crystal energy landscape in which only one crystal structure is calculated to be more stable than any other by over a few kJ mol^{-1} typical of polymorphic energy differences, the molecule should only be found in that global minimum structure. Experiments or more demanding searches are needed to determine the stability relative to multicomponent forms such as hydrates[138], solvates[139] or cocrystals[140,141], etc. However, in most generated crystal energy landscapes, there is typically a cluster of structures within the energy range of plausible polymorphs. These structures have to be qualitatively assessed to see whether the barriers between the different structures are so low that thermal motion in the crystal or molecular rearrangement during nucleation and growth will ensure that only the most thermodynamically stable structure is seen. Closely related structures may also give rise to disorder[142,143] or plastic phases[144] at higher temperatures.

The crystal energy landscape may suggest that alternatives to the known structures could be found either as more stable or as kinetically trapped, metastable polymorphs. In this case, the calculations can complement experimental screening[145] by suggesting alternative methods[146], such as a templating, to produce new polymorphs[147]. The knowledge of the crystal energy landscape can help avert problems in pharmaceutical industries during the manufacturing process and guide the experimentalist search for undiscovered structure. In fact, the study of the crystal energy landscape in Chapter 4 helped my colleague to discover a new polymorph (section 4.3.8). A way of testing methodologies developed for predicting crystal structures is through ‘blind tests’.

2.8.6. The Blind Tests

The early lack of progress in CSP was termed a ‘continuing scandal’ in Nature in 1988 by John Maddox[148]. In response, the Cambridge Crystallographic Data Centre (CCDC) organised five ‘blind tests’[10,11,66,67,149] of organic CSP, with the first ‘blind test’ held in

1999. The blind test provided many researchers the opportunity to test their CSP methods. The fourth blind test results showed a dramatic improvement in the rates of success over previous blind tests, with one group (GRACE) correctly predicting four crystal structures as their first ranked choice, albeit at a considerable computational expense[10].

In the fifth blind test, our research group successfully predicted the crystal structure of molecule XX[11,17], which has eight flexible degrees of freedom (internal degrees of freedom), using *CrystalPredictor* and *CrystalOptimizer* algorithms. I use a similar approach in Chapters 3 and 4. The fifth blind test showed that methodologies such as DFT-D could reliably predict the structures of small organic molecules, but GRACE was unsuccessful for molecule XX[11]. CCDC released the sixth 'blind test' molecules in mid-September 2014[15], too late for me to take part. By chance, an informal 'blind test' challenge did arise during the second year of my research, which is the focus of Chapter 3.

2.9. Crystal Structure Comparison

Collins et al.[150] argued that one of the problems modern scientists face is the issue of structural similarity and difference. They noted "criteria for deciding when two crystal structures are the same is an issue that divides many structural analysts". Indeed, this problem of determining how similar or different two or more crystal structures are was what inspired researchers to create methods for comparing crystal structures that exist today. The visual inspection of similarities and differences between crystal structures does have its benefits, but the method is slow and susceptible to human bias. Thus, researchers need a practical computational method that enables them to identify identical or similar structures.

Identifying the similarities among crystal structures is important when deciding whether a newly discovered structure is the same as an existing structure[151]. Sometimes there is also a need to determine similarities when comparing many structures that are related, as in the case of 50 structures of carbamazepine (CBZ)[150]. Another reason for identifying similarities lies in the field of CSP[67,152], where thousands of hypothetical structures are generated and many structures converge to the same or approximately the same structure during lattice optimization. Therefore, the removal of duplicates and the production of unique lists of crystal structures are essential.

The *Mercury*[153] material module, which has *COMPACK*[151] embedded in it, allows graphical representation of the similarity among the polymorphs, while standalone *COMPACK* scripts can compare hundreds of thousands of crystal structures. There are

three main tools in the material module[153]: motif search, crystal packing feature and crystal packing similarity tools. The crystal packing similarity tool calculates the level of packing similarities between crystal structures that have the same or similar compound.

COMPACK computes these calculations by analysing the geometry of the cluster of molecules. There are two basic steps in comparing two crystal structures using *COMPACK*. The first step renders the chemical search query subject to several geometric constraints using the reference molecular structure. The second step is the structure similarity search, which involves searching the 3-D coordinates of a target for an arrangement that matches the query to within specified tolerances. The search is performed by 3DSEARCH[154], which is an efficient general search algorithm designed to find chemical queries within 3D crystal structures information. Once 3DSEARCH finds a match, the two structures can be superimposed to obtain a visual impression of their similarity.

The default values for a cluster size of 15 molecules, a distance tolerance of 20%, and angle tolerance of 20° are usually used. The default cluster size is 15 because the same clustering results of 30 molecules, instead of the default 15 molecules for 1,000 predicted structures of 2-amino-3-nitropyridine led Chisholm and Motherwell[151] to the conclusion that a cluster of 15 molecules is a sufficient size for identifying the same crystal structure.

Other methods of comparing crystal structures do exist. For example, XPac[155,156] identifies similar sub-components of complete crystal structures of 0-, 1-, or 2-D in polymorphs or crystal structures of similar molecules, as well as of 3-D in isostructural assemblies. Hirshfeld surface (a way of exploring packing modes[157]) and fingerprint plots (a visual representation of all intermolecular interactions simultaneously[158]) are examples of Hirshfeld-based methods[150,159,160] of analysing similarities. The graph–set analysis for H-bonds compares and contrasts sets of molecules[161-163] and is based on graph theory for categorising H-bond motifs. Moments of inertia tensors is a technique for comparing crystal packing based on moments of inertia tensors irrespective of chemical connectivity and hence, the comparison of compounds of diverse chemical composition is possible[164]. Simulated powder X-ray diffraction[165] is another example of comparing crystal structures. In short, the comparison of compounds of diverse chemical composition is possible.

COMPACK's ability to give just one number (the distance between matched atoms in a cluster-root-mean-square-deviation) that shows the relationship between two structures makes this similarity tool one of the simplest and most effective ways of quantitatively

rationalising similarity. This is why *COMPACK* will be the primary choice for comparing crystal structures in this thesis.

Chapter 3. An Informal Blind Test: A Crystal Structure Prediction of 1-Benzyl-1*H*-tetrazole

"*One of the continuing scandals in the physical sciences is that it remains in general impossible to predict the structure of even the simplest crystalline solids from a knowledge of their chemical composition.***"**

— John Maddox, *Nature* (1988)

3.1. Introduction

3.1.1. Background

This chapter reports an informal 'blind test' of flexible 1-benzyl-*1H*-tetrazole, BT (Figure 3.1), using similar CSP methodology that successfully predicted the very flexible molecule XX in the fifth blind test[11,17]. In essence, I will test the viability of the CSP method, (sections 1.1 and 2.8), which has evolved over the years from predicting the crystal structures of rigid molecules[166-168] to molecules that are so flexible that their conformational flexibilities are accounted for at the start of CSP[11,17].

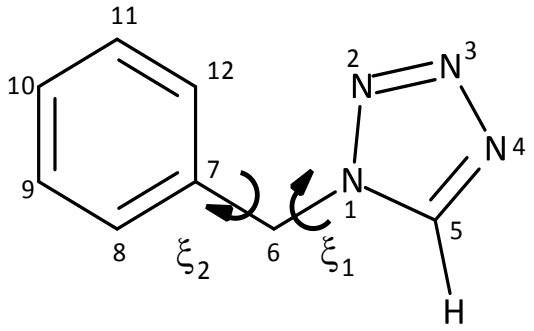


Figure 3.1: The chemical diagram of BT determined from the crystal structure. Torsion angles ξ_1=C_7-C_6-N_1-C_5 and ξ_2 = C_{12}-C_7-C_6-N_1.

In the previous chapters, I mentioned that the emergence of a new polymorph of ritonavir two years after market lunch (section 1.1) as well as John Maddox's[148] "continuing scandal" remark in 1988 (section 2.8.6) were the catalysts of the 'blind test' challenge organised by the Cambridge Crystallographic Data Centre (CCDC) in 1999. These 'blind tests' provided researchers all over the world an opportunity to test their CSP methods. But unfortunately, the CCDC—my industrial sponsor—did not organise any 'blind tests' during the course of the research reported in this thesis[15]. By chance, my colleague *Rex Palmer* observed an unusual packing of BT crystal structures, and subsequently issued me a 'blind test' challenge. This challenge was to predict the crystal structure of BT given just the chemical diagram (Figure 3.1) and given that Z′=1.

The drug discovery literature frequently mentions the motifs of 5-substituted *1H*-tetrazole as bioisosteres for the carboxylic acid group[169-171], due to the motifs' similar pK_a values, shapes, and steric interactions. Bioisosteres are structural moieties with broadly similar shapes and functions. For pharmaceutical industries, the ability to model the interactions of tetrazoles, including BT, is of great interest due to tetrazole's bioisosteres feature.

3.1.2. Experimental Determination of 1-Benzyl-1H-Tetrazole

My collaborators (*John Spencer*, *Hiren Patel*, *John J. Deadman*, *Rex Palmer, Louise Male*, and *Simon J. Coles*) carried out the following experiments (their findings in sections 3.1.3 and 3.1.4 were not made available to me until the end of the 'blind test,' i.e. the submission of my candidate structures):

Synthesis and X-Ray Crystallography

John J. Deadman synthesized BT on a ca. 50 mmol scale from benzylamine, sodium azide, and triethyl formate. *Hiren Patel,* under the supervision of *John Spencer,* grew the single crystals of BT from dichloromethane–hexane.

Louise Male and *Simon J. Coles* collected the X-ray data and did the preliminary analysis. A colourless crystal fragment of dimensions 0.09 × 0.03 × 0.02 mm^3 was mounted on a glass fibre and flash-frozen to 120 K. The X-ray data were collected on a Bruker diffractometer using Mo Kα radiation. The crystal structure was solved by direct methods. SHELXL-97[172-174] was used to refine the crystal structure.

3.1.3. Molecular Geometry of the Experimental Structure

My collaborators discovered that BT crystallised in the monoclinic space group $P2_1$ with unit cell dimensions a = 7.6843(5), b = 5.5794(4), c = 9.4459(7) Å, β = 100.949(4) °, V = 397.61(5) $Å^3$, Z' = 1, and density = 1.338 g cm^{-3}. Overall, the benzyl and tetrazole rings were essentially planar, including the H atom on the CH group (see C_5 in Figure 3.1) belonging to the 5-membered tetrazole ring. The individual rings were also co-planar with the inter-ring link C_6 atom (Figure 3.1). The N-N bonds (see bond lengths in Table 3.1) were of three different types, consistent with Figure 3.1: N_1-N_2 = 1.343(2) Å; N_2-N_3 = 1.301(2) Å is longer than a typical N = N bond[175] but shorter than a typical aromatic bond; N_3-N_4 = 1.368(2) Å was also largely aromatic in character[175]. The two C-N bonds have different characteristics: C_5-N_4 = 1.314(2) Å was a double bond; C_5-N_1 = 1.336(2) Å was typically aromatic[175]. The phenyl rings of BT have a normal geometry, with average C–C bond length of 1.386(4) Å, average bond angle of 120.0(7)° and ring planarity root mean square (*rms*) deviation of 0.0057 Å. The bond angles in the tetrazole

ring have an average of 108.0(1.7)°, which is consistent with the planarity of the ring having an *rms* deviation of 0.012 Å, including the link C atom and H_5 on C_5. These bond lengths and angle measurements are similar to those of 5-azido-*1H*-tetrazole[176].

Table 3.1: Angles and bond lengths of the tetrazole ring in BT. No information about the bonds or angles were given to me prior to submission of the candidate structures.

Angles	X-ray °	Bond lengths	X-ray Å
N_1-N_2-N_3	106.9(2)	N_1-N_2	1.343(2)
N_2-N_3-N_4	110.2(2)	N_2-N_3	1.301(2)
N_3-N_4-C_5	105.2(2)	N_3-N_4	1.368(2)
N_4-C_5-N_1	109.6(2)	N_4-C_5	1.314(2)
C_5-N_1-N_2	108.0 (2)	C_5-N_1	1.336(2)
C_5-N_1-C_6	130.5(2)	N_2-C_1	1.472(2)
N_2-C_1-C_6	121.5(2)		

3.1.4. Crystal Packing: Weak Hydrogen Bonding

Rex Palmer noticed that the mode in which the molecules in the experimental crystal structure BT were packed was unusual. He noted that a molecule such as BT, based on linked delocalized rings, might be expected to form crystal structures involving π···π bonding (ring stacking). However, no such interactions occurred in the experimental structure (Figure 3.2). Instead, the structure was held together by a large number of weak intermolecular contacts[177,178]: 12 of these were of the CH···N interaction, and 3 were of the weaker CH···C interaction, which involve atoms in the phenyl ring. This unusual packing resulted in a structure composed of infinite S-shaped layers, and produced a sequence of benzyl and tetrazole layers (Figure 3.3). Another unexpected feature of the packing was the formation of tetrazole clusters (Figure 3.2a), each having a central tetrazole ring coordinated by six other tetrazoles.

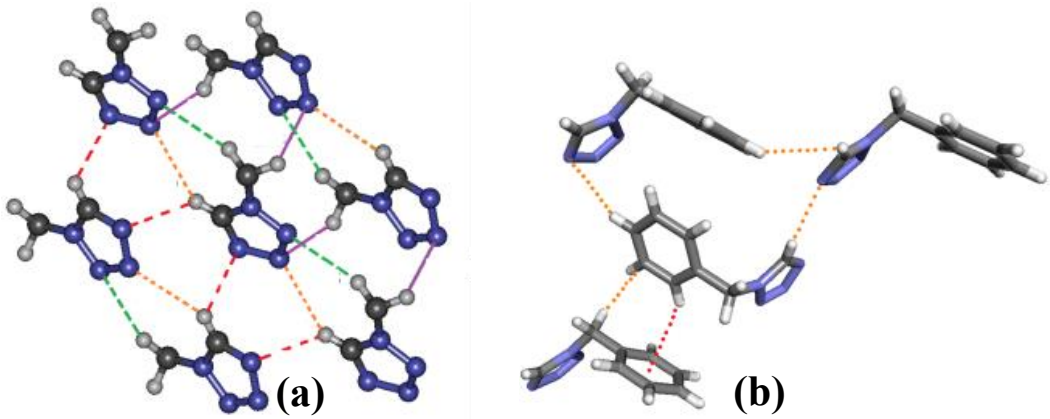


Figure 3.2: The crystal structure of BT: (a) Partial view of the crystal packing showing close CH···N contacts between a central tetrazole and six surrounding tetrazoles, (b) Partial view showing CH···N, CH···C and CH···π contacts.

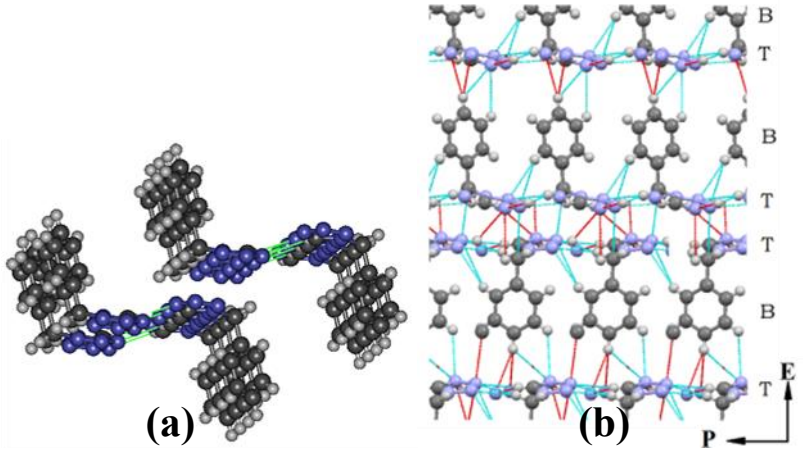


Figure 3.3: BT crystal structure: (a) View of the infinite S-shaped layers held together by weak CH···N and CH···C H-bonds; (b) Partial view of the layer structure showing alternative B(benzyl) and T(tetrazole) layers, and the repeating layer sequence [BTBT | TBTB].

3.1.5. Overview of the Challenge

After my experimental collaborators grew single crystals of BT from a non-polar solvent mixture and determined the crystal structure at 120 K, I was issued a 'blind test' challenge by *Rex Palmer* to predict the crystal structure of BT because of the unusual mode of crystal packing (section 3.1.4). No experimental information was given prior to the submission of the candidate structures. The subsequent part of this chapter presents the computational work I carried out independently upon receiving the chemical diagram of BT (Figure 3.1).

This chapter uses CSP methodology (section 3.2.2) to predict the crystal structure of a molecule that can adopt a wide range of molecular shapes due to the conformational flexibility of the molecule. Consequently, the conformational flexibility must be accounted for at the start of CSP (*CrystalPredictor* 1.6, section 2.8.1). In addition, this method relies on the systematic improvement of the evaluation of the lattice energy, E_{latt}:

$$E_{latt} = U_{inter}(\mathbf{X}; \boldsymbol{\theta}) + \Delta E_{intra}(\boldsymbol{\theta}) \qquad 3.1$$

where U_{inter} and ΔE_{intra} are the intermolecular and intramolecular contributions. $\mathbf{X}$ and vector $\boldsymbol{\theta}$ denote intermolecular and intramolecular degrees. These systematic improvements increase the computing time. As a result, it is essential to reduce the number of crystal structures that will be examined as the confidence in the lattice energy ranking improves in each step.

This chapter will also test the transferability of the intermolecular energy potential to BT since this molecule contains a functional group that differs from those used in determining the parameters of the intermolecular potentials (section 2.3)[38,39].

To determine whether this study is a success, the predicted structures of BT will be compared with the experimental crystal structures using a similarity tool. In addition, I will compare and contrast the most stable generated structures of BT with the similar tetrazole fragment structures to determine whether the packing of the experimental structure of BT was unusual. In this chapter, I will show that the crystal structure of BT can be successfully predicted starting from just the chemical diagram (Figure 3.1).

3.2. Computational Methodology

3.2.1. Conformational Analysis

To restrict the search space of BT to energetically meaningful regions, I carried out quantum mechanical (MP2/6-31G(d,p) single-point calculations on HF/6-31G(d,p) optimised geometries) scans of the two flexible torsions, ξ_1 and ξ_2 (Figure 3.1), using the GAMESS[179] program.

3.2.2. Crystal Structure Prediction

Step 1: The search was carried out using the *CrystalPredictor*[126] program, which covered the 59 most commonly occurring space groups. The lattice energy was minimised by varying the cell parameters and torsion angles ξ_1 and ξ_2 within energetically meaningful regions (section 3.2.1). At this stage, the lattice energies (equation 3.1) were crudely estimated, and then, using the grid of the *ab initio* results from section 3.2.1, the ΔE_{intra} contribution were interpolated. The intermolecular electrostatic interactions were modelled using the atomic charges derived from section 3.2.1. All other intermolecular energy terms were derived from an empirical *exp*-6 potential, FIT.[40] Approximately a quarter of a million lattice energy minimisations were performed. This produced approximately 170,000 distinct structures, of which 44,000 structures were unique. These generated structures were labelled by their energy ordering. The result of step 1 is the *CrystalPredictor* energy landscape.

Step 2: Next, the lattice energy model was improved. For each of the lowest 10,000 crystal structures, the energy of the isolated molecule and its charge distribution were calculated at the PBE0/6-31G(d,p) level of theory, using GAUSSIAN03[53]. This calculation provided a better estimate of ΔE_{intra}, and a more accurate representation of the charge density in terms of a distributed multipole model[51]. These two contributions were then combined with an atom–atom *exp-6* repulsion-dispersion potential using parameters that had been fitted to azahydrocarbons[39] and polar crystal structures[40], which were assumed to be transferable to this tetrazole. At this stage, the lattice energy was minimised, with the molecule held rigid using DMACRYS[34].

Step 3: Approximately 100 of the most stable structures were then further refined at the PBE0/6-31+G(d) level of theory by allowing the molecular conformation to adjust more accurately to the intermolecular forces using the program *CrystalOptimizer*[71,72] (section 2.8.2). This algorithm allowed the intermolecular and intramolecular energy to be minimised simultaneously. The *CrystalOptimizer* program uses the Quasi-Newton algorithm coupled with a line search to solve the minimisation program. It also makes use of GAUSSIAN03[53] calculations on the molecular conformations and DMACRYS[34] optimisations of the crystal structure.

Step 4: The effect of the crystal environment on the conformational energies ΔE_{intra} and charge densities was estimated by calculating the conformational energy and charge distribution in a Polarizable Continuum Model (PCM)[180] with ε=3, a value typical of organic molecules[131]. This calculation was implemented in GAUSSIAN03[53] at the PBE0/6-31+G(d) level of theory. The result of step 4 is the PCM crystal energy landscape.

3.3. Results and Discussion

3.3.1. Conformational Energy Grid

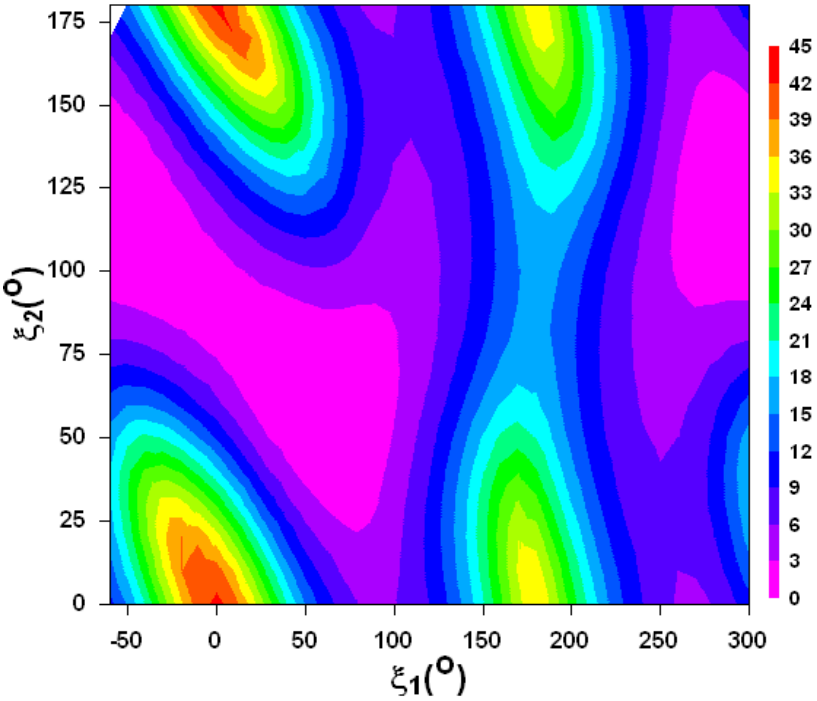


Figure 3.4: The conformational energy grid used in the search. Relative intramolecular energies, ΔE_{intra} (kJ mol^{-1}), for isolated molecule calculations at the MP2/6-31G(d, p) level of theory.

Prior to the generation of hypothetical structures, it was necessary to investigate the possible steric hindrances to rotation about the two flexible torsion angles ξ_1 and ξ_2 (Figure 3.1). The result, shown in Figure 3.4, indicates a high degree of flexibility in the molecule. The purple colours represent regions with the smallest energy penalty, i.e. the

most stable conformations. This high degree of flexibility makes possible a wide range of crystal structures with different molecular shapes. For this reason, the generation stage (step 1 of sections 3.2.2 and 3.3.2) incorporates conformational flexibility from the start.

3.3.2. The CrystalPredictor Energy Landscape

The analyses presented here are post-process, i.e. after submitting the ten most likely crystal structures on the PCM (final) energy landscape (**step 4**, section 3.3.3). The energy landscape after **step 1** (*CrystalPredictor*) is displayed in Figure 3.5. This step estimated the lattice energies, E_{latt}, very crudely by interpolating the atomic charges and intramolecular energies (Figure 3.4). Yet, the generated structure that matched the experimental structure of BT at this stage was only ranked 23rd. This structure, #BT_23, which is highlighted in red on Figure 3.5, has only a 3 kJ mol^{-1} energy difference in comparison with the global minimum structure.

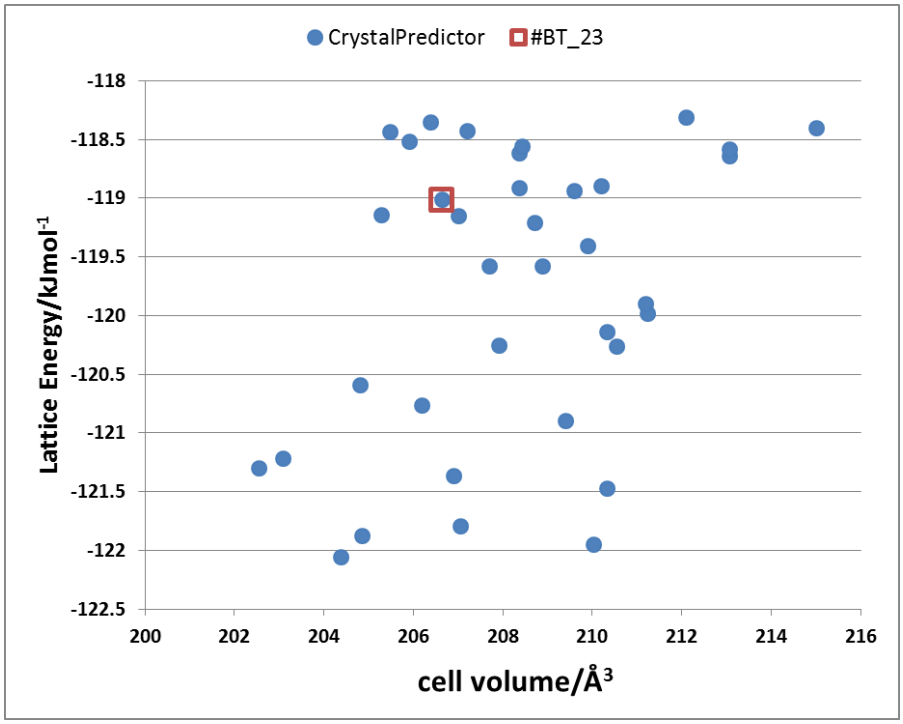


Figure 3.5: The *CrystalPredictor* energy landscape of BT generated with a crude lattice energy model. The generated structure that matched the experimental structure is highlighted in the open red symbol.

3.3.3. The CrystalOptimizer and PCM Energy Landscapes

The two most stable crystal structures on the PCM energy landscape (Figure 3.6b) have similar lattice energy: less than 0.1 kJ mol^{-1}. The 'blind test' was deemed a success because the experimental structure matched the second most stable crystal structure, #2BT_23, on the PCM energy landscape (Figure 3.6b and Table 3.2). This PCM energy landscape was generated after *CrystalPredictor* (**step 1**, Figure 3.5), lattice energy refinement (**step 2**), *CrystalOptimizer* (**step 3**, Figure 3.6a), and PCM calculations (**step**

4, Figure 3.6b). Compared to the first step (**step 1**, Figure 3.5), the lattice energy rank of the experimental structure improved significantly, from 23rd to 2nd.

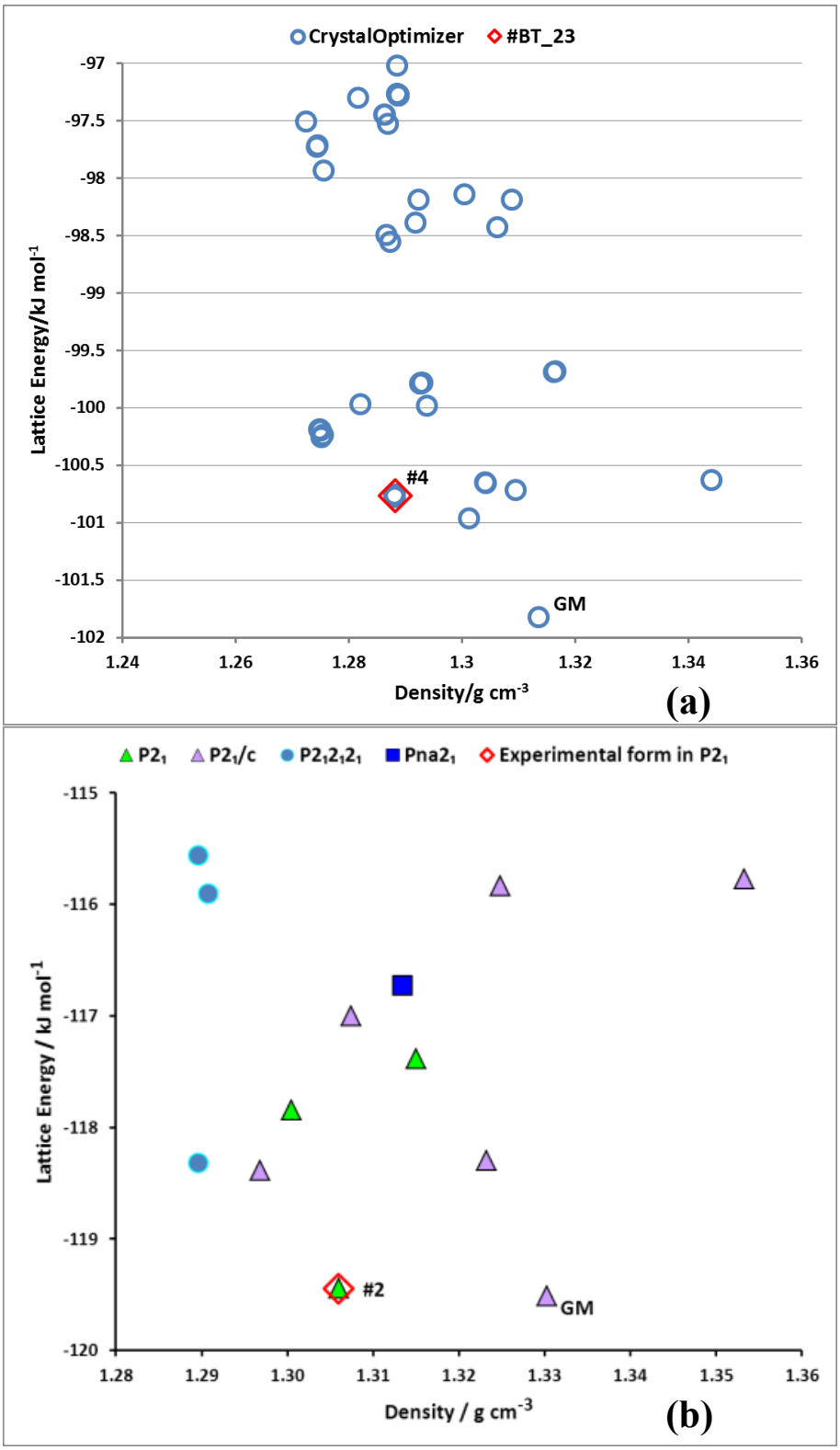


Figure 3.6: (a) The *CrystalOptimizer* (**step 3**), and (b) PCM (**step 4**) energy landscape of BT. GM is the global energy minimum. The open red symbol represents a match between the experimental crystal structure and the generated structure. The energy ranking of the crystal structure that matches the experimental has improved from fourth most stable, #4, to second most stable, #2.

I submitted a list of 10 structures in which the #2BT_23 structure was selected as the first choice, while the global minimum (labelled GM on Figure 3.6b) was selected as the second choice[181]. The main reason for selecting #2BT_23 as the first choice was the notable improvement of energy rank, from fourth (Figure 3.6a, **step 3)** to second (Figure 3.6b, **step 4)**. In addition, the two most stable crystal structures on the PCM energy landscape (Figure 3.6b, **step 4)** were only 0.1 kJ mol^{-1} apart.

Table 3.2: Summary of experimental (highlighted in blue) and computed crystal structures on the PCM energy landscape (Figure 3.6). The computed structure that matched the experimental structure (section 3.1.2) is highlighted in bold. For label #*x*BT_*y*, *x* and *y* are energy ranked after steps 4 and 1 of CSP respectively (section 3.2.2).

Label	Space Group	*a*	*b*	*c*	*β*	*ρ*	U_{inter}	ΔE_{intra}	E_{latt}
		Å			°	g cm^{-3}	kJ mol^{-1}		
#1BT_120	$P2_1/c$	9.99	8.17	11.04	117.50	1.33	-121.76	2.25	-119.52
#2BT_23	**$P2_1$**	**7.93**	**5.55**	**9.44**	**101.32**	**1.31**	**-121.52**	**2.08**	**-119.44**
Experiment	$P2_1$	*7.68*	*5.58*	*9.45*	*100.95*	*1.34*	-	-	-
#3BT_165	$P2_1/c$	10.86	5.54	15.97	121.41	1.30	-121.09	2.70	-118.39
#4BT_335	$P2_12_12_1$	9.38	15.81	5.56	90.00	1.29	-120.32	2.00	-118.33
#5BT_767	$P2_1/c$	9.41	5.53	15.80	102.34	1.32	-120.44	2.14	-118.30
#6BT_167	$P2_1$	5.68	8.39	8.81	103.22	1.30	-122.89	5.04	-117.85
#7BT_912	$P2_1$	5.57	8.69	8.61	103.65	1.32	-118.82	1.43	-117.39
#8BT_34	$P2_1/c$	8.09	5.53	18.59	101.66	1.31	-119.19	2.19	-117.00
#9BT_491	$Pna2_1$	16.66	5.66	8.58	90.00	1.31	-122.99	6.26	-116.73
#10BT_250	$P2_12_12_1$	19.04	5.57	7.78	90.00	1.29	-119.16	3.25	-115.91
#11BT_55	$P2_1/c$	10.80	5.56	15.82	122.32	1.32	-119.79	3.96	-115.83
#12BT_490	$P2_1/c$	6.53	8.66	13.94	94.28	1.35	-120.98	5.21	-115.77
#13BT_178	$P2_12_12_1$	8.33	5.61	17.64	90.00	1.29	-117.89	2.32	-115.57

3.3.4. Did the Predicted Structure Really Match the Experimental?

Rex Palmer provided the *.cif* file of the experimental structure once ten of the most stable structures on the PCM landscape were submitted (Figure 3.6b). Then, I compared the experimental structure with the submitted crystal structures using COMPACK[151]. The computed #2BT_23 structure, my first choice (section 3.3.3), gave an excellent overlay with 15 molecules of the experimental structure with a root mean square difference, RMSD, value of only 0.148 Å (Figure 3.7). This computed #2BT_23 structure has similar unit cell dimensions (Table 3.2) and is in the same space group, $P2_1$, as the experimental structure.

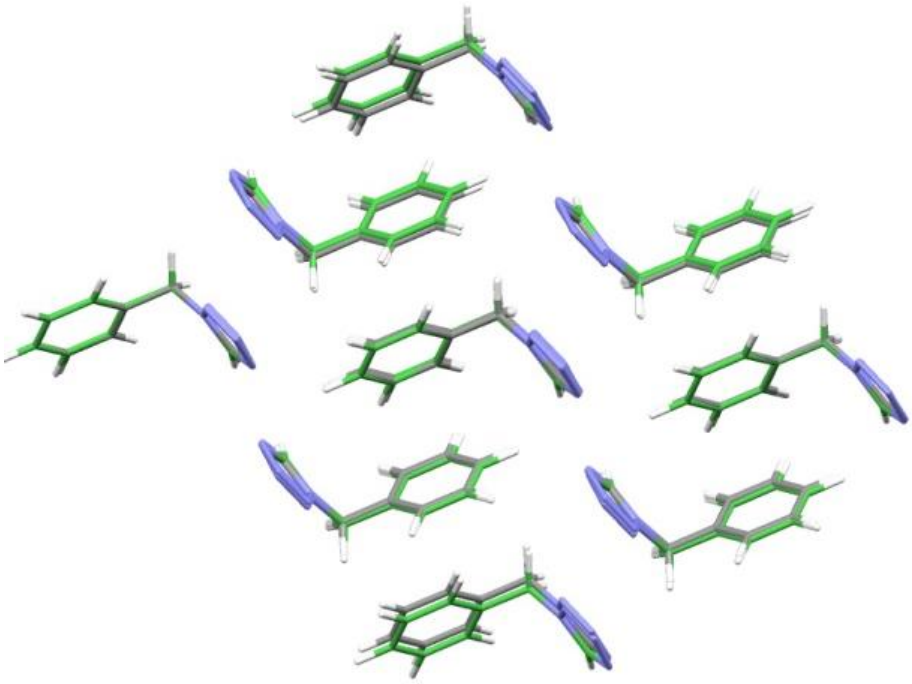


Figure 3.7: The 15-molecule overlay between the experimental (atomic colours) and the second most stable computed structure, #2BT_23 (green). The direction of the view obscures some of the molecules.

In addition to excellent $RMSD_{15}$ overlay, the #2BT_23-generated structure provided a starting model that was easily refined in SHELXL-97, which was further proof that the structure was a genuine reproduction of the experimental crystal structure (section 3.1.2). Hence, it is clear that the CSP method outlined in this chapter, which uses only the chemical diagram as prior information, can successfully predict the crystal structure of a flexible molecule.

3.3.5. The Global Minimum Structure on the PCM Energy Landscape

The global minimum, GM, structure labelled #1BT_120 on the PCM energy landscape (Figure 3.6) was very different from the experimental crystal structure as only two molecules overlaid (Figure 3.8) during similarity calculations. From the overlay of the similarity calculations (Figure 3.8), it is clear that the GM and the experimental structure have different conformations. The GM structure, #1BT_120 (Table 3.2), is in space group $P2_1/c$, and has a very different spatial arrangement (as shown in Figure 3.9) when compared to the experimental structure (Figure 3.2 and Figure 3.3). This structure also has dominant weak H-bonds: 8 CH···N and 1 CH···C interactions. There was also one weak π···π, benzyl···benzyl interaction (Figure 3.9).

The similar lattice energy values of the two most stable structures generated (Figure 3.6b) suggest that the GM structure could a polymorph. This GM structure could be

synthesised at high pressure because it is the denser of the two most stable structures. Also, growing the experimental structure using a non-polar solvent (section 3.1.2) may have prevented H-bonding formation.

After the generation step (*CrystalPredictor*), the initial rank of the most stable crystal structure after **step 4**, GM, was 120th. The eventual rank was first, despite the fact that only about 100 structures were optimised; this shows the importance of the novel intermediate **step 2** in the CSP method. After the *CrystalPredictor* **step 1**, the GM structure was 9 kJ mol^{-1} less stable than the experimental #2BT_23 structure and had distinctly fewer close contacts than the experimental #2BT_23 structure. Therefore, the rearrangement of lattice energy depends on the method used.

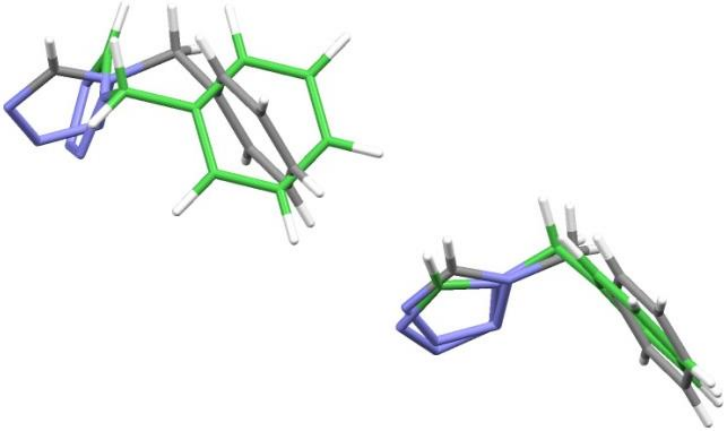


Figure 3.8: Crystal Packing Similarity between experimental (grey) and the most stable predicted structure (#1BT_120, green), with only 2 molecules matching ($RMSD_2$ = 1.072 Å).

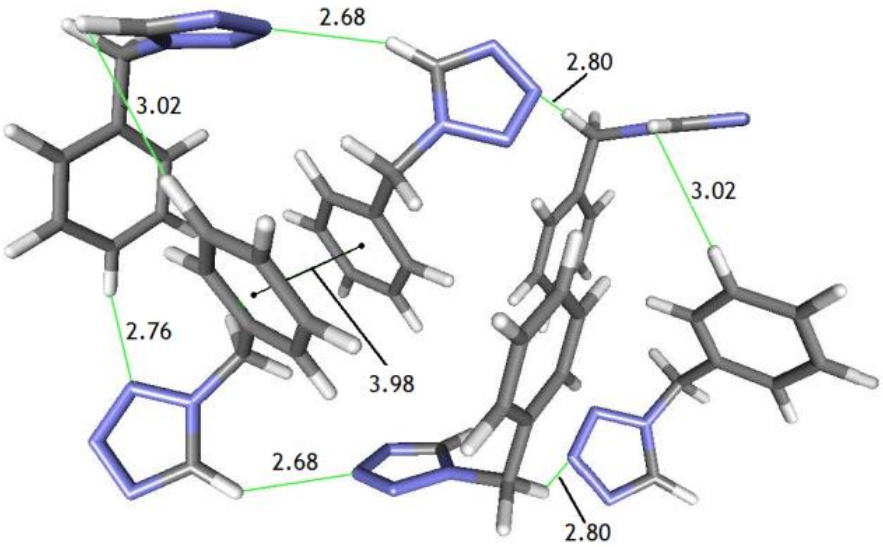


Figure 3.9: The structure corresponding to the global minimum in lattice energy (#1BT_120) showing the dominant weak H-bonds (Å): eight of type CH···N; and one CH···C and the weak π···π interaction (benzyl···benzyl).

3.3.6. Post Analysis Comparison of Structure with CSD

The motivation for the 'blind test' challenge was the unusual packing of the experimental structure. Thus, a database containing over half-million crystal structures was searched in an effort to investigate the breadth of variation in the packing of the experimental structures of BT. A Cambridge Structural Database (CSD)[182] survey was carried out to compare BT with other 1-substituted *1H*-tetrazole crystal structures (Figure 3.10). The CSD version 5.33 March 2012 was searched for fragments, defined in Figure 3.10. The bond lengths, bond angles, and torsion were analysed and compared with the experimental structure using histogram plots.

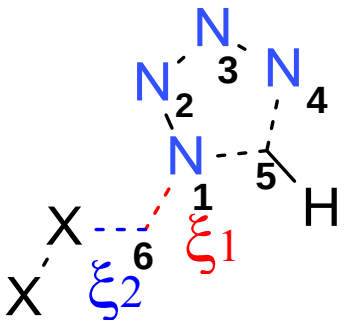


Figure 3.10: Query used in Conquest to search the CSD for similar fragments.

The query of the CSD yielded 34 structures, including two redeterminations. The refcodes of these 32 structures were ACIROU, ACIRUA, BEGRIP, CAZWIK, DOKQEA, DOKQIE, EKAJOQ, HISYEO, IQUJEK, ABMIL, LUVPIC, MEVWOZ, MUKMAG, PHTETZ01, PUGLUZ, QALNOH, QALNUN, QALPAV, QALPID, QALPOJ, QALPUP, QALQAW, QALQEA01, QALQIE, QEBPUI01, REVMOV, REZNUG, RIZXEE, SOLHAD, TIJLIH, VUNNUO, and OBAREQ.

The experimental bond lengths were typical of those in the other crystal structures containing the 1-substituted *1H*-tetrazole group (Figure 3.11). Some significant differences (≤ 0.02 Å) were observed between the *ab initio* computed values and the experimental bond lengths, but both were within the observed ranges. The variations in tetrazole bond angles (Figure 3.12) were consistent with the bond length variations for this moiety[176], which was always close to planar. The external torsions angles, ξ_1 and ξ_2, exhibited a wide range of values (Figure 3.12), a result that was consistent with the large range of low-energy conformations (Figure 3.4) despite the range of functional groups bonded to C_6.

No tetrazole clusters were observed for the 32 1-substituted *1H*-tetrazole structures found in the CSD. There were, however, some tetrazole layers in #1BT_120 and the

experimental structures DOKQIE, EKAJOQ, QALPAV, QALPOJ, QALPUP, and REVMOV. All 32 structures had the tetrazole bonded to a sp^3 carbon and were tetrazole hydrocarbons (except QALPAV, which also contained an iodine atom). Hence, there was no strong driving force for the tetrazole cluster. Still, a layer structure seemed to be quite favourable in cases where there were no competing strong interactions. There were more similarities between the CSD tetrazole structures and the GM structure than there were between the experimental structures. This finding suggests that the use of informatics of CSD in CSP, such as Desiraju's supramolecular synthon approach[183], would weight the global minimum structure more than the experimental structure.

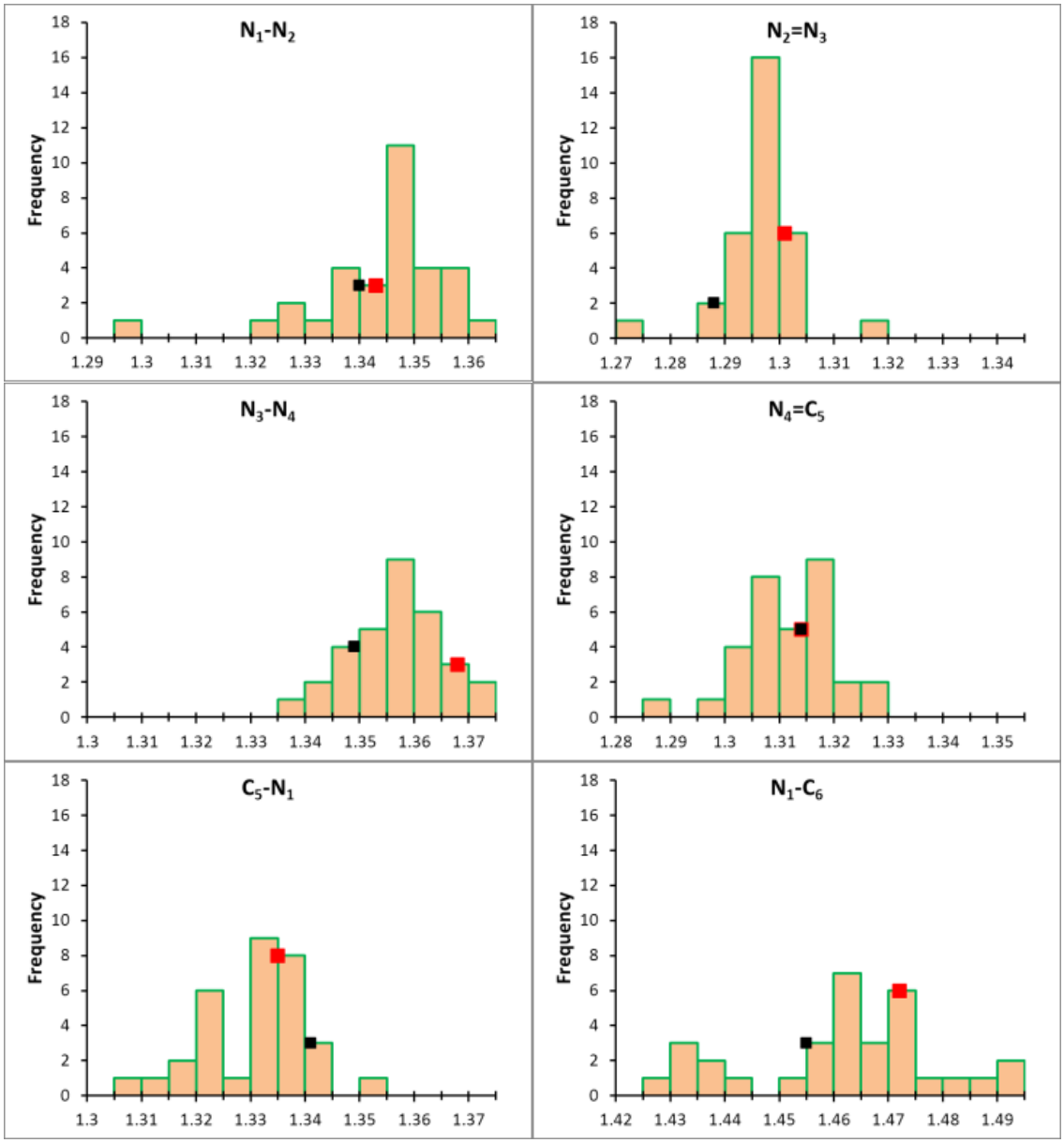


Figure 3.11: Histograms showing the bond length distributions of 32 1-susbtituted *1H*-tetrazole rings in the CSD. The red square denotes the experimental bond lengths in BT, while the black square denotes the *ab initio* bond lengths in the corresponding calculated structure #2BT_23.

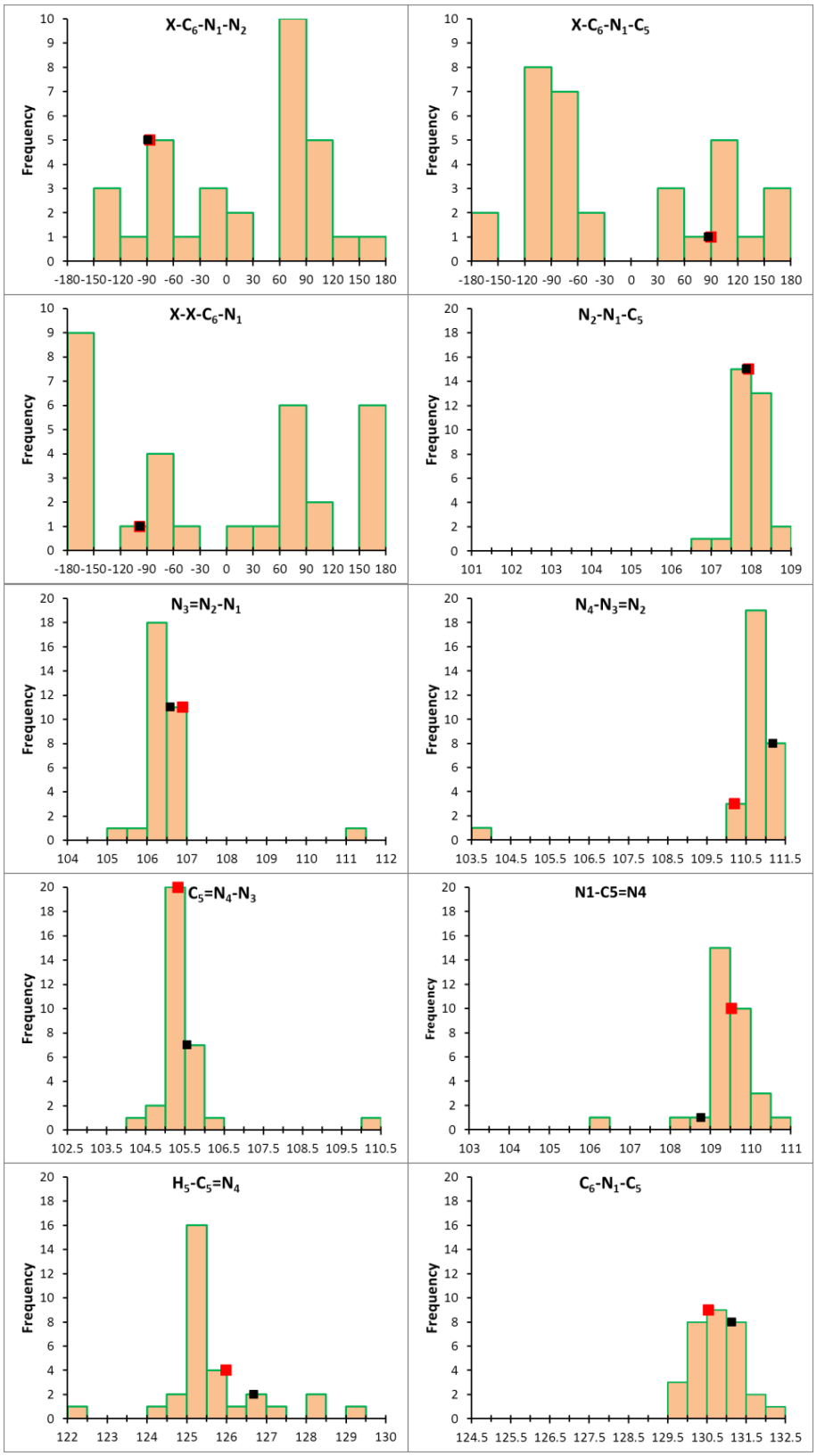


Figure 3.12: Histograms showing the distributions of the torsions and angles of the 32-tetrazole rings found during the CSD survey. The experimental bond length in BT is denoted by the red square, while the *ab initio* bond length in the corresponding calculated structure #2BT_23 is denoted by the black square.

3.4. Summary

I successful predicted the unusual crystal packing of BT given just the chemical diagram and the Z′ number. The packing was predicted by using a model that accurately determines the electrostatic forces arising from molecular charge density including the anisotropic forces from the lone pair and π electrons, but otherwise was not tailored to the tetrazole···tetrazole interactions. Thus, the intermolecular energy potential is transferable to BT crystal structure even though the functional group used in determining the parameters of the intermolecular potentials did not include a single tetrazole molecule.

The successful prediction of the experimental structure as one of the two distinct most stable structures shows that the unusual layers do present an optimal compromise between the many different weak hydrogen bonds and other intermolecular interactions. The most stable crystal structure, GM, could well have been an undiscovered polymorph. In fact, informatics weighting may have selected this structure. Thus, this study shows the potential value of CSP for generating unexpected structures.

This study also showed that the level of computational chemistry modelling that was used is suitable for modelling the molecular recognition of this functional group. This level of modelling could become increasingly useful in pharmaceutical design due to tetrazoles' bioisosteres properties. In the next chapter, more will demanded of CSP. Will this method deliver?

Chapter 4. Is the fenamate group a *polymorphophore*? Contrasting the crystal energy landscapes of fenamic and tolfenamic acids

"*It's the little details that are vital. Little things make big things happen.***"**

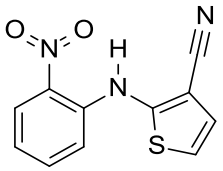
— John Wooden

4.1. Introduction

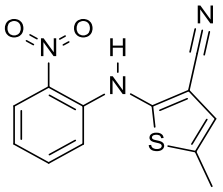
This chapter builds on Chapter 3's description of a procedure that successfully predicted the experimental crystal structure of a flexible molecule, in order to address two specific questions. First, in a given molecular structure, is the presence of a *polymorphophore* (section 1.1) substructure sufficient to ensure polymorphism? Second, what role do substituents play in determining the range of polymorphs that are present in a molecule?

In this investigation of the concept of *polymorphophore,* I will use CSP to contrast the crystal energy landscapes of two fenamates (Figure 1.3): monomorphic fenamic acid **(**2-(phenylamino)-benzoic acid, FA**)** and one of its highly polymorphic derivatives, tolfenamic acid **(**2-[(3-chloro-2-methylphenyl)amino]-benzoic acid, TA**)**. TA is a good test case for this investigation into whether the fenamate group is a *polymorphophore* because at the start of this study, TA had five known polymorphs[19].

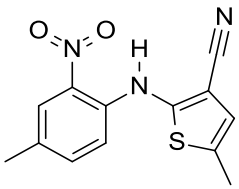
The term "fenamate" can refer to a range of molecules that vary widely in molecular structure, including niflumic acid, which contains an aromatic nitrogen. Although some researchers who have conducted thermodynamic and structural investigations of fenamate molecular crystals have included an extraordinarily wide range of molecules, even dichlofenac[184], the present study is restricted to the fenamates shown in Figure 1.3. Essentially, I will focus on cases where the dominant interactions that determine the crystal or drug binding are inherent in the proposed *polymorphophore* families of fenamic acid (Figure 1.3).



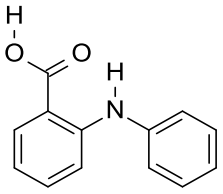
ROY fragment
WIFHUP 2[185]
R=H



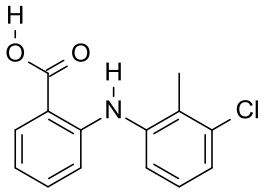
ROY
QAXMEH 7(3)[20]
R_2=CH_3



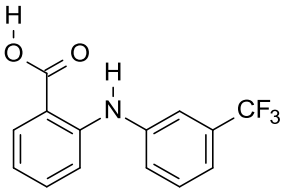
OGIDEN 3[186]
R_2=R_5=CH_3



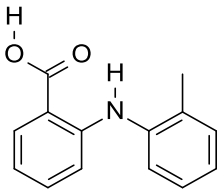
Fenamate fragment
Fenamic acid (FA)
QQQBTY 1[187]
R=H



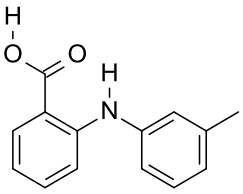
Tolfenamic acid (TA)
KAXXAI 5[19]
R_1=CH_3, R_2=Cl,



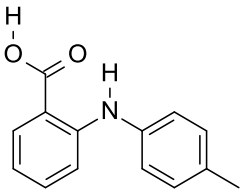
Flufenamic acid (FFA)
FPAMCA
2(1)[22]
8(1)[18] in 2012
R_2=CF_3



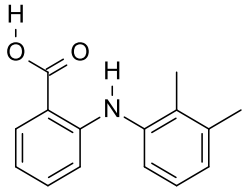
PEFMET 1[188]
R_1=CH_3



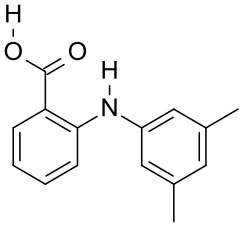
PEFNAQ 1[188]
R_2=CH_3



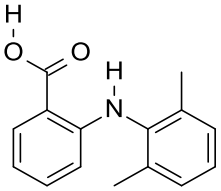
SURMEY 1[187]
R_3=CH_3



Mefenamic acid
XYANAC 2(1)[189,190]
R_1=R_2=CH_3



SURMOI 1[187]
R_2=R_4=CH_3



LAHLOW 1[191]
R_1=R_5=CH_3

Figure 4.1: The *polymorphophore* families of ROY and fenamic acid. The numbers of polymorphs in the Cambridge Structural Database (CSD) follow the refcodes for each specific molecule, and (*x*) denotes the number of additional polymorphs that whose crystal structures were not available. The lowest two rows are fenamates whose structures are used in Table 4.6.

4.1.1. Polymorphism and Polymorphophores

Polymorphism (section 1.1), the occurrence of multiple crystal structures with the same chemical content, is now known to be a widespread phenomenon among organic molecules[192], and is of great importance in the industrial manufacture of specialty chemicals and pharmaceutical products. Relatively few systems have several polymorphs whose crystal structures and relative stabilities have been determined, the most thoroughly investigated[20] being the polymorphs of 2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile, commonly known as ROY (Figure 1.3), after the red, orange and yellow colour spectrum of the first six polymorphs to be reported[193]. Some families of related molecules[23,194], such as the sulphonamides[194], ROY derivatives[23], barbiturates[195], carbamazepine derivatives[147,196], and fenamates[19] appear to have a strong tendency toward polymorphism. This fact has given rise to the study of a *polymorphophore*[19,23,196], a term coined by Matzger[23] that refers to a structural element that, when incorporated into a molecule, favours the formation of polymorphic crystal forms. Hence, *polymorphophore* is a descriptive term and not a quantifiable hypothesis.

The focus of this chapter is the *polymorphophore* concept. Some families of molecules contain a common substructure (the *polymorphophore*) wherein many members, but not necessarily all, exhibit polymorphism. For example, removing or adding a methyl group to ROY creates molecules (Figure 1.3) that are also polymorphic, indicating that ROY is a *polymorphophore* (Figure 1.3). However, the number of related molecules is limited, probably because ROY is only a precursor in the synthesis of the blockbuster drug olanzapine. Studying *polymorphophores* is, like studying *pharmacophores*, a particularly promising way of finding new leads in drug discovery, especially in cases when a receptor's 3D structure is unknown[197].

This chapter considers the fenamates that conserve the H-bonding groups and dominant molecular shapes by having small, non-polar substituent groups for the aromatic protons, as exemplified in Figure 1.3. These types of substituent(s) might be expected to have minimal effects on the strongest intermolecular interactions, and hence on either crystal packing or drug receptor binding. The fenamates contain a proposed *polymorphophore*[18,19] that is related to that of ROY (Figure 1.3) in having a phenyl group attached via an N-H group to a further aromatic ring.

4.1.2. Conformations in known Crystal Structures

The overlay of the conformers of FA and TA in the experimental crystal structures (Figure 4.2) confirms that the main variation in conformation is the ξ_1 torsion angle. The

two most stable polymorphs of TA, forms I and II (blue and red in Figure 4.2), have the greatest difference in ξ_1. The benzoic acid ring and the N-H groups are coplanar, with the ξ_2 torsion angle being approximately zero, concordant with the intramolecular H-bond.

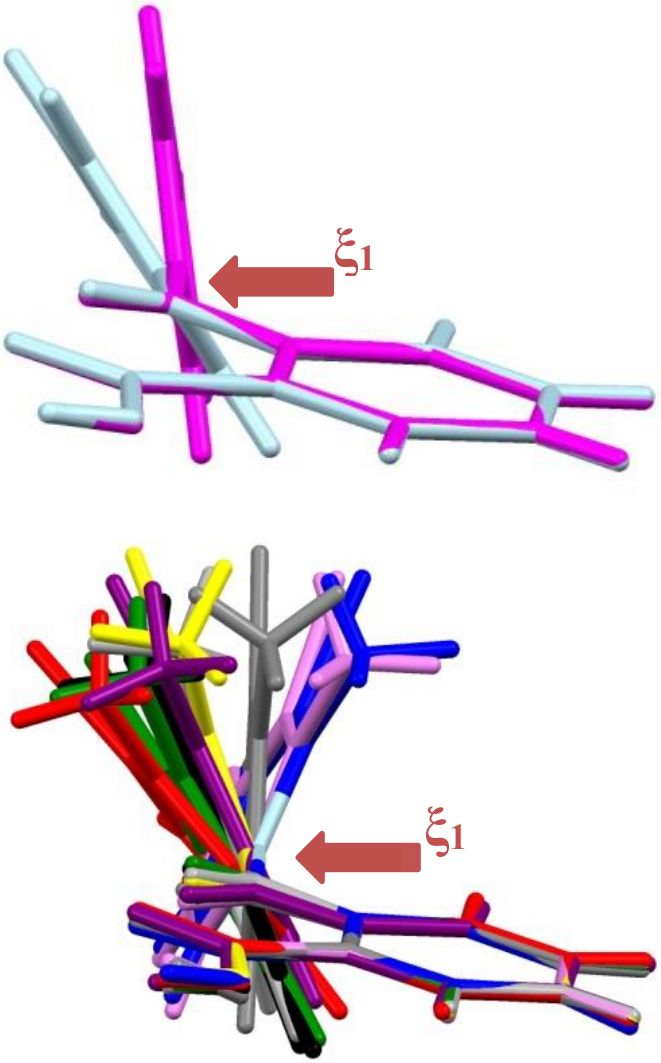


Figure 4.2: Overlay of the experimental conformers of FA (top) and TA (bottom), viewed with the overlaid benzoic acid fragment arranged horizontal. FA: QQQBTY02 (Z′=2; light blue and magenta). TA: KAXXAI01 (Form I; Z′=1; blue), KAXXAI (Form II; Z′=1; red), KAXXAI02 (Form III; Z′=2; green and purple), KAXXAI03 (Form IV; Z′=3; yellow, brown (obscured by yellow) and black), and KAXXAI04 (Form V; Z′=1; disordered on both sites (section 4.1.5); violet and grey). The arrows point to the main variation (ξ_1) between the conformers of FA and TA in the experimental crystal structures.

4.1.3. Experimental Data of FA and TA

FA has just one known form[187], which has two molecules in its asymmetric unit, Z. FA serves as a parent structure for several fenamates including TA. TA is a non-steroidal anti-inflammatory drug used for the treatment of rheumatic conditions[198]. At the start of this study, TA had at least five polymorphs, forms I-V[19,199]. It currently has six polymorphs (see section 4.3.8). Unlike ROY, the polymorphs of TA vary in the number

of molecules in the asymmetric unit, Z' (Table 4.1). The variation in colour that is observed between form I (colourless) and II (yellow) is due to the conformational difference, as observed for ROY[20,23]. There were three undisputedly metastable polymorphs, including form V, whose disorder is analysed further in section 4.1.5.

Here it is necessary to discuss further the relative stability of forms I and II (Table 4.1). The enthalpy and temperature of fusion, ΔH_{fus} and T_{fus}, measurements provide a way of classifying polymorphs as either being monotropic or enantiotropic. Four different papers[19,199-201] have presented evidence for the thermodynamic relationship between forms I and II, but each paper gives slightly different conclusions. Surov et al. showed that the yellow form (form II) is more stable than the white form (form I)[201] and that these polymorphs have a monotropic relationship because the form with the higher melting temperature, form II, has a higher enthalpy of fusion, ΔH_{fus} (by the enthalpy of fusion rule)[202], as illustrated in Figure 4.3(a). However, Mattei and Li[200] measured an endothermic transformation going from II to I with temperature increasing to 141.8 °C with a Differential Scanning Calorimetry (DSC), implying that forms I and II have an enantiotropic relationship, as shown in Figure 4.3(b). The observed transformation from form II to I in a slurry experiment indicated that the thermodynamic transition point was below room temperature. All three sets of measurements[199-201] that were extrapolated to 0 K had form II more stable than form I.

Table 4.1: The experimental data of the known forms of FA and TA. Crystal structures and ΔH_{fus} values were taken from ref. [19] unless otherwise specified. [a]The most widely varying torsion angle, ξ_1, is defined in Figure 4.9. [b]Temperature of crystal structure determination. [c]Melting temperature.

REFCODE (form)	Space Group	Z'	ξ_1[a] / °	T[b] / K	ΔH_{fus} / kJ mol^{-1}	T_m[c] / K
FA						
QQQBTY02	*P-1*	2	47.21, 70.45	283-303[187]	-	-
TA						
KAXXAI01(I)	*P2$_1$/c*	1	107.74		39.37	486.25-488.6
				110	42.50	484.05[199]
					41.0±0.5	484.18±0.2[201]
					41.27	485.85[200]
KAXXAI(II)	*P2$_1$/n*	1	42.23		38.7	486.67
				110	48.40	487.65[199]
					49.0±0.5	485.78±0.2[201]
					42.40	414.95[200]
KAXXAI02 (III)	*P2$_1$/c*	2	44.19, 57.64	85	-	
KAXXAI03 (IV)	*P-1*	3	67.34,57.58,47.84	85	31.88	
KAXXAI04 (V)	*P-1*	1	55.62	85	-	

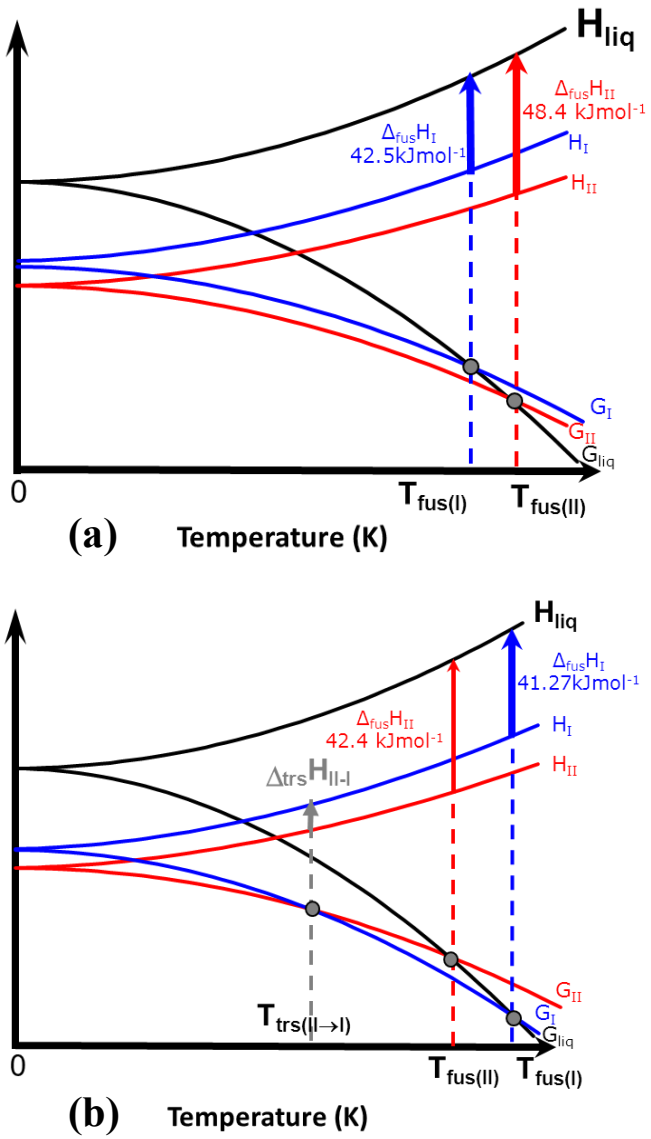


Figure 4.3: Temperature dependence of the enthalpy (H) and free energy (G) diagrams for forms I and II of TA, showing a monotropic (top) relationship from data from Surov[201], which are qualitatively the same as data from Andersen[199], and an enantiotropic (bottom) relationship from data from Mattei and Li[200].

Lopez-Mejias et al.[19] provided the only thermodynamic data available for forms III–V. The researchers noted that heating forms III and V at 80°C caused a transformation to form I within a period of minutes to hours. In solution, forms III, IV and V transformed to form I or a mixture of form I and II. Hence, forms III, IV and V were shown to be metastable polymorphs. The heats of fusion given in Lopez-Mejias' supplementary information[19] were listed by the initial form, and forms III and V may well have transformed to form I rather than form II. Therefore, these sets of data were inconclusive as to the relationship between forms I and II.

4.1.4. Similarities between the known forms of FA and TA

The known polymorphs of TA and FA showed some similarities (Table 4.2), as all were based on the H-bonded dimer (Figure 4.4a). The graph-set analysis[161,162] for the (two) types of H-bonds are identical for all the crystal structures of TA and FA: intermolecular dimers with 2 donors, 2 acceptors and 8 atoms in total ($R_2^2(8)$) and intramolecular with 1 donor, 1 acceptor and 6 atoms in total ($S(6)$)[201].

Form II had the most distinctive packing (Figure 4.4b) as well as conformation (Figure 4.2b) and was the only structure with a short contact involving the chlorine atom and the aromatic ring. The most striking result from Table 4.2 was the similarity between the metastable forms III and IV, where 11 molecules could be overlaid (Figure 4.4c). The two independent molecules in form III had similar conformations to two of the three independent molecules in form IV (compare the open circle and closed triangle symbols in Figure 4.12a). The polymorphs had sufficiently similar simulated powder patterns to have a PXRD similarity measure in the grey area between polymorphism and redeterminations[203], and for this reason, they exemplify the challenges in defining polymorphism[204,205].

Table 4.2: Quantification of the similarities of FA and TA crystal structures showing the packing similarity ($RMSD_n$) and powder X-ray diffraction (PXRD) similarity[206]. The bold numbers indicate the number of molecules, n, that match within distance and angle tolerances of 20% and 20° respectively, with the $RMSD_n$ values in brackets. See section 4.2.5.

	n(RMSD$_n$/Å)				
	FA	**TA(I)**	**TA(II)**	**TA(III)**	**TA(IV)**
FA	-	**6**(0.66)	**2**(0.48)	**3**(0.74)	**5**(1.08)
TA(I)	0.78	-	**1**(0.87)	**3**(0.78)	**3**(1.07)
TA(II)	0.85	0.86	-	**2**(0.77)	**2**(0.74)
TA(III)	0.78	0.83	0.90	-	**11**(0.45)
TA(IV)	0.78	0.84	0.92	0.98	-
TA(V)	0.86	0.89	0.93	0.91	0.93
	PXRD Similarity				

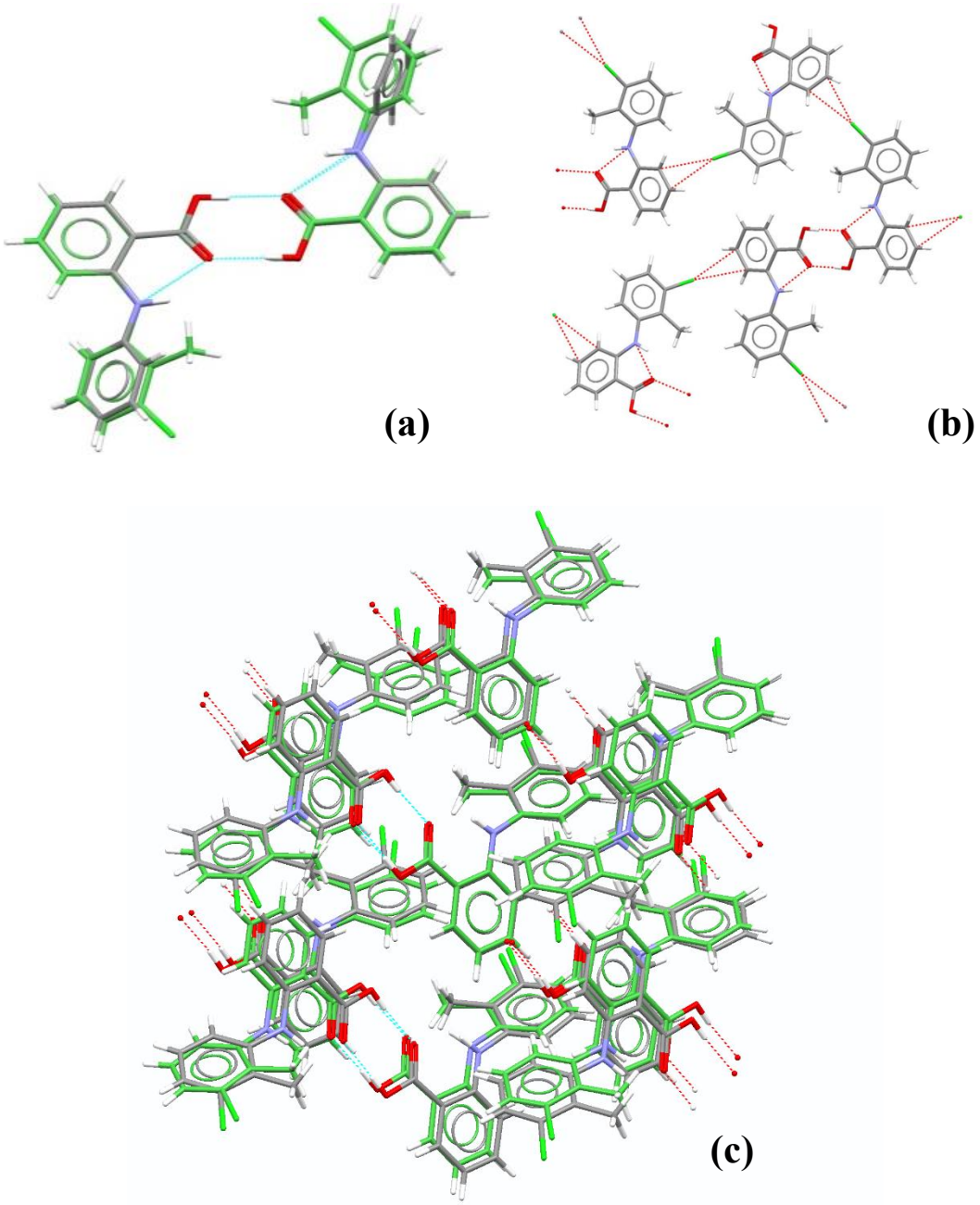


Figure 4.4: Features of the experimental crystal structures: (a) The hydrogen-bonded dimer motif found in all structures with variations in the phenyl ring orientations (ξ1). The specific illustration shows the packing similarity between FA (QQQBTY02) and TA form II (KAXXAI). (b) The unique packing of TA form II with short contacts between chlorine and the aromatic ring. (c) Crystal Packing Similarity between form III and form IV of TA showing the 11-molecule overlay, which includes only two dimers.

4.1.5. Solving the Connectivity Problem of a Disordered Form

The crystal structure of Form V of TA was solved with an R-factor of 6.58%; the whole molecule was equally disordered over two sites[19]. I found that separating the two components of disorder gave the wrong connectivity between the C_8 and N_1 atoms (green in Figure 4.5b). The C_8-N_1 bond length of 1.70 Å was unusually long, and this caused the breakdown of the trigonal planar relationship of nitrogen, C_7-$N_1(H_6)$-C_8, connectivity

(Figure 4.5c). A model with the two components both corresponding to the expected molecular geometry can be made by redefining the connectivity (Figure 4.5b) to link half of one disordered molecule (the bonded molecule in Figure 4.5a) with half of the other disordered molecule (the atoms-only component in Figure 4.5a). The second component was formed by combining the remaining halves. The resulting molecules (red and blue in Figure 4.5b) have the expected C-N bond length of 1.43 Å and a trigonal planar geometry around the nitrogen atom.

The above procedure results in three ordered Z=2 crystal structures with the unit cell of form V by using the two resulting molecules (Figure 4.6a, b for Z′=1 and Figure 4.6c for Z′=2).

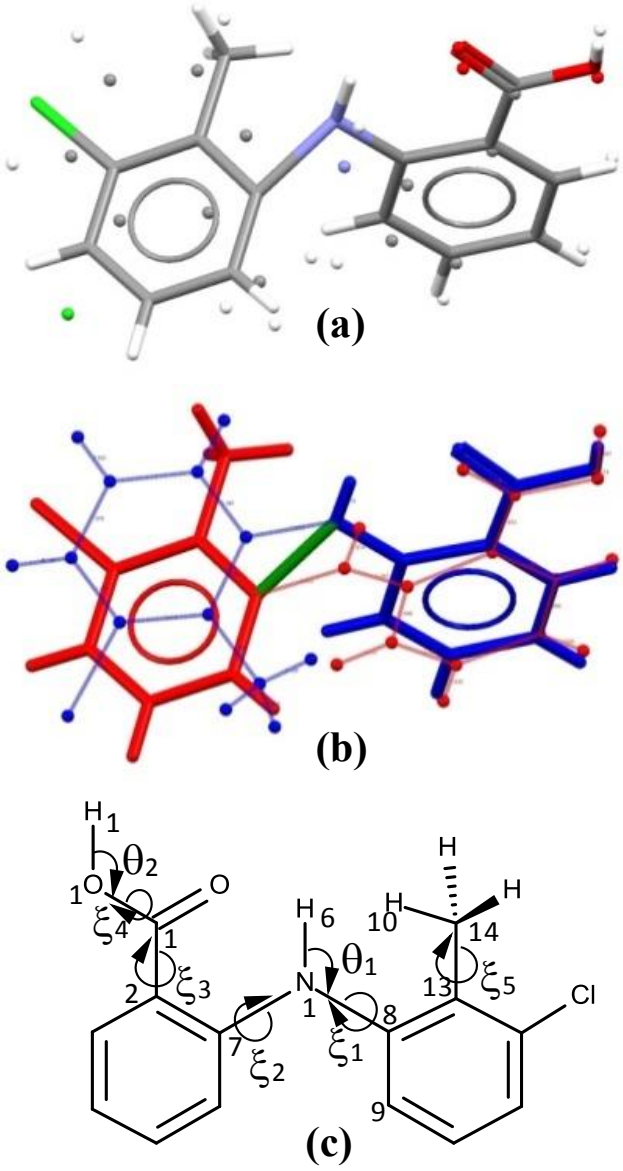


Figure 4.5: (a) Crystal structure KAXXAI04 of form V of TA (b) The atom connectivity of the two new components (colour red and blue); the original C-N connectivity is labelled green (c) Chemical diagram showing the atomic connectivity of TA.

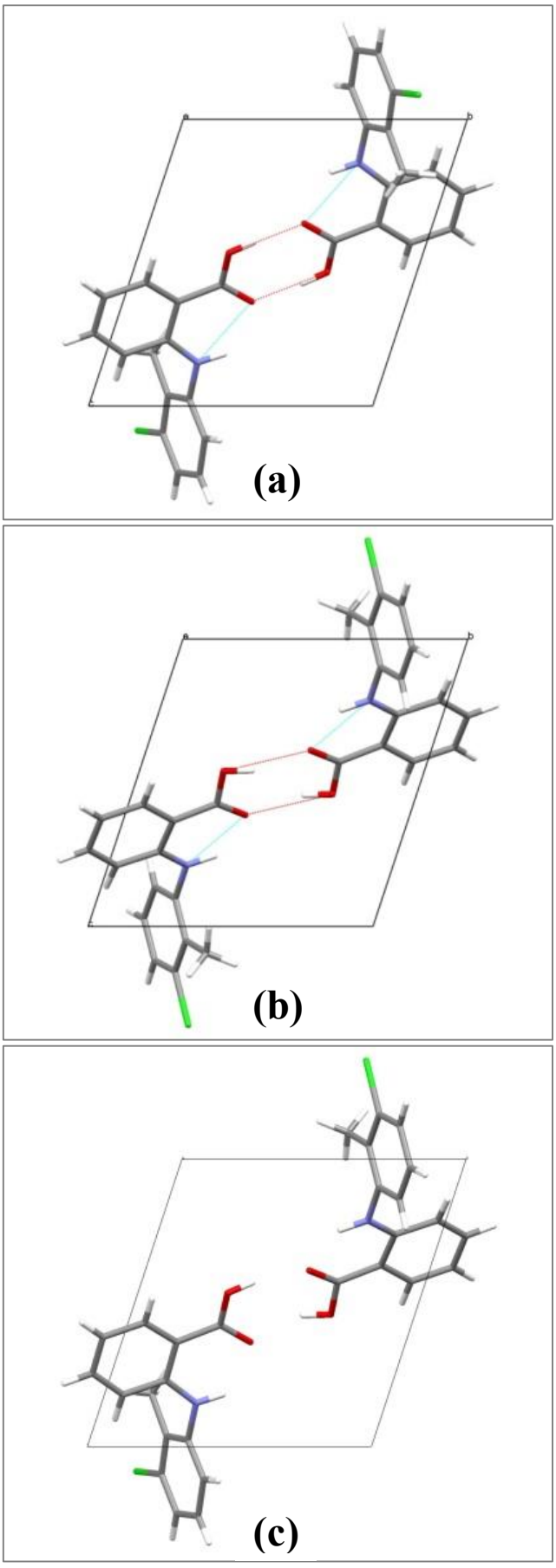


Figure 4.6: Ordered Z=2 crystal structures formed from the disordered components of TA form V. (a) Form V_*a*, (b) Form V_*b*, and (c) Form V_*c*.

4.1.6. Overview

The focus of this chapter is the *polymorphophore* concept, which I will investigate by generating crystal energy landscapes (section 2.8.5) of FA and TA. These crystal energy landscapes are a valuable tool for examining the known forms of different molecules because the same level of polymorph screening is rarely applied to entire families of molecules. This is likely the case for FA and TA, since TA is a drug molecule while FA is not. In addition, there were far more publications for TA than FA (see section 4.1.3). For these reasons, FA and TA provide a good test for the concept of *polymorphophore*.

In this chapter, the energy landscapes of FA and TA will be generated by first selecting a suitable *ab initio* level of theory for the conformational energy of the isolated molecule. I will achieve this by performing *ab initio* scans at various levels of theory. Then, these *ab initio* scans will be compared with histogram plots of crystal structures whose fragments are similar to FA and TA. The analysis of these scans and plots will guide the methodology I will use to generate the hypothetical structures. The final energy landscapes of FA and TA will be generated by systematically increasing the accuracy of evaluating the lattice energies while reducing the number of generated structures. Finally, I will compare the crystal structures on the energy landscapes for FA and TA using a similarity tool.

I will show the conformational flexibility of fenamate fragments allows the potential for a large number of structures, and that it is the substituents that determine if a molecule is polymorphic or not. In other words, *polymorphophore* promotes but does not guarantee polymorphism. Concisely, CSP goes beyond the 'blind test' described in Chapter 3.

4.2. Computational Methods

4.2.1. Ab initio Conformational Profile of TA

To test the effects of *ab initio* methods on the conformation profile of TA, I compared relaxed scans at various levels of theory of the main torsion, ξ_1 (Figure 4.9), with single-point energy calculations at the MP2/6-31G(d,p) level of theory. The effect of rotating the H-bonded torsion, ξ_2 (see Figure 4.9), was investigated by performing a relaxed scan at the PBE0/6-31+G(d,p) level of theory. The results from these scans helped in selecting the appropriate level of theory as well as the torsion that would be allowed to rotate during the generation of hypothetical structures (step 1, section 4.2.3).

4.2.2. CSD Survey for Fenamates

The relaxed conformational scans of the main torsion angle, ξ_1 (section 4.2.1), at the PBE0/6-31+G(d) level of theory were complemented by a Conquest[207] search of the Cambridge Structural Database (CSD [208]) for the crystal structures containing the

molecular fragment shown in Figure 4.7. This comparison allowed us to see the variation in ξ_1 when there was an internal H-bond (Z···H < 2.6 Å) that constrained part of the molecule to planarity.

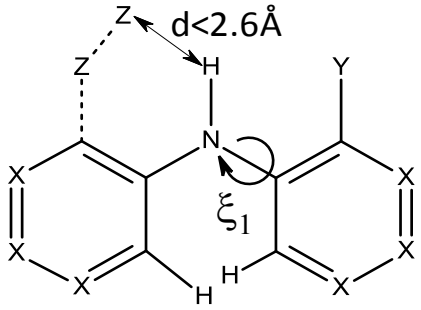


Figure 4.7: Query fragment and the torsion angle ξ_1 measured in the Conquest[207] survey of the CSD structures. There were 131 organic structures where Y=H, and 94 where Y≠ H.

4.2.3. Crystal Structure Prediction

I calculated the crystal energy landscapes[14] of FA and TA using a method[209] that was adapted from that used for the highly flexible molecule XX in the fifth blind test of CSP[11], a method that called for an approximate conformational energy surface in the initial search stage (see section 2.8.1). The final energy landscapes of FA and TA were generated in four steps, with the quality of the methods increasing, and estimated the conformational energy penalty ΔE_{intra} and the intermolecular contribution U_{inter} to the lattice energy, E_{latt}:

$$E_{latt} = U_{inter} + \Delta E_{intra} = U_{inter} + U_{elec} + \Delta E_{intra} \qquad 4.1$$

Step 1 (Generation of hypothetical crystal structure**)**: Crystal structures with Z′=1 were generated in 15 common space groups: *P1*, *P-1*, *P2₁*, *P2₁/c*, *P2₁2₁2*, *P2₁2₁2₁*, *Pna2₁*, *Pca2₁*, *Pbca*, *Pbcn*, *C2/c*, *Cc*, *C2*, *Pc* and *P2/c*, using *CrystalPredictor*[126]. The grid was used to extrapolate the atomic charges and intramolecular energy ΔE_{intra}. This grid was pre-computed as a function of just ξ_1 (Figure 4.8) using GAMESS[179] at the MP2/6-31G(d,p) level of theory (single-point calculation), after the molecular geometries were optimized at the HF/6-31G(d,p) level. The intermolecular contributions to the lattice energy were calculated crudely from the *atomic charges* and the Williams[135] *exp*-6 repulsion-dispersion potential. The *CrystalPredictor* program carried out 250,000 lattice energy minimizations. This resulted in 162,288 and 182,615 distinct structures for TA and FA respectively, of which 16,543 and 22,884 were unique for TA and FA respectively.

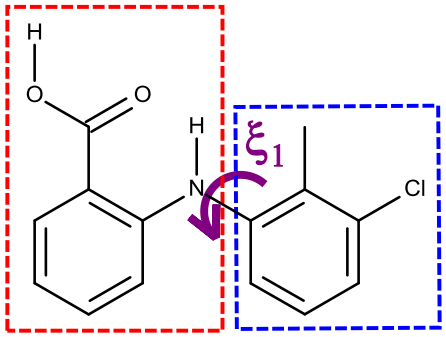


Figure 4.8: The fragment used in the *CrystalPredictor* search for TA. Each fragment was kept rigid during the search for hypothetical crystal structures.

Step 2 (Accurate evaluation of the electrostatic contribution (U_{elec})**)**: Single-point *ab initio* calculations were carried out on the conformation for each of the unique structures from **step 1**, using GAUSSIAN03[53] at the PBE0/6-31G(d,p) level of theory, to improve the estimate of ΔE_{intra} and provide the *atomic multipoles*, by distributed multipole analysis[51] using GDMA. The crystal structures were re-optimized using DMACRYS[34], keeping the molecule rigid, with the intermolecular energy U_{inter} calculated from the distributed multipoles and the FIT parameterization[38-40] of the *exp-6* atom–atom repulsion–dispersion potential.

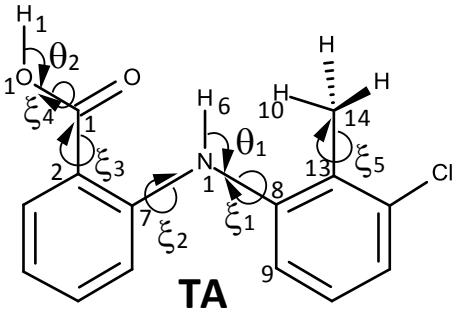


Figure 4.9: The degrees of freedom optimized by *CrystalOptimizer* for crystal structures of FA and TA ($\xi_1 \equiv C_7\text{-}N_1\text{-}C_8\text{-}C_9$, $\xi_2 \equiv C_2\text{-}C_7\text{-}N_1\text{-}C_8$, $\xi_3 \equiv O_1\text{-}C_1\text{-}C_2\text{-}C_7$, $\xi_4 \equiv H_1\text{-}O_1\text{-}C_1\text{-}C_2$, and $\xi_5 \equiv C_8\text{-}C_{13}\text{-}C_{14}\text{-}H_{10}$, $\theta_1 \equiv H_6\text{-}N_1\text{-}C_8$, $\theta_2 \equiv H_1\text{-}O_1\text{-}C_1$).

Step 3 (Accurate evaluation of lattice energy ($E_{latt}=U_{inter}$ and ΔE_{intra})**):** Approximately 100 of the most stable crystal structures from **step 2** were re-minimized due to the limited time constraints of submitting this study as my first year report. Six and seven conformational degrees of freedom, for FA and TA respectively (Figure 4.9), were allowed to change, as was the crystal structure, using *CrystalOptimizer*[71,72] and calculating ΔE_{intra} and the distributed multipoles at the PBE0/6-31+G(d) level of theory for the isolated molecule.

Step 4 (Modelling the effect of the crystalline environment**):** Finally, I estimated the effect of the crystal environment using a Polarizable Continuum Model (PCM)[180] that I

obtained by calculating the wavefunction of the molecular structure in a dielectric of ε=3 (typical of molecular organic crystals)[131], as implemented in GAUSSIAN03. The crystal structures were then re-minimized using the distributed multipoles calculated in the PCM environment while keeping the molecule rigid; the resulting U_{inter} was combined with the ΔE_{intra} to evaluate the lattice energies. The CSP-generated crystal structures for FA were labelled #*x*FA_*y*, where *x* was the final energy ordering and *y* the ordering after the *CrystalPredictor* stage. A comparison of *x* and y shows the extent of re-ranking by increasing the accuracy of the lattice energy model.

4.2.4. Experimental Crystal Structures with Z´>1

The complexity of the task of generating Z′>1 crystal energy landscapes (**steps 1-4**) for flexible molecules limited the search (section 4.2.3) to Z′=1. Analogous *CrystalOptimizer* (step 3) and PCM calculations (step 4) were performed starting from all the experimental crystal structures of TA and FA, which included Z′>1 structures: FA form I; and TA forms III, IV and V (section 4.1.5). This allowed comparison between Z′=1 generated crystal structures and the experimental crystal structures of FA and TA.

4.2.5. Comparison of Crystal Structures

The experimental crystal structures, the corresponding lattice energy minima, and the structures on the crystal energy landscapes were compared using the Crystal Packing Similarity[151] module (section 2.9) implemented in Mercury[210]. This crystal structure similarity tool returns the highest number of molecules *n* (where $n \leq 15$) that can be overlaid between two different structures when all non-hydrogen atom–atom distances are within a 20% distance tolerance and angles are within 20°. The calculated $RMSD_n$ is the root mean square deviation of all non-hydrogen atom positions in the clusters of *n* molecules. The program allows for comparisons of crystal structures of different molecules, with the $RMSD_n$ being calculated from only the common non-hydrogen atoms. This Crystal Packing Similarity calculation determines the similarity in the coordination environment of the two crystal structures, rather than of the crystallographic cell.

4.3. Results

4.3.1. Dependence of Conformational Energy on Method

An appropriate grid of intramolecular energies was essential in order to search all the likely conformations of TA and FA and thereby generate all likely Z′=1 crystal structures. To observe the effects that changing the level of theory had on the relative intramolecular energy, the ξ_1 torsion angle was scanned at HF, PBE0, MP2, and B3LYP levels of theory using GAUSSIAN03[53] with the rest of the molecule allowed to relax.

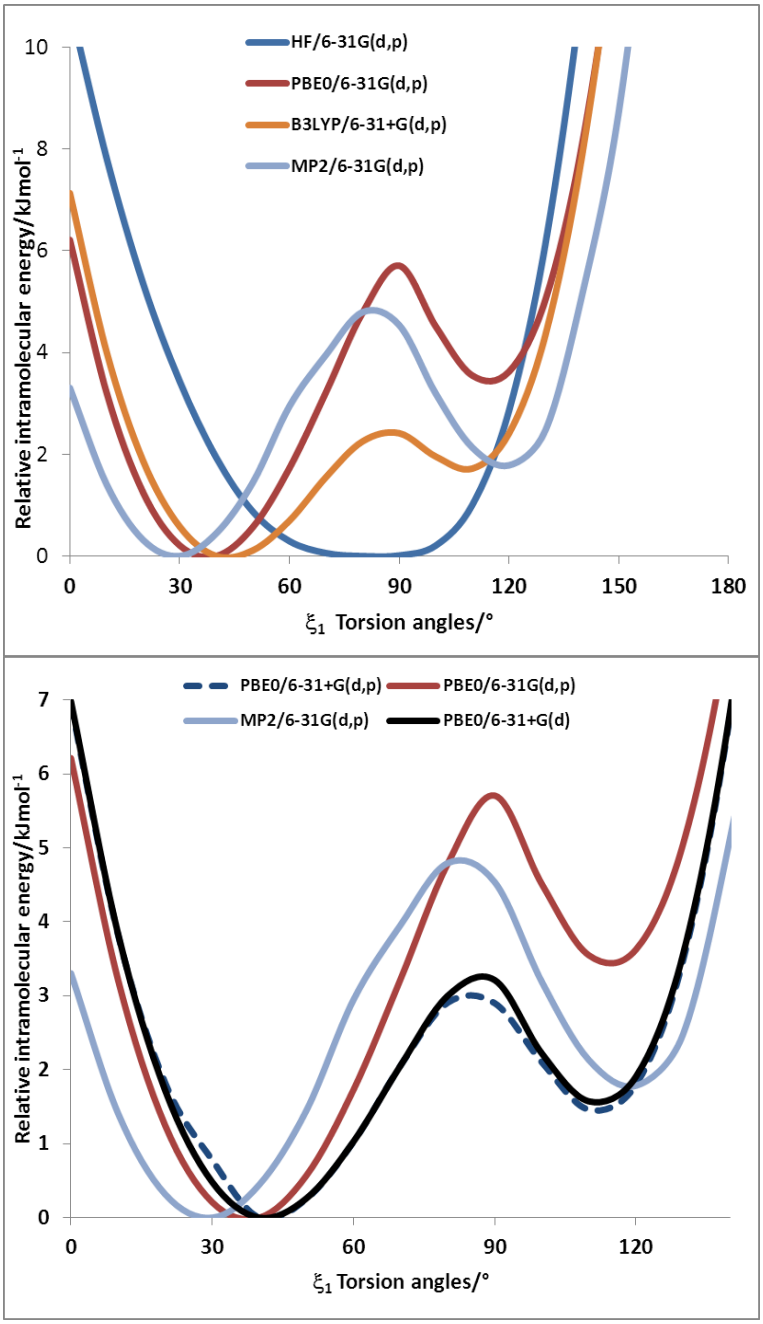


Figure 4.10: The relative conformational energy of isolated TA molecules as a function of the ξ_1. The MP2 and PBE0 method at 6-31G(d,p) basis set compared with the other method (top) and the other basis sets (bottom) for the PBE0 method.

The HF/6-31G(d,p) level of theory produced one minimum for the ξ_1 torsion angle (Figure 4.10a). This result was qualitatively incorrect, so the PBE0 level was selected as the best compromise because it reproduced the correlation effect present in TA, its minima coincided in the same torsion angle region as that of MP2, and was computationally cheaper than MP2. Since the conformational profile of TA converged at the 6-31+G(d) basis set (Figure 4.10b), I selected PBE0/6-31+G(d) as the appropriate level of theory for the final lattice energy calculation (steps 3 and 4 of section 4.2.3) of the crystal structures. This level of theory produced the standard plots of Chapters 4 to 6.

4.3.2. Consideration of H-bonded Torsion

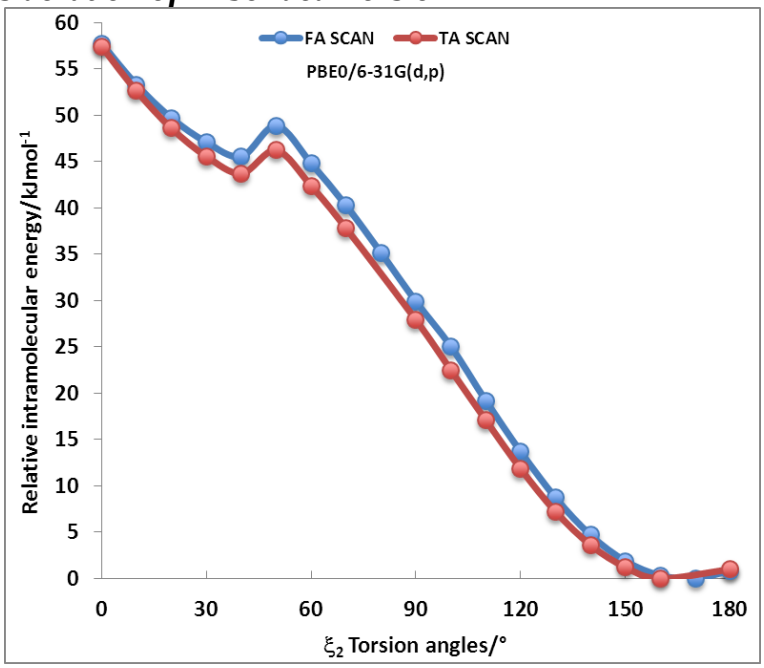


Figure 4.11: The relative conformational energy as a function of the ξ_2 (Figure 4.9) for FA and TA, calculated at the PBE0/6-31G(d,p) level of theory.

The conformational energy scans of ξ_2 (Figure 4.11) of TA and FA are very similar. The decrease in the torsion angle from 180° strained—and eventually broke—the intramolecular H-bond between the carbonyl group and N-H. This H-bond strain accounts for the huge energy penalty observed when this torsion angle changed. Therefore, ξ_2 torsion was kept rigid during the search for hypothetical crystal structures, even as the torsion was optimized within the crystal structures to enable an accurate evaluation of the balance between intermolecular and intramolecular energy in step 3 (section 4.2.3). Chapters 5 and 6 present a more thorough investigation of the conformational energy scans of fenamates.

4.3.3. CSD Survey of Fenamates

The conformational potential energy surfaces of isolated TA and FA molecules (Figure 4.12a) were similar for a wide range of conformations around those found in experimental structures (Figure 4.2), but were different as the bulky methyl substituent of TA approached the benzoic acid ring (Figure 4.12a). For FA, the energy penalty for conformational change, ΔE_{intra}, was less than 6 kJ mol^{-1} for a complete rotation of the phenyl ring, whereas for TA, there was a conformational region that was far too high in energy to occur in crystal structures. The observed conformations of FA and TA in their crystal structures corresponded to low-energy isolated molecule conformations (Figure 4.12a), with a relative intramolecular energy, ΔE_{intra}, of less than 3 kJ mol^{-1}. The conformational profile for TA was in qualitative agreement with that calculated by other high-quality *ab initio* methods (see MP2 and B3LYP in Figure 4.10), but the profile was quantitatively affected by the subtle balance between intramolecular dispersion and intramolecular basis set superposition error[60]. Other studies have also suggested that the barriers to conformational change would be lower in solvents from calculations in a dielectric continuum simulating CCl_4 and methanol[211].

The crystal structures in the CSD that contain the search criterion of Figure 4.7 also show the same preference for the lower energy conformations. That is, the distribution maxima coincide with the conformational minima (Figure 4.12b, c), with very few structures having the two rings near coplanar or perpendicular (ξ_1=0, 90°) around the local maxima in the conformational energy scans. For the Y=H search, because both ortho substituents were hydrogen atoms, two torsion angle values arose per conformation in a crystal structure (one value between 0 and 90° and the second value between 90° and 180°), giving a symmetric plot (Figure 4.12b). For Y≠H, on the other hand, each structure had only one torsion angle value and an asymmetric distribution (Figure 4.12c), reflecting the steric hindrance of the ortho substituent.

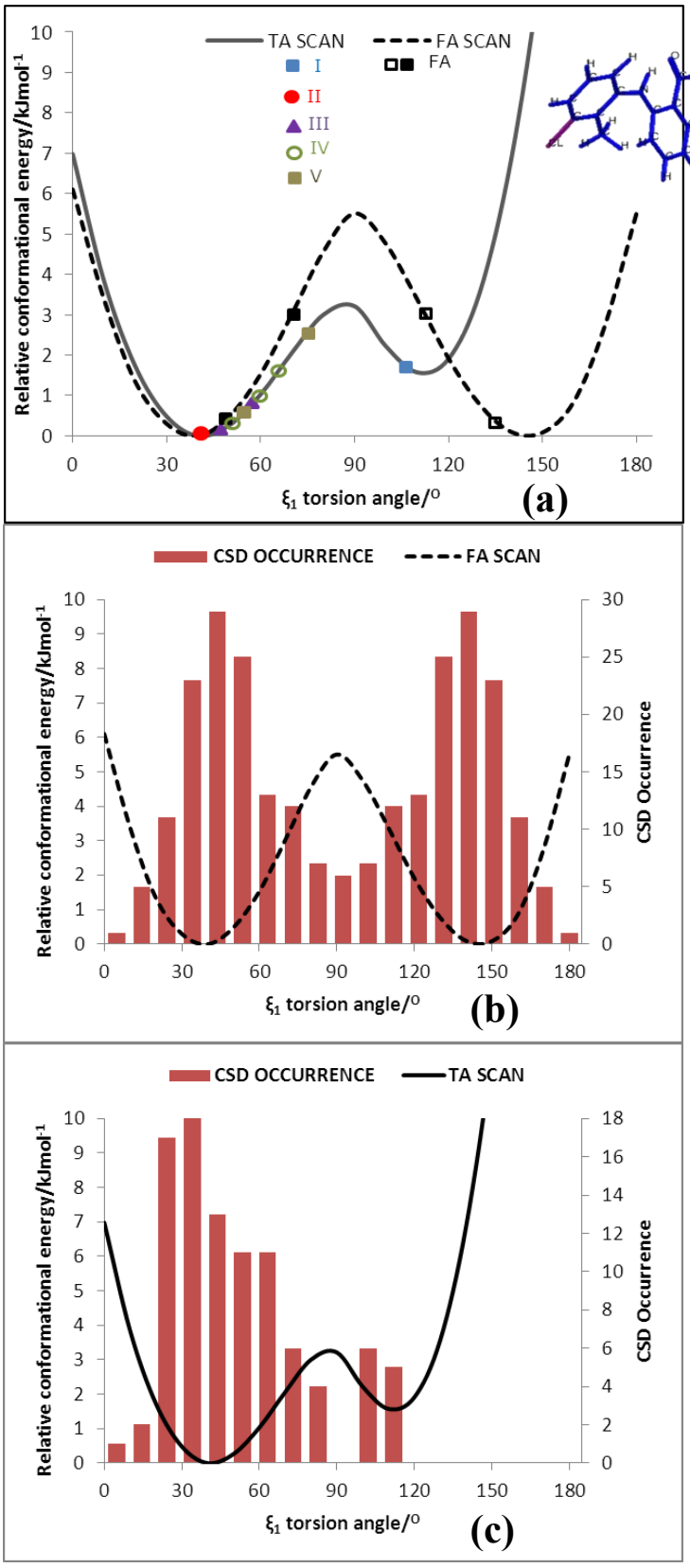


Figure 4.12: The relaxed conformational energy scan of FA and TA as a function of ξ_1 at the PBE0/6-31+G(d) level of theory. (a) Includes the experimental values of ξ_1 marked for TA and FA, with the symmetry equivalent FA conformations denoted by open black squares and an insert diagram showing the steric hindrance for TA in the high-energy region. A histogram of (b) 131 crystal structures found in the CSD search (Figure 4.7) with Y=H plus symmetry equivalent structures and (c) 94 crystal structures with Y≠H.

4.3.4. Validation of the CrystalOptimizer plus PCM Models

The computational model (steps 3 and 4 of section 4.2.3) is able to reproduce all ordered crystal structures satisfactorily as minima in the lattice energy (Table 4.3), with these static 0 K structures overlaying the finite temperature crystal structures with similar accuracy. The range of lattice energies is less than 2 kJ mol^{-1}, consistent with both the small energy range expected for polymorphs and with the thermodynamic data (Table 4.1). Form I is the most stable at 0 K, in disagreement with the monotropic relationship between forms I and II, as determined by DSC measurements[199,201] and other data (section 4.1.2). Forms III and IV are metastable and very close in energy to form II. Of the three form V Z=2 ordered model crystal structures, two have a nearby lattice energy minimum. Because all are quite high in energy, the disordered structure is likely to be more complex than a combination of these structures. However, a full symmetry adapted ensemble study of a large supercell[143,212] would not be appropriate without more detailed experimental studies.

Table 4.3: Comparison of the lattice energy minimum with the experimental structure, used as a starting point for the final lattice energy model (PBE0/6-31+G(d), PCM), for the ordered structures of FA and TA. The three form V Z=2 model structures derived from the disorder components of TA (section 4.1.5) are also compared with the starting model. RMSD$_n$ corresponds to n=15 unless (n) is given.

	Space Group	ρ gcm^{-3}	E_{latt} kJ mol^{-1}	a b c Å	α β γ °	RMSD$_n$ Å
FA						
Expt	P-1	1.33		8.08; 9.81;14.04	85.96; 88.64; 73.45	
Opt	P-1	1.34	-136.63	8.48; 9.91;13.33	90.94; 88.16; 71.30	0.46
TA						
Expt I	$P2_1/c$	1.44		4.83;32.13; 8.04	90.00;104.88; 90.00	
Opt I	$P2_1/c$	1.39	-147.24	4.86;31.54; 8.32	90.00;102.00; 90.00	0.33
Expt II	$P2_1/n$	1.45		3.84;22.00;14.21	90.00; 94.11; 90.00	
Opt II	$P2_1/n$	1.41	-145.21	3.86;22.06;14.60	90.00; 96.21; 90.00	0.26
Expt III	$P2_1/c$	1.44		7.64;11.31;28.07	90.00; 93.03; 90.00	
Opt III	$P2_1/c$	1.39	-145.55	7.83;11.64;27.48	90.00; 93.32; 90.00	0.30
Expt IV	P-1	1.44		7.52;14.33;17.59	103.68; 98.25; 93.04	
Opt IV	P-1	1.38	-145.14	7.65;14.00;18.28	102.56; 99.32; 91.52	0.37(*14*)‡
Expt V	P-1	1.44		7.65; 9.02; 9.42	107.40; 92.06;101.66	
Opt V_*a*	P-1	1.38	-139.54	7.67; 9.19; 9.61	107.56; 93.99;100.87	0.21
Opt V_*b*	P-1	1.37	-141.77	6.78;10.77; 8.97	92.75; 85.03;103.48	0.89(*6*)
Opt V_*c*	P-1	1.39	-144.37	7.68; 9.28; 9.49	106.95; 92.51;102.49	0.19

‡The Expt and Opt IV crystal structures overlaid 14 out of 15 molecules.

It is clear from Table 4.3 that the lattice energy minimization of the ordered structures V_*a* and V_*c* produces a fairly small change in the cell constants, but the nearest minimum to V_*b* is rather different, probably because the initial structure contained a close Cl···Cl contact. These ordered structures correspond to the most stable of three models of form V. All three structures have a lattice energy that is 6-8 kJ mol^{-1} above form I, and so these ordered models of form V are not competitive in energy with the other known polymorphs of TA.

4.3.5. Energy Landscapes of Intermediate Steps

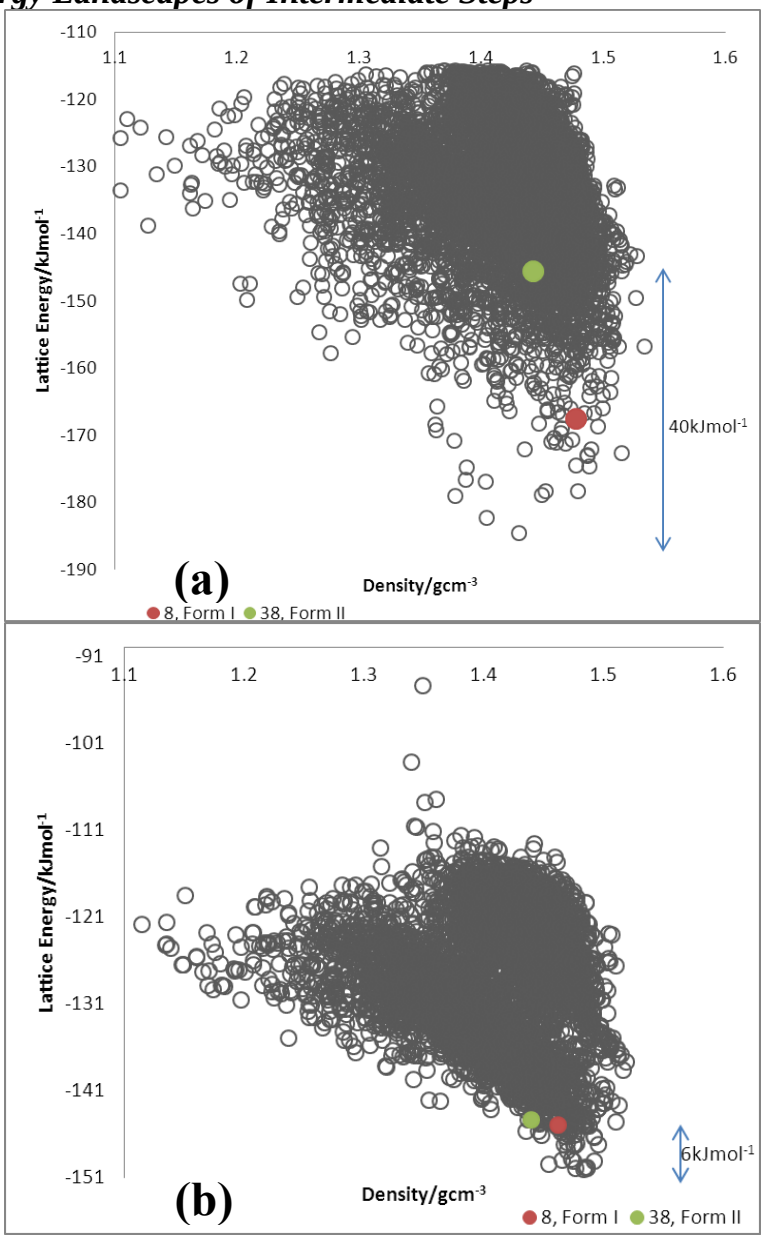


Figure 4.13: Crystal energy landscape of TA for the Z′=1 structure after (a) *CrystalPredictor* (**Step 1**) (b) the single conformer calculation of ΔE_{intra} and the atomic multipoles at the PBE0/6-31G(d,p) level of theory (**Step 2** of section 4.2.3).

I refined the lattice energy calculation in stages and systematically reduced the number of crystal structures as my confidence in the energy ranking increased. After conducting single-point calculations to get accurate ΔE_{intra} and distributed multipoles (which improved the U_{elec} and thus the U_{inter}, see equation 4.1) on all of the unique crystal structures (**step 2** of section 4.2.3), the gap between the known forms (I and II of TA) and the global minima reduced from 40 kJ mol^{-1} to 6 kJ mol^{-1} (Figure 4.13). This intermediate step was vital in reducing the computing time of the subsequent **step 3** (section 4.2.3, only about 100 crystal structures were refined) due to the increased confidence in ranking of the known forms (Figure 4.13).

4.3.6. Finding and Explaining the Known Structures with CSP

The known forms of FA and TA were at, or close to, the global minimum on the final crystal energy landscapes (Figure 4.14), which generated all the known $Z'=1$ ordered polymorphs. The crystal energy landscapes of FA and TA were similar in that they each had an energy gap of approximately 2 kJ mol^{-1} between the global minimum cluster and the other structures, although the cluster of low-energy structures was far larger for TA. Although this difference suggested that TA would be more polymorphic than FA, I compared the structures to see whether it was plausible that TA and FA could crystallize as distinct polymorphs, allowing for thermal motion at crystallization temperatures.

All the crystal structures of FA and TA whose energies were plotted in Figure 4.14 (and tabulated in Table 4.4 and Table 4.5) contain the carboxylic acid dimer (Figure 4.4a). However, the orientation of the phenyl substituents could differ so markedly that this dimer did not always satisfy the distance criterion to be overlaid by the default Crystal Packing Similarity comparison. This observation suggests that these structures on the final energy landscape for FA and TA differ by the packing of the aromatic rings. It was necessary to perform a qualitative assessment of the barriers that prevented the rearrangement of the molecules into a more stable form.

The lowest-energy hypothetical structure on the crystal energy landscape for FA was a $Z'=1$ structure (#1FA_22) that was very similar in density and lattice energy to the only known experimental $Z'=2$ form (Figure 4.14a), with a PXRD similarity of 0.85. This #1FA_22 hypothetical crystal structure overlaid 5 molecules (Figure 4.15a) with the experimental form, and it matched the conformation of one of the independent molecules (Figure 4.15b). The difference between these two structures lay in the packing of the layers, a difference that is seen in many polymorphs, such as progesterone[213].

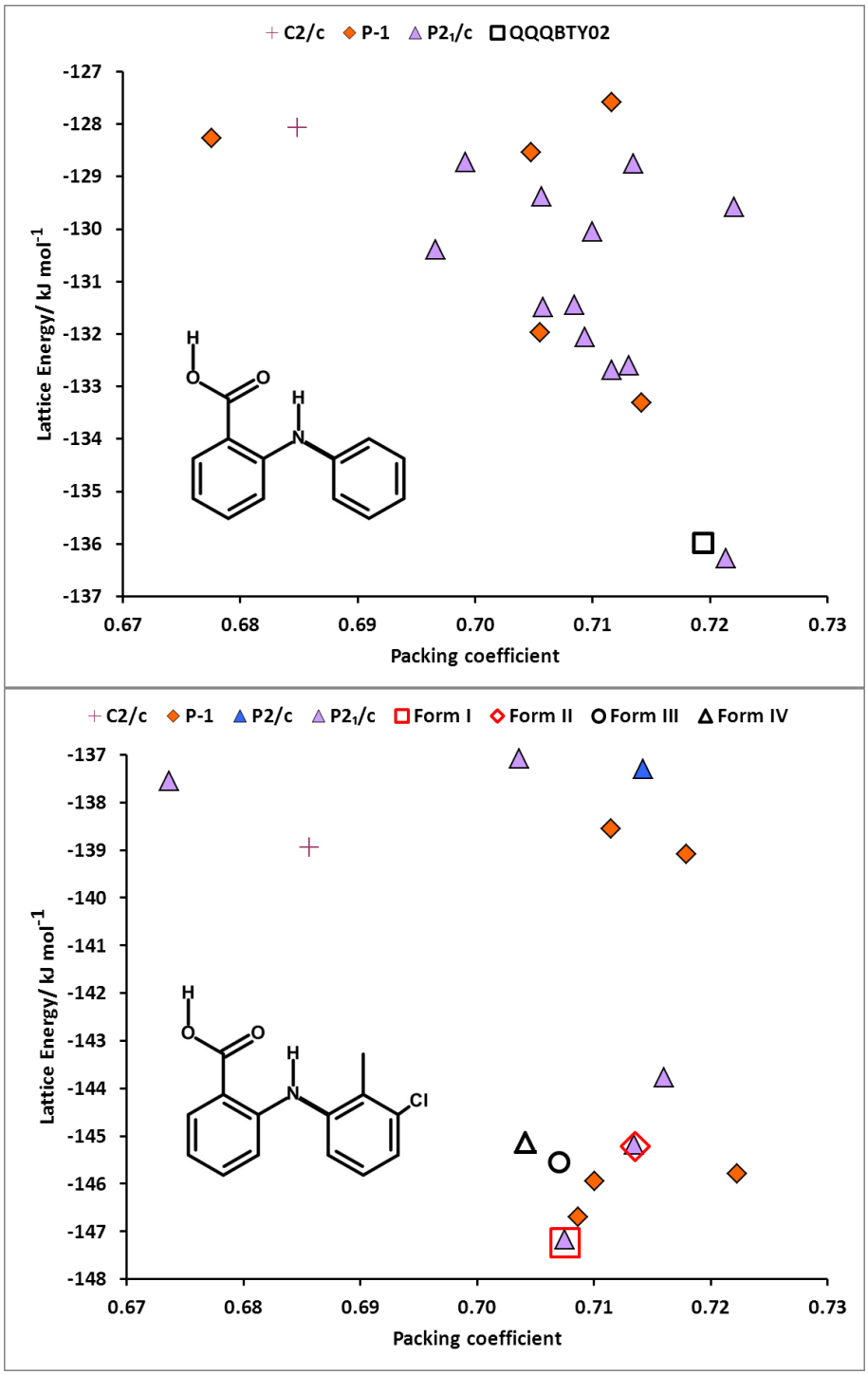


Figure 4.14: The crystal energy landscapes of FA (top) and TA (bottom). Each symbol represents a crystal structure of the specified space group, which is a minimum in the lattice energy (calculated within the polarizable continuum). The open symbols correspond to the minima, starting from the experimental structures (Table 4.3). The packing coefficient is the proportion of the cell volume occupied by the molecule[214], calculated using a grid spacing of 0.1 Å.

Table 4.4: The lattice energy minima of FA shown in Figure 4.14a. The minimum found starting from the experimental structure is in bold.

Label	Space Group	a b c / Å	α β γ / °	ρ / g cm^{-3}	U_{inter} ΔE_{intra} E_{latt} / kJ mol^{-1}
#1FA_22	*P2$_1$/c*	8.68; 4.68;27.09	90.00; 74.04; 90.00	1.34	-137.65; 1.37;-136.28
QQQBTY02	***P-1***	**8.48; 9.91;13.33**	**90.94; 88.16; 71.30**	**1.34**	**-139.20; 3.16;-135.99**
#2FA_2	*P-1*	4.54; 9.27;14.93	69.10; 92.01; 68.58	1.32	-134.83; 1.52;-133.31
#3FA_59	*P2$_1$/c*	4.57; 8.80;26.95	90.00; 99.56; 90.00	1.32	-133.81; 1.13;-132.69
#4FA_1	*P2$_1$/c*	4.51; 8.99;26.82	90.00; 78.77; 90.00	1.33	-134.35; 1.75;-132.60
#5FA_10	*P2$_1$/c*	4.58;26.86; 8.80	90.00; 85.26; 90.00	1.31	-133.50; 1.44;-132.06
#6FA_151	*P-1*	12.85; 7.44; 6.66	66.29;102.74;111.27	1.31	-134.18; 2.21;-131.97
#7FA_4	*P2$_1$/c*	17.88; 5.34;12.83	90.00; 61.36; 90.00	1.32	-145.12;13.63;-131.49
#8FA_16	*P2$_1$/n*	13.95; 5.20;15.08	90.00; 78.09; 90.00	1.32	-145.05;13.61;-131.44
#9FA_287	*P2$_1$/n*	14.29; 4.59;17.70	90.00; 70.60; 90.00	1.29	-132.79; 2.40;-130.39
#10FA_192	*P2$_1$/c*	4.60;19.18;12.31	90.00; 97.92; 90.00	1.32	-136.41; 6.36;-130.05
#11FA_19	*P2$_1$/c*	5.01;21.79;10.04	90.00; 73.77; 90.00	1.35	-143.77;14.19;-129.58
#12FA_23	*P2$_1$/c*	6.44;34.25; 6.53	90.00;131.52; 90.00	1.31	-138.55; 9.17;-129.38
#13FA_2880	*P2$_1$/n*	14.90; 4.93;14.56	90.00; 88.25; 90.00	1.33	-136.66; 7.91;-128.75
#14FA_33	*P2$_1$/c*	4.30;26.00; 9.97	90.00;102.50; 90.00	1.30	-130.39; 1.66;-128.73
#15FA_18	*P-1*	8.70; 7.99; 9.35	105.83;102.20;113.01	1.32	-130.40; 1.87;-128.53
#16FA_5226	*P-1*	10.21;13.25; 9.90	141.09;116.96; 43.47	1.27	-130.13; 1.86;-128.27
#17FA_484	*C2/c*	38.64; 8.97; 7.50	90.00; 58.74; 90.00	1.28	-129.62; 1.55;-128.07
#18FA_7	*P-1*	4.99;11.13; 9.95	84.46;104.67; 92.66	1.33	-141.67;14.09;-127.58

Table 4.5: The lattice energy minima of TA shown in Figure 4.14b. Minima found starting from the experimental structures, or ordered models, are in bold.

Label	Space Group	a b c / Å	α β γ / °	ρ / g cm^{-3}	U_{inter} ΔE_{intra} E_{latt} / kJ mol^{-1}
FORM I	***P2$_1$/c***	**4.86;31.54; 8.32**	**90.00;102.00; 90.00**	**1.39**	**-148.74;1.50;-147.24**
#1TA_8	*P2$_1$/c*	4.85;31.60; 8.33	90.00; 77.87; 90.00	1.39	-148.65;1.48;-147.17
#2TA_15	*P-1*	17.72; 8.47; 4.84	79.49;112.37;111.71	1.40	-148.32;1.63;-146.69
#3TA_876	*P-1*	16.98; 7.44; 7.34	94.40; 93.76; 41.81	1.41	-149.52;3.57;-145.95
#4TA_6243	*P-1*	12.06;10.15; 8.71	71.76; 61.70;117.37	1.42	-147.74;1.96;-145.79
FORM III	***P2$_1$/c***	**7.83;11.64;27.48**	**90.00; 93.32; 90.00**	**1.39**	**-148.53;2.98;-145.55**
FORM II	***P2$_1$/n***	**3.86;22.06;14.60**	**90.00; 96.21; 90.00**	**1.41**	**-147.84;2.63;-145.21**
#5TA_38	*P2$_1$/n*	3.86;22.05;14.60	90.00; 83.76; 90.00	1.41	-147.81;2.63;-145.18
FORM IV	***P-1***	**7.65;14.00;18.28**	**102.56; 99.32; 91.52**	**1.38**	**-147.73;2.59;-145.14**
#6TA_82	*P2$_1$/c*	3.90;14.33;23.74	90.00; 68.05; 90.00	1.41	-146.11;2.34;-143.77
FORM V_*b*	***P-1***	**6.78;10.77; 8.97**	**92.75; 85.03;103.48**	**1.37**	**-145.05;3.28;-141.77**
FORM V_*c*	***P-1***	**7.68; 9.28; 9.49**	**106.95; 92.51;102.49**	**1.39**	**-144.37;3.13;-141.24**
FORM V_*a*	***P-1***	**7.67; 9.19; 9.61**	**107.56; 93.99;100.87**	**1.38**	**-142.19;2.65;-139.54**
#7TA_45	*P-1*	14.44;12.31; 3.85	109.50; 83.30; 76.31	1.42	-141.71;2.63;-139.08
#8TA_88	*C2/c*	26.97; 4.79;20.67	90.00;105.67; 90.00	1.35	-141.13;2.18;-138.95
#9TA_64	*P-1*	7.56; 9.00; 9.86	75.31;101.45; 80.29	1.41	-147.44;8.89;-138.55
#10TA_153	*P2$_1$/c*	4.92;12.39;22.17	90.00; 75.74; 90.00	1.33	-139.51;1.96;-137.55
#11TA_80	*P2/n*	14.92; 3.98;21.50	90.00; 75.36; 90.00	1.41	-139.10;1.80;-137.30
#12TA_3630	*P2$_1$/c*	16.55; 3.84;23.38	90.00; 57.14; 90.00	1.39	-139.30;2.22;-137.08

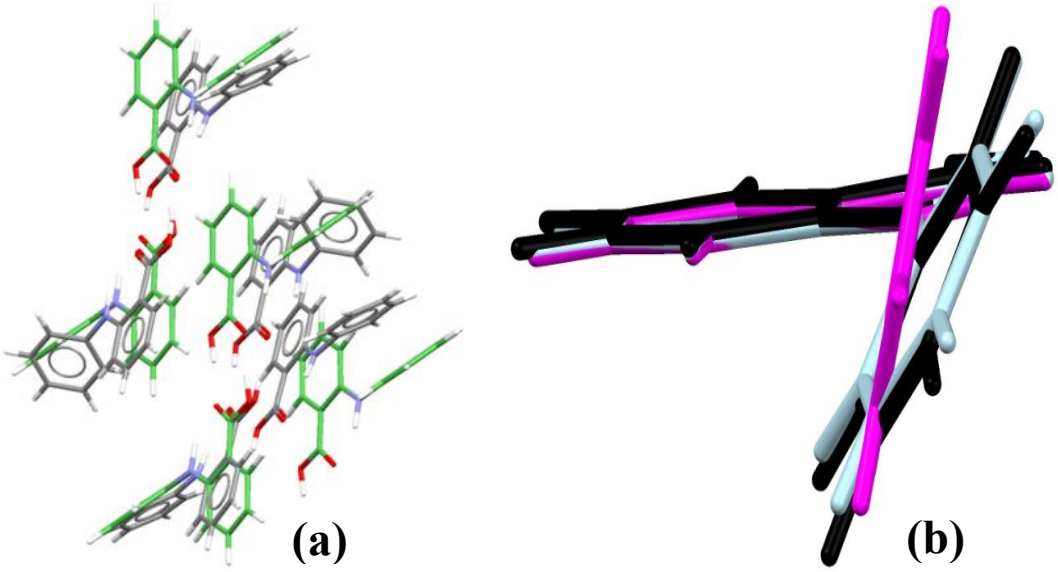


Figure 4.15: (a) Overlay of the only known form of FA (grey) and the most stable structure on the Z′=1 crystal energy landscape (Figure 4.14a) #1FA_22 (green) with an $RMSD_5$ value of 1.75Å. (b) Overlay of the conformers of form I (pink and light blue) and #1FA_22 (black).

A significant energy gap was observed between the experimental and lowest energy Z′=1 structure and the other structures generated for FA (Figure 4.14a). Remarkably, this gap was between two hypothetical structures that had a 12-molecule overlay and only differed in the packing of the layers (Figure 4.14a). The lower-energy structure (#1FA_22) had more close contacts and a higher packing coefficient than the higher-energy structure. The second most stable hypothetical structure (#2FA_2) was also closely related to the experimental structure; in it, half of the molecules had changed the phenyl torsion angle to a less stable conformation, changing the π–π stacking with the phenyl ring in the neighbouring molecule to a herringbone[215] (C-H···π) interaction (Figure 4.16b,c). Thus, there were two ways that the #2FA_2 structure could gain 2 kJ mol^{-1} of stabilisation energy (Figure 4.16). The relationship between the two most stable structures (#1FA_22 and the experimental structure) means that it may be very difficult to produce the Z′=1 structure as a polymorph. Clearly, Z′=1 and Z′=2 structures are very close in energy and are related (via #2FA_2). It is therefore difficult to establish whether the observed form is in fact a "crystal on the way"[216] that is trapped by a barrier from rearranging into the Z′=1 form, or whether the observed form is the more stable structure[217].

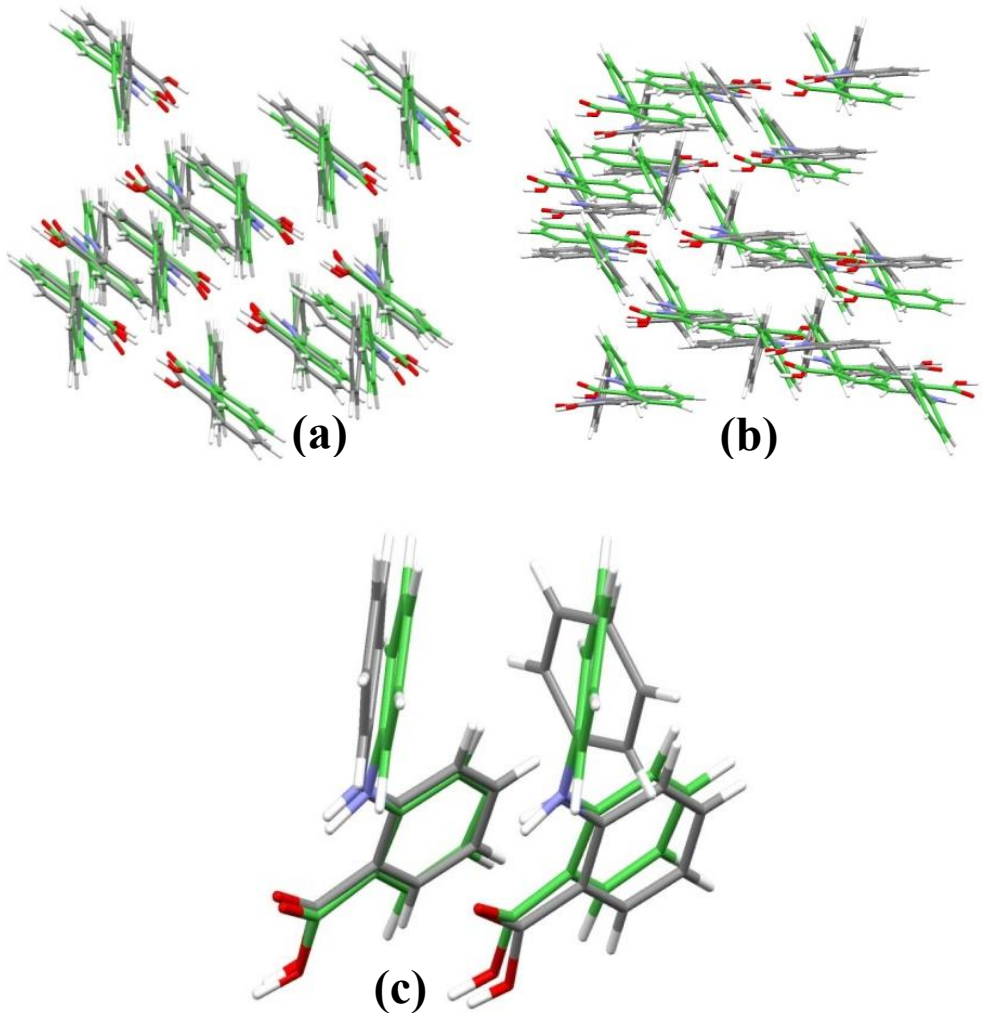


Figure 4.16 (a) Overlay of the two most stable hypothetical crystal structures on the crystal energy landscape; global minimum #1FA_22 (grey), and #2FA_2 (green) with ΔE_{latt}=2.0 kJ mol^{-1}. The crystal structures overlaid 12 molecules with $RMSD_{12}$ of 0.51 Å, and the conformations had an $RMSD_1$ of 0.09 Å. (b) Overlay of #2FA_2 (green) and the experimental crystal structure of FA (grey) with a $RMSD_{15}$ of 1.23 Å, obtained with a distance and angle tolerance of 50% and 50°. (c) Overlay of the two independent molecules in the experimental structure (grey) with #2FA_2 (green).

For TA, the crystal energy landscape successfully found the Z′=1 polymorphs, forms II and I, to be the first and fifth most stable crystal structures, within 2 kJ mol^{-1} of the global minimum (Figure 4.14b); the landscape also provided a good reproduction of the experimental crystal structures (Figure 4.17 and Table 4.3). Four other structures were observed, all of which were competitive in lattice energy with the known metastable forms. These crystal structures had somewhat similar packing to the known forms of TA (Table 4.8), some being more similar to the form V Z=2 model structures constructed from the disorder components of TA than to the other polymorphs (contrast Table 4.2 and Table 4.8). Some of the similarities (Table 4.6) between the hypothetical structure and the other fenamates in Table 4.1 were even more striking. One of the computed structures

#6TA_82 (highlighted in Table 4.5) matched 15 molecules of SURMOI (Table 4.1), suggesting that crystals of this dimethyl-substituted molecule might template the nucleation of a novel polymorph of TA. This observation led to the joint experimental and computational project discussed in section 4.3.8.

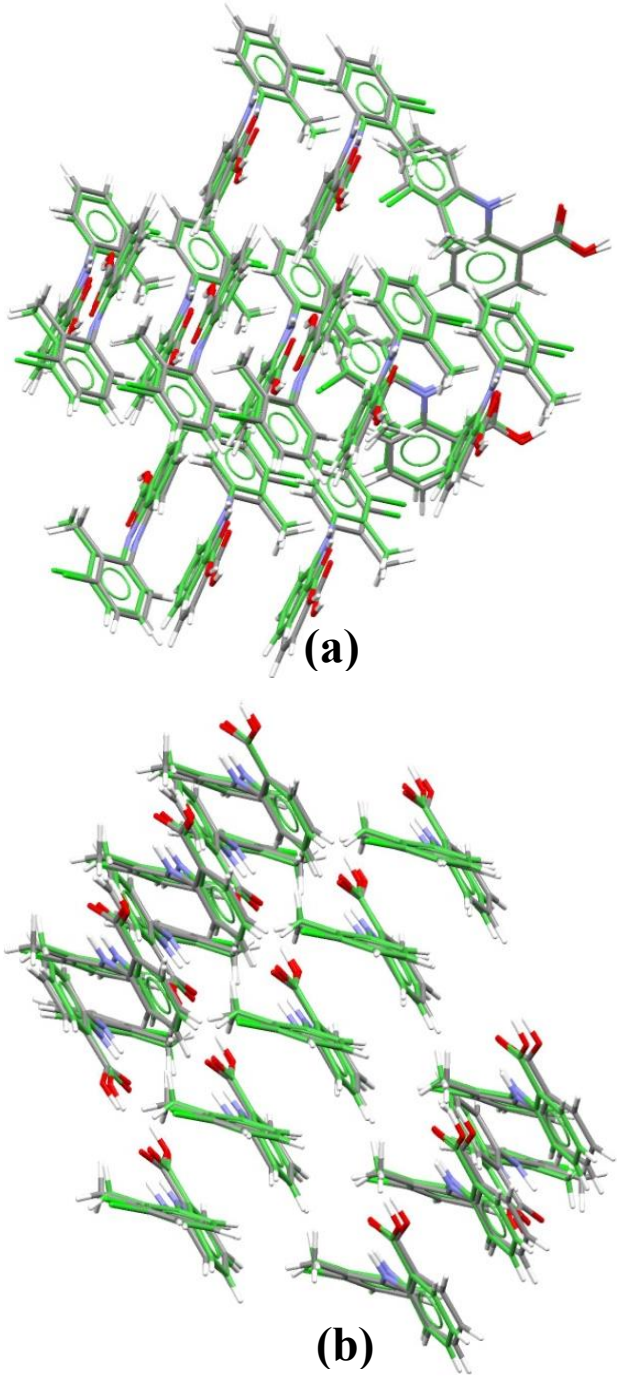


Figure 4.17: The overlay of the experimental Z′=1 crystal structures (grey) of TA with the closest structure found in the search (green) (a) Form I and #1TA_8 overlaid with an $RMSD_{15}$ value of 0.32Å (b) Form II and #5TA_38 overlaid with an $RMSD_{15}$ value of 0.26Å.

4.3.7. Other Known Fenamates and their Hypothetical Structures

Further marked similarities were observed between the structures on the crystal energy landscapes of FA and TA (Figure 4.18). The hypothetical crystal structures on the crystal energy landscape of FA in Figure 4.18 were similar to some of the known forms of TA (forms I-IV), with some of these hypothetical crystal structures overlaying up to 13 molecules (forms I and III of TA; Figure 4.18). Many of the low-energy structures of FA had more in common with a polymorph of TA than with the only known form of FA; that is, with a few exceptions, the second numbers are greater than the first bold numbers (Figure 4.18). There exist distinct crystal structures of FA with the planar molecules (highlighted in Figure 4.18).

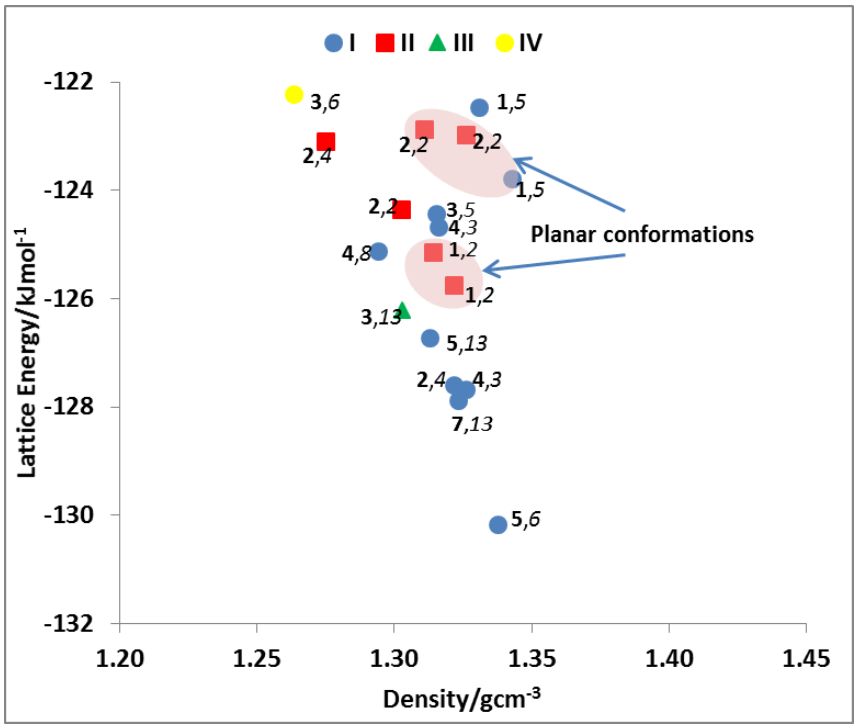


Figure 4.18: Classification of the crystal energy landscape of FA in terms of the similarity of the structures to those of known TA polymorphs (I-IV) *without* PCM (hence the different y-axis compared to Figure 4.14a). Numbers in bold indicate the number of molecules that overlay with the known structure of FA. Hence, a label **5**, *6* on a structure represented by a blue circle means that the structure overlaid 5 molecules with the experimental form of FA but 6 with form I of TA. Each point is coloured by the TA polymorph with which it overlaid the greatest number *n* of molecules, in a Crystal Packing Similarity calculation that ignored the CH_3 and Cl.

The similarities between the known forms of TA and the hypothetically generated structures of FA confirmed that, because of the effect of TA's substituents (CH_3 and Cl) on the molecules' relative energies, the energy landscape of TA was more polymorphic than that of FA (Figure 4.14). The common clusters produced by the similarity overlays between different fenamates usually had the differing substituents in the exterior region

where there was no overlay of the coordinating molecule, consistent with the substituents causing the differences in the packing. This in marked contrast to the large common cluster between forms III and IV of TA (Figure 4.4c), where the packing of the aromatic rings is similar.

Table 4.6: Crystal Packing Similarity comparisons of the experimental structures of FA and TA and their most stable hypothetical crystal structures with the crystal structures of the other fenamates shown in Figure 1.3. The similarity is given in the form n(RMSD$_n$), with $n > 4$ molecules highlighted.

	n(RMSD$_n$/Å)							
	FPAM-CA	FPAM-CA11	LAH-LOW	PEF-MET	PEF-NAQ	SUR-MEY	SUR-MOI	XYA-NAC
FA								
QQQBTY02	2(0.71)	**4(1.33)**	**7(0.84)**	3(1.28)	**6(0.62)**	3(0.5)	3(1.2)	**4(0.49)**
#1FA_22	2(0.41)	**5(1.91)**	**6(1.31)**	2(0.13)	**11(0.96)**	**4(0.98)**	**4(1.08)**	3(1.8)
#2FA_2	2(0.43)	3(0.56)	**7(1.01)**	2(0.1)	**11(0.51)**	3(1.07)	**7(0.98)**	3(0.86)
TA								
II	1(0.89)	2(0.15)	2(0.59)	2(0.11)	1(0.61)	2(0.12)	**12(1.87)**	2(0.29)
I	2(0.59)	2(0.48)	3(0.61)	1(0.7)	**13(0.38)**	**6(1.78)**	**4(1.2)**	2(0.62)
III	1(0.81)	**7(0.75)**	2(0.48)	**13(0.33)**	2(0.55)	**13(0.61)**	3(0.41)	**13(0.45)**
IV	2(0.87)	**6(0.71)**	3(1.09)	**11(0.49)**	2(0.46)	**13(0.84)**	3(0.56)	**15(0.3)**
V_a	2(0.57)	2(0.46)	**5(0.42)**	1(0.68)	3(0.39)	2(0.22)	2(0.18)	2(0.57)
V_b	2(0.75)	3(0.32)	**7(0.64)**	2(0.47)	**4(1.10)**	**5(0.41)**	**4(1.19)**	**5(0.72)**
V_c	2(0.95)	2(0.84)	**5(0.41)**	1(0.42)	**4(1.10)**	**4(0.51)**	2(0.29)	3(0.46)
#2TA_15	2(0.52)	2(0.53)	3(0.91)	1(0.77)	**14(0.4)**	**4(0.82)**	**4(1.22)**	2(0.7)
#3TA_876	1(0.93)	**4(1.56)**	**7(1.42)**	3(0.43)	1(0.66)	3(0.71)	**4(0.47)**	2(0.37)
#4TA_6243	2(0.54)	3(1.15)	3(1.76)	1(0.69)	**4(0.16)**	**4(1.49)**	2(0.17)	3(0.91)
#6TA_82	1(0.92)	2(0.17)	2(0.61)	2(0.12)	2(0.64)	2(0.08)	**15(0.28)**	2(0.28)

Some fenamate crystal structures are specific to the molecule. One observed example (Table 4.6) is the FPAMCA polymorph of flufenamic acid. In 2012, through the use of polymer-induced heteronucleation and solid–solid transformations, flufenamic acid set a new record[18] for a polymorphic compound with eight solved structures, which was recently equalled by aripiprazole[218]. These fenamate structures are all based on the carboxylic acid dimer, with some having more similarity to the TA and FA low-energy structures (Table 4.7). The specific substituents affect the range of packing of the aromatic rings. For example, the low-energy unobserved structures of FA, which are based on a planar conformation of the entire molecule (Figure 4.14), are probably also specific to FA, because the lowest-energy crystal structure of TA with a planar molecule is 14 kJ mol^{-1} above the global minimum structure. The energy difference between crystal structures of TA is caused by substituent intermolecular interactions that destabilize the

structures containing planar molecules; the intramolecular energy penalty, ΔE_{intra}, for this planar conformation is similar for FA and TA (~6 kJ mol^{-1}, Figure 4.12).

Table 4.7: An extension of Table 4.6 showing Crystal Packing Similarity comparisons of the latest structures of flufenamic acid[18] and the experimental structures of FA and TA and their most stable hypothetical crystal structures.

	***n*(RMSD$_n$/Å)**					
	FPAMCA_*z*					
	z=2	*z*=4	*z*=5	*z*=6	*z*=7	*z*=8
FA						
QQQBTY02	**5(1.49)**	3(0.79)	2(0.73)	**4(1.60)**	2(0.43)	3(0.82)
#1FA_22	**4(1.62)**	**5(1.42)**	2(1.05)	**5(2.02)**	2(0.22)	2(0.26)
#2FA_2	**4(1.51)**	**4(0.34)**	3(1.19)	**5(0.80)**	2(0.21)	3(1.30)
TA						
II	2(0.10)	2(0.16)	2(0.4)	2(0.17)	1(0.71)	2(0.81)
I	3(1.58)	2(0.54)	2(0.69)	4(2.10)	2(0.34)	2(0.41)
III	2(0.22)	2(0.24)	**4(0.52)**	2(0.25)	1(0.64)	2(0.77)
IV	2(0.20)	2(0.23)	**4(1.68)**	3(0.84)	2(0.64)	2(0.70)
V_*a*	3(1.24)	2(0.48)	2(0.64)	2(0.50)	3(0.68)	3(0.80)
V_*b*	**5(0.54)**	3(0.67)	2(0.45)	3(0.63)	1(0.59)	2(0.78)
V_*c*	2(0.85)	2(0.41)	2(0.79)	2(0.42)	2(0.77)	3(1.27)
#2TA_15	**4(3.04)**	2(0.80)	2(0.17)	2(0.62)	3(0.42)	**4(2.80)**
#3TA_876	**4(1.34)**	2(0.18)	**4(2.10)**	2(0.19)	1(0.76)	2(0.82)
#4TA_6243	**4(1.10)**	2(0.58)	2(0.40)	2(0.60)	2(0.31)	2(0.38)
#6TA_82	2(0.18)	2(0.19)	2(0.25)	2(0.20)	1(0.73)	2(0.81)

Table 4.8: Crystal Packing Similarity calculations of the known forms of TA and unobserved, thermodynamically competitive crystal structures. The similarity between the structures is given in the form *n*(RMSD$_n$) where *n* is the number of molecules overlaid, and the value in brackets is the RMSD$_n$ overlay. Structures with the highest similarity are highlighted.

	n(RMSD$_n$/Å)			
	#2TA_15	#3TA_876	#4TA_6243	#6TA_82
KAXXAI	1(0.95)	**4(0.33)**	1(0.88)	**9(0.43)**
KAXXAI01	1(0.49)	1(0.96)	**5(1.32)**	1(0.90)
KAXXAI02	2(0.87)	3(1.08)	2(0.95)	3(0.59)
KAXXAI03	2(0.83)	3(0.81)	3(2.55)	3(0.81)
KAXXAI04_a	**4(1.25)**	1(0.92)	3(0.16)	1(0.86)
KAXXAI04_b	**4(1.28)**	3(0.92)	**6(1.07)**	2(0.69)
KAXXAI04_c	1(0.13)	2(1.20)	**4(1.05)**	1(0.59)

4.3.8. Update on a Recently Discovered Polymorph of TA

Section 4.3.6 theorized that experimentalists could nucleate a novel polymorph of TA by using the SURMOI molecule as a template. However, my colleague at UCL used mefenamic acid (Figure 1.3) because SURMOI is very expensive. A novel polymorph of TA was nucleated by solution evaporation after dissolving TA (form I) in ethanol and

seeding it with mefenamic acid (form I)[219]. This procedure resulted in the nucleation of form VI of TA, which was disordered (65% and 35%) with same unit cells as form I of mefenamic acid.

Although the major component of this novel form VI was generated in **step 1**, #137_TA, the form was not among the structures on the final energy landscape of Figure 4.14b. The major component of form VI and #137_TA overlaid 15 molecules with a $RMSD_{15}$ value of 0.30 Å (Figure 4.19). This structure #137_TA had a final lattice energy of -144.97 kJ mol^{-1}, which was competitive with form IV of TA (Table 4.5, -145.14 kJ mol^{-1}). It should be noted that **steps 3** and **4** of CSP (section 4.2.3) involved just over 100 crystal structures. Clearly, lattice energy refinements (**steps 3** and **4**, section 4.2.3) of more hypothetical crystal structures are vital.

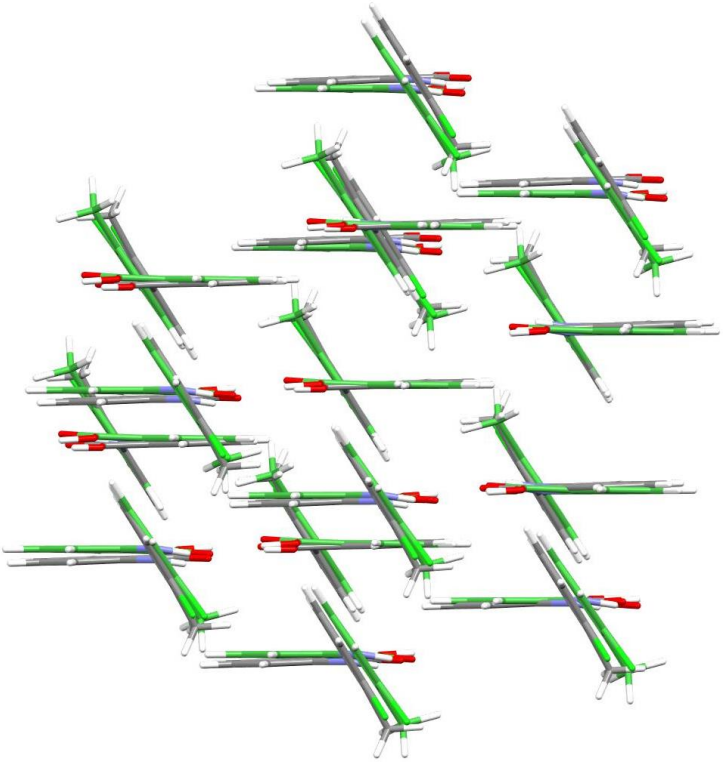


Figure 4.19: Overlay of #FA_137 (green) and the major component of experimental crystal structure of TA form VI (grey) with a $RMSD_{15}$ of 0.30 Å, obtained with a distance and angle tolerance of 50% and 50°.

4.3.9. The Cost of Lattice Energy Minimisation

The lattice energy minimization (*CrystalOptimizer*; **step 3**) of one crystal structure accounted for as much as 75% of the computing time and took as long as 4 days (Figure 4.20a) because multiple *ab initio* calculations were required to balance U_{inter} and ΔE_{intra}. The computing time was approximately half as long for the optimisation of the last structure (form VI of TA, Figure 4.20b). This last optimisation (Figure 4.20b) was faster

because there were about hundred structures stored in the Local Approximation Models (LAMs)[71] database. (LAMs[71] database stores and reuses conformational energy and multipole representations of the charge density.) This final calculation was performed after the discovery of a novel polymorph of TA, as discussed in section 4.3.8. Even with the help of LAMs[71] database, the computational cost was still very high.

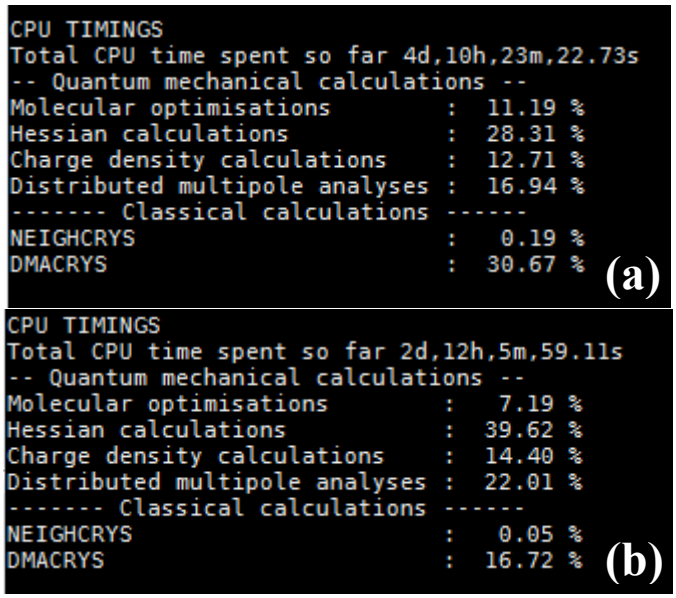


Figure 4.20: CPU timings of (a) #38_TA (form II) and (b) #137_TA (form VI, section 4.3.8) after lattice energy minimisation using *CrystalOptimizer* (**step 3**).

4.4. Summary

This chapter described the successful use of the CSP method (see Chapter 3) to examine the computer-generated low-energy crystal structures of the unsubstituted structural fragment, FA, and a derivative, TA. This chapter showed that all structures were based on the same $R_2^2(8)$ H-bonded dimer, and that the structures differed only in the packing of the phenyl rings. The crystal energy landscape of TA had several crystal structures, including the observed polymorphs, which were tightly clustered around the global minimum, whereas the corresponding cluster for FA contained only the observed and a closely related structure. This observation is supported by past experiments.

Many of the crystal structures on the energy landscape for FA overlaid a significant portion of the coordination environment with other observed or calculated structures. Therefore, the phenyl ring substituents determined the relative energies of the crystal structures and hence whether a specific fenamate would be very polymorphic or not. The FA fragment can be described as a *polymorphophore*, despite only two structures clustered around the global minimum, since the molecular fragment packed in a wide range of distinct crystal structures with different conformations that were energetically favourable. This finding indicates that being a *polymorphophore* correlates with but does

not guarantee polymorphism. It is perhaps not surprising that this term has not been more widely accepted.

The similarities between the structures on the final energy landscapes of TA and other fenamates led my colleague to conduct a templating experiment on TA (section 4.3.8). The subsequent discovery of a new polymorph of TA (form VI, section 4.3.8) exemplified one problem with the CSP discussed earlier. Restricting the number of crystal structures that were accurately optimised (**step 3**, *CrystalOptimizer*) to just over 100 structures meant excluding the recently discovered crystal structure of TA (section 4.3.8) from the final crystal energy landscape (Figure 4.14b). As a result, current intermediate steps like the single-point calculations used in this chapter (**step 2**), although successful (section 4.3.5), had their limitations.

The findings of Chapters 3 and 4 indicate the need for a cheaper and more accurate approach to modelling lattice energy. If the evaluation of ΔE_{intra} (the bulk of computer time, section 4.3.9) can be realized through a less expensive method (Chapter 6), then molecular dynamics calculations and the temperature effects (e.g. free energies) can be better accounted for. However, we must first understand the origin of intramolecular forces and energies (Chapter 5).

Chapter 5. The Origin of conformational energy barriers in Fenamates

"…'*Begin at the beginning,***' *the King said, gravely, '**and go on till you come to the end; then stop***'…"**

— Lewis Carroll, *Alice in Wonderland*

5.1. Introduction

5.1.1. Background

Chapters 3 and 4 established that lattice energy, E_{latt}, could be partitioned into intermolecular, U_{inter}, and intramolecular components, ΔE_{intra}:

$$E_{latt} = U_{inter} + \Delta E_{intra} \qquad 5.1$$

where U_{inter} is the sum of the dominant electrostatics, U_{elec}, plus the contributions of repulsion-dispersion, $U_{repul\text{-}disp}$, and ΔE_{intra} is the energy required to distort the molecule from its gas phase conformation[71]. The conformation determines the overall molecular shape, so different conformers may lead the crystal to arrange itself in different ways, ultimately influencing the properties of the crystal[220] (Chapter 1).

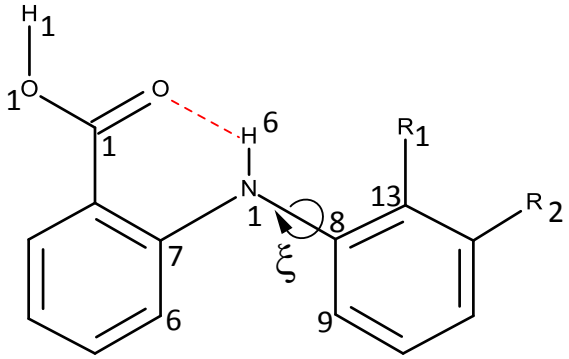


Figure 5.1: The fenamate family, showing the low barrier torsion angle (ξ=C_7-N_1-C_8-C_9). ξ=0 when the aromatic rings are coplanar as drawn. The fenamates family mentioned in this chapter includes fenamic acid (FA) R_1= R_2= H, tolfenamic acid (TA) R_1= CH_3, R_2= Cl, mefenamic acid (MA) R_1= R_2= CH_3, clofenamic acid (ClFA) R_1=H, R_2=Cl, and flufenamic aid (FFA) R_1=H, R_2=CF_3. The dotted line represents an intramolecular hydrogen bond.

This chapter will investigate the conformational flexibility of fenamate molecules in isolation. Through this investigation, we will come to a better understanding of the origin of torsional barriers, which will enable the derivation of an analytical model for the intramolecular energies of the fenamate molecules FA and TA (Chapter 6). To put this problem in context, there is still an ongoing debate about the origin of the torsional barrier of the ethane molecule, which has only 8 atoms, as compared to the 27 and 30 atoms of FA and TA, respectively.

5.1.2. Torsional Barrier for Ethane

The torsional barrier is defined as the difference in energy between the least stable and the most stable conformations in rotation around a bond. In order for one full rotation around a bond to occur, this energy barrier must be overcome (Figure 5.2). The energy barrier for fenamic acid, FA, is smaller (Chapter 4) than ethane (Figure 5.3).

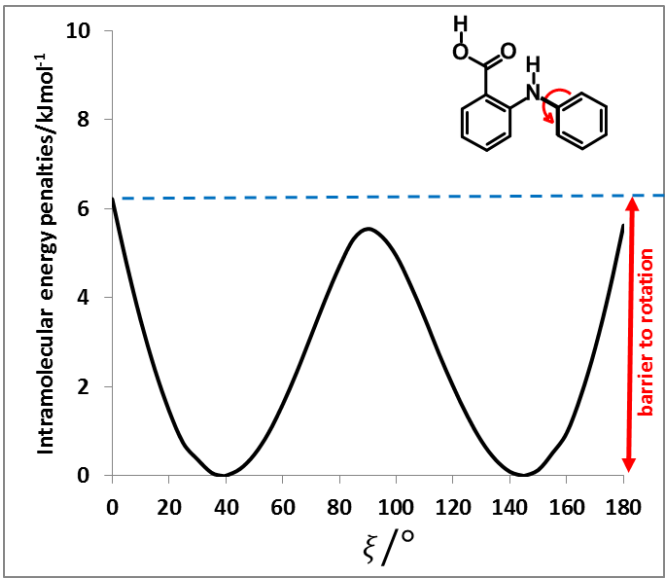


Figure 5.2: Illustration of the barrier to rotation with the relaxed conformational energy profile of FA.

Previous studies of torsional barriers have focussed on small molecules such as ethane. The torsional barrier for the ethane molecule is one of the most fundamental puzzles in conformational analysis, and researchers have great interest in understanding its origin[221]. The experimental (12.03 to 12.25 kJ mol^{-1}) and quantum mechanical (11.24 to 13.03 kJ mol^{-1}) calculations of ethane's torsional barrier are in good agreement (Table 5.1).

However, researchers disagree upon the origin of the torsional barrier in ethane. The root of the controversy is whether the origin of the torsional barrier in ethane results from stronger *hyperconjugation* stabilization[222] in the staggered conformation (Figure 5.3) than the eclipsed form, or whether the barrier originates from greater steric repulsion[221] in the eclipsed conformation (Figure 5.3) due to *electrostatic interaction* and *Pauli exchange repulsion*. Three physical factors may play a role in ethane's structural preference: *hyperconjugation*, *electrostatic* interactions, and *exchange* interactions.

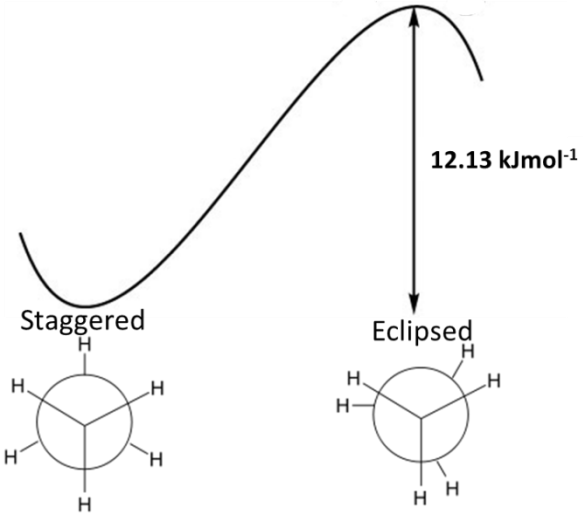


Figure 5.3: The staggered and eclipsed conformations of the ethane molecule. (Sources: Ref. [222] and en.wikipedia.org/wiki/Alkane_stereochemistry [Accessed 11/01/2015]).

Table 5.1: Torsional barrier height of ethane. (Source: Reproduced from Ref. [223]).

Method	E_{barr} (cm^{-1})	
Heat capacity	1006	**Experimental** 12.03-12.25 kJ mol^{-1}
IR	1024	
Microwave	1008	
Raman	1009	
Raman	1012	
MP2/6-31++G**	1089	**Quantum Theory** 11.24-13.03 kJ mol^{-1}
HF/cc-pVDZ	1130	
HF/cc-pVTZ	1066	
HF/cc-pVQZ	1067	
HF/CBS	1067	
CCSD(T)/cc-pVDZ	1017	
CCSD(T)/cc-pVTZ	971	
CCSD(T)cc-pVQZ	952	
CCSD(T)/CBS	941	
CCSD(T)/CBS + core correlation correction	942	
CCSD(T)/CBS (X^{-3})	946	
CCcf/CBS	940	
QCISD(T)/cc-pVTZ// MP2/6-311G(3*df*,3*pd*)	980	
CCSD(T)/cc-pVQZ + core correlation correction	957	
1VBSCF/6-31G(*d*)// CCSD(T)/6-31G(*d*)	945	

Hyperconjugation and *exchange* repulsion are quantum mechanical effects arising from orbital overlap. The *hyperconjugation* effect for the ethane molecule is constituted of the vicinal interactions between the occupied σ_{CH} bond orbitals of one of the methyl group and the virtual antibonding σ_{CH}* orbitals of the other methyl group[221]. The staggered

conformation (Figure 5.4a) is stabilised through the delocalization of electron charge density involving these orbitals, an attractive contribution.

Pophristic and Goodman[222] used natural bond orbitals (NBOs) to argue that *hyperconjugation*, or stabilization of the staggered ethane conformation, is the origin of the barrier to rotation in ethane. By showing that if the σ_{CH}-σ_{CH}* hyperconjugative interactions were removed, and by adding electrons to the unoccupied σ_{CH}* virtual orbital, the steric repulsion favoured the less stable eclipsed conformation (Figure 5.3). This model of *hyperconjugation* interactions relies on NBOs and is not so obvious if molecular orbitals (MOs) are used, let alone if correlated wave functions are used.

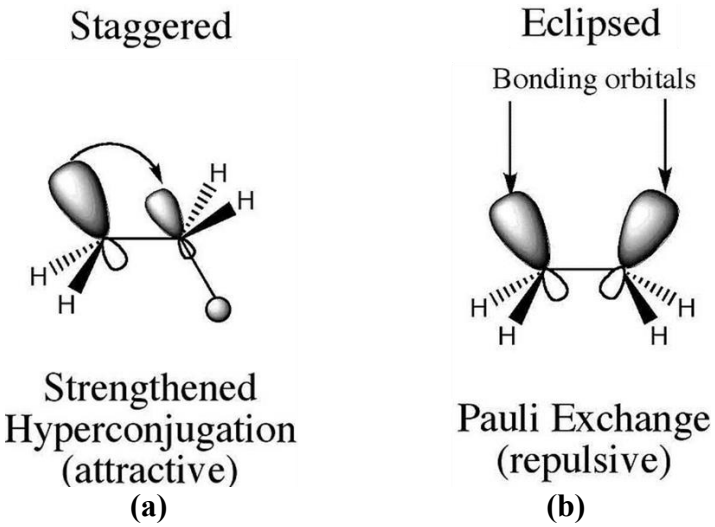


Figure 5.4: (a) *Hyperconjugation* effects on the staggered, (b) Pauli *exchange* repulsion in the eclipsed conformation. (Source: en.wikipedia.org/wiki/Hyperconjugation [Accessed 11/01/2015])

Pauli *exchange* repulsion, which requires that pairs of electrons do not occupy the same spatial region, occurs in the eclipsed conformation (Figure 5.4b). *Electrostatic* (Figure 5.5, also see section 2.2.3) repulsion involves the classical R^{-1} repulsion between charges. At short distances, the 1-4 (third nearest neighbour) interactions still have an electrostatic effect, but there are also considerable penetration-effects (section 2.2.5). The Pauli *exchange* and *electrostatic* repulsion, frequently regarded as the steric factor, destabilise the eclipsed conformation of ethane. These interactions are the same as intermolecular interactions, but at shorter distances.

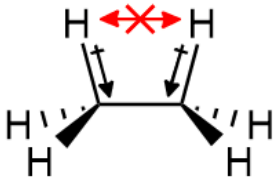


Figure 5.5: Electrostatic steric repulsion (between electrons on the bond, black arrow) for the eclipsed conformation of ethane. The red arrow indicates exclusion of van der Waals repulsion since the van der Waals radii of the hydrogen are too small to make contact. (Source: en.wikipedia.org/wiki/Conformational_isomerism [Accessed 11/01/2015])

Mo and Gao[221] made three arguments in favour of *steric* repulsion being the origin of the barrier to rotation. One argument was that *hyperconjugation* is a second-order effect, and hence should have little or no direct role in restricting free rotation[224]. Second, these researchers argued that overestimation of the *hyperconjugative* effects led others to misidentify *hyperconjugation* as the origin of the rotation barrier. Third, Mo and Gao argued that the experimental work on heats of hydrogenation for propene supported a specific energy decomposition scheme[225]. In short, the researchers argued that their three points supported the claim that steric repulsion is the origin of the torsional barrier.

The origin of the torsional barrier remains controversial[226], even for a small molecule such as ethane. The question remains: Is the torsion defined by the origin *hyperconjugation*[221,222] or by a similar electronic effect from changes in MOs on the atoms? Or is the barrier equivalent to intermolecular non-bonded repulsion and electrostatic effects[227] at short distances, i.e. as though there were no single covalent bond? The quantitative distinction can be highly dependent upon the precise definitions and upon the type of charge density calculation that is used.

5.1.3. Overview

The conformational energy barrier of FA (6.22 kJ mol^{-1}) is smaller than that of the ethane molecule (12.3 kJ mol^{-1}). The goal of this chapter is to investigate the qualitative origin of the conformational barrier to rotation for fenamate molecules in isolation. First, I will investigate the effects that different definitions of the same flexible bond have on the calculated conformational profile of TA. The step size of torsion angles, various *ab initio* methods, and relaxation effects will be investigated for various fenamates (Figure 5.1). Following this, I will investigate the electronic contributions using constrained model molecules that avoid pyramidization (Figure 5.7) with minimal or no steric effects. Finally, an *exp*-6 atom-atom model, which contains both a repulsive and a dispersive part, will be used to investigate the steric effect of FA and TA.

This chapter will show that the barrier to rotation for FA and TA (Figure 5.1) depends significantly upon small correlated changes in the entire molecule rather than on the atomic position of the atoms that define the rotating torsion. It will also show that there is an intrinsic dispersion and electronic energy barrier when the aromatic rings are perpendicular to one another; the barrier to rotation for small model molecules can be represented by a simple cos2ξ term.

5.2. Method

In section 4.3.1, it was established that the PBE0/6-31+G(d) level of theory is the best compromise between the cheap but inaccurate HF method and the expensive but accurate MP2 method for calculating the packing (intermolecular) and conformational (intermolecular) energies (equation 5.1) of FA and TA. This chapter examines the *ab initio* conformational profiles of fenamate molecules in more detail. All the *ab initio* calculations reported in this chapter were performed using GAUSSIAN03[53].

5.2.1. Determining Conformation Energy Scan for Isolated Molecule

For TA, there are 4 definitions of the N_1-C_8 rotatable bond (Figure 5.1): (1) ξ= C_7-N_1-C_8-C_9, (2) ξ^2= C_7-N_1-C_8-C_{13}, (3) ξ^3= H_6-N_1-C_8-C_9, and (4) ξ^4= H_6-N_1-C_8-C_{13}. The effects of these different definitions upon the conformational profile of TA were investigated using the PBE0/6-31+G(d) level of theory. Preliminary relaxed conformation energy scans were also performed on FA using 2 and 1° step sizes around the maxima and minima (section 4.3.1) respectively, at the same level of theory, in order to investigate the origin of the slight asymmetry in FA.

5.2.2. Ab initio Conformational Energy Profiles for Fenamates

To test the sensitivity to method of the different fenamate molecules' conformational profiles (Figure 5.1), I carried out a relaxed scan at the HF/6-31+G(d) level of theory, as well as single-point energy calculations at the MP2/6-31+G(d) level of theory on the optimized PBE0/6-31+G(d) conformations. The PBE0/6-31+G(d) energies were then revaluated within a polarizable continuum model (PCM)[180] with dielectric continuum, ε, value of 3, a typical dielectric environment of organic crystals. PCM has been shown to help the relative energy ranking of some conformational polymorphs (Chapter 4)[131].

The primary aim of this research was not to seek a definitive *ab initio* conformational profile for all molecules, as this would have been beyond the capability of the available computational resources. Indeed, many studies have emphasized the difficulty in evaluating the relative energies of fenamate polymorphs[184,211,228]. Instead, the objective

of this chapter is to present qualitatively correct scans that can be compared among the fenamate molecules.

5.2.3. Investigation of Electronic Effects

In an attempt to distinguish the steric effects from the electronic effects, I studied a series of model molecules where the steric effects had been minimized (Figure 5.6).

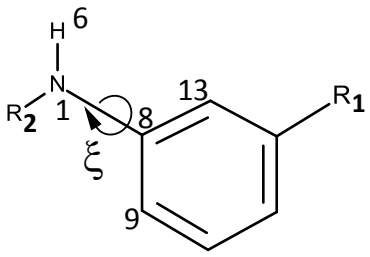


Figure 5.6: Chemical diagram of model molecules used to investigate the electronic contributions. The bonds around N_1 are constrained to be planar (i.e. R_2-N_1-H_6-C_8=180°): for phenylamine (PA) R_1=R_2=H, *N*-methylaniline R_1=H, R_2= CH_3, *N*-vinylaniline R_1=H, R_2 = CH=CH_2 and 3-anilinoprop-2-enoic acid R_1=H, R_2 = CH=$CHCO_2H$; for 3-chloroaniline R_1= Cl, R_2=H and 3-methylaniline R_1=CH_3, R_2=H. (ξ=R_2-N_1-C_8-C_9).

The benzoic acid group was replaced by a series of smaller molecular fragments (Figure 5.6), for example a hydrogen atom. The three bonds to the nitrogen atoms were constrained to be planar by fixing the torsion involving these atoms (e.g. H_7-N_1-H_6-C_8=180° for PA). This constraint prevents the normal pyramidalization at the nitrogen in the torsional potential of phenylamine[227] (Figure 5.7) or the major rearrangement of the second hydrogen that occurs in a relaxed scan of H_7-N_1-C_8-C_9.

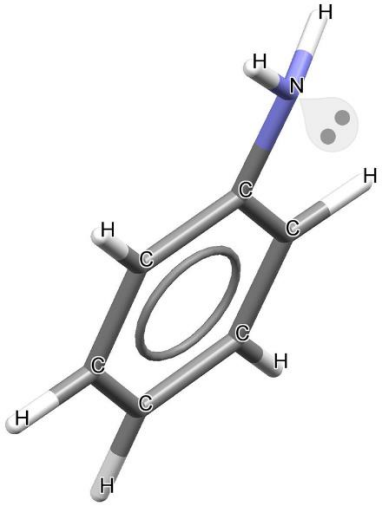


Figure 5.7: Pyramidalization in unconstrained phenylamine (PA) molecule caused by lone pair on nitrogen atom.

The difference in electronic structure with torsion angle was estimated by using the lowest occupied molecular orbital **(**i.e. the highest occupied molecular orbital (HOMO)-*n*, where *n* is an integer**)** and HOMO electron density for the PA model, FA and TA. For

selected constrained (fixed ξ) conformations in the PBE0/6-31+G(d) scan, a single-point calculation with GAUSSIAN03 was used to generate a density file, which was then used to produce the molecular orbitals (MOs) isovalue surfaces of 0.02 au. The graphical interface used was GaussView 5[229].

5.2.4. Investigation of the Steric (Non-Bonded) Effects

Within organic molecules that are typically used in pharmaceuticals, many of the atom-atom distances are similar to intermolecular distances, although some are considerably shorter. The intramolecular steric interactions can be represented as the *exp-6* atom-atom intermolecular parameters, which were derived by fitting to crystal structures and heats of sublimation of hydrocarbons, oxahydrocarbons, azahydrocarbons, chlorohydrocarbons and nitro compounds[42]. This provides the parameters for all intramolecular interactions involving C, N, O, Cl, and H, but the same parameters are used for all C and H atoms, whether aromatic or in the methyl or carboxylic acid groups, which is an approximation to the most force field, e.g. Sybyl typing[230]. This approximation is better justified for FA than for TA (Figure 5.1) because of the importance of distinguishing C in different chemical environments. That is, the chemical environment of aromatic C is different from that of methyl C. There is a polar hydrogen type HB, which was used for both polar hydrogens (H_1 and H_6 in Figure 5.1) because the intramolecular energy is expected to be insensitive to the interactions of the carboxylic acid group.

Table 5.2: Gavezzotti's *exp*-6 atom-atom parameters: $E=Aexp(-BR)-CR^{-6}$. ε is the well depth and R_m is the distance at the minimum[41,231,232]. Italics represent the hydrogen-bonded interactions. [a]From the UNI intermolecular potential function of Mercury[233].

Interactions	A/kJ mol^{-1}	B/Å^{-1}	C/kJ mol^{-1}Å^{-6}	ε/kJ mol^{-1}	R_m/Å
H···H	24158.4	4.01	109.2	0.04	3.36
H···C	120792.1	4.10	472.8	0.21	3.29
H···N	228279.0	4.52	502.1	0.39	2.99
H···O	295432.2	4.82	439.3	0.51	2.80
H···Cl	292963.7	4.09	1167.3	0.50	3.30
C···C	226145.2	3.47	2418.4	0.39	3.89
C···N	491494.5	3.86	2790.7	0.84	3.50
C···O	393086.8	3.74	2681.9	0.67	3.61
C···Cl	390660.1	3.52	3861.8	0.67	3.83
N···O	268571.0	3.86	1523.0	0.46	3.50
O···O	195309.1	3.74	1334.7	0.33	3.61
O···Cl	338297.3	3.63	2782.4	0.58	3.72
HB···O (acids)	26416395.3	8.75	857.7	29.29	1.60
HB···N (-NH···OH)	23867340.0[a]	7.78[a]	1577.4[a]	26.61	1.80

The Gavezzotti parameters used for FA and TA are given in Table 5.2. Although *exp*-6 does not model the electrostatic contributions explicitly, some of the parameters have absorbed some electrostatic contributions during the fitting process. The energies were partitioned into repulsion **(**A*exp*(-*B*R)**)** and dispersion **(**-C/R^6**)** contributions.

The extension of the *exp-6* parameterization to hydrogen bonding interactions required potentials with a particularly deep well at the short hydrogen bonding distances to absorb the electrostatic interactions in hydrogen bonding[231]. The polar hydrogens were assumed not to have significant intermolecular interactions with any atoms other than the hydrogen bond acceptor interactions, which is a reasonable assumption for intermolecular interactions. The parameters for the H···HB and C···HB that were not determined by Gavezzotti may be important.

Interactions that are three atoms or more apart from each other (i.e. 1-4 and above) were then investigated. The longest bond path for FA and TA is 1-11. The atom···atom interactions that dominated the intramolecular contributions were analysed with the *exp*-6 potential. For each fenamate molecule, interactions were classed as dominant if the difference between the minimum and maximum total energy ($E_{repul\text{-}dis}$) of the conformers was greater than or equal to 1 kJ mol^{-1}. It is expected that some interactions in FA and TA will vary with conformations; however, a few interactions will have constant contribution (e.g. HB···O (acid), an important intramolecular H-bond). The interactions with constant contributions are not classed as dominant interactions.

5.3. Results

5.3.1. Effect of Different Definitions of the Same Flexible Bond

The N_1-C_8 rotatable bond can be defined in four ways: C_7-N_1-C_8-C_9, C_7-N_1-C_8-C_{13}, H_6-N_1-C_8-C_9, and H_6-N_1-C_8-C_{13} (Figure 5.1). The conformational energy profile of TA (Figure 5.8) is sensitive to the atoms used in defining the N_1-C_8 rotatable bond. For example, the intramolecular energy penalty at ξ= 0° for the C_7 torsions (ξ=C_7-N_1-C_8-C_9 and ξ^2=C_7-N_1-C_8-C_{13}) are approximately half that of the H_6 torsions (ξ^3=**H**$_6$-N_1-C_8-C_9 and ξ^4=**H**$_6$-N_1-C_8-C_{13}) (Figure 5.8). The relaxation of the hydrogen atom (H_6) was clearly playing a vital role in stabilizing the molecular conformation at ξ= 0°. There were some differences in intramolecular energy penalties as ξ approached 180° for the C_7 torsions (ξ and ξ^2 in Figure 5.8).

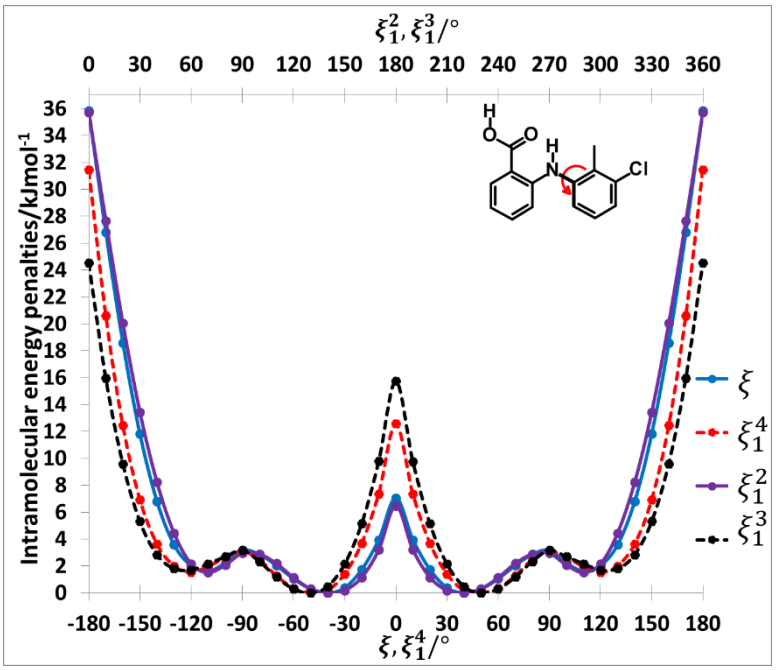


Figure 5.8: Relaxed scan of TA at the PBE0/6-31+G(d) level of theory with the different definitions of flexible N_1-C_8 bond (ξ= C_7-N_1-C_8-C_9, ξ^2= C_7-N_1-C_8-C_{13}, ξ^3=H_6-N_1-C_8-C_9 and ξ^4= H_6-N_1-C_8-C_{13}) used as the constrained geometric parameter.

The torsional barriers for TA are sensitive to the atoms used in defining the N_1-C_8 rotatable bond (Table 5.3). The C_7 torsional barriers (Group 2 of Table 5.3) are larger than the H_6 torsional barriers (Group 2 of Table 5.3). For the C_7 torsions, the H_6 proton relaxes to maintain the intramolecular H-bond, while for the H_6 torsions, the benzoic acid (including the C_7 atom) deviates minimally to prevent deformation of aromatic ring and form better pyramidalization (Figure 5.7) at N_1 for TA. These results show that the torsional barrier and conformational energy profile of TA (Figure 5.8 and Table 5.3) depend upon the selected 4 atoms that define the flexible torsion.

Table 5.4 shows no correlation between shorter (stronger) intramolecular H-bond distances and smaller energy penalties. However, a closer inspection of the atom-atom distances between the H atom of the secondary amine and the H atom of the methyl group (Group 1 of Table 5.5) suggest that steric hindrance also contributes to the intramolecular energy because conformations with higher energy penalties correlate with shorter atom-atom distances.

Table 5.3: Torsional barrier height of the relaxed conformational energy profiles in Figure 5.8 for TA.

	Torsion	Barrier	Barrier relative to ξ
		kJ mol^{-1}	
Group 1	ξ	35.85	0
	ξ^2	35.72	-0.13
Group 2	ξ^3	24.53	-11.32
	ξ^4	31.48	-4.37

Table 5.4: Comparison of intramolecular energies and hydrogen bond distances of TA for the near-planar conformer.

	Torsion	N-H···O=C/Å	ΔE_{intra}/kJ mol^{-1}
Group 1	ξ= 0°	1.850	7.06
	ξ^2= 180°	1.847	6.48
	Δ	**-0.003**	**-0.58**
Group 2	ξ^3= 180°	1.880	15.74
	ξ^4= 0°	1.881	12.57
	Δ	**0.001**	**-3.17**
	$\xi^3(180)$-ξ(0)	*0.030*	*8.68*

Table 5.5: Comparison of the intramolecular energies and the N-H_7···H_X-C_{14} bond distances of TA for the near-planar conformer, where H_X-C_{14} is the methyl group.

Group 1	N-H_7···H_{10}-C_{14}	N-H_7···H_{11}-C_{14}	N-H_7···H_{12}-C_{14}	ΔE_{intra}
	(Å)			kJ mol^{-1}
ξ= 0°	3.491	2.284	2.194	7.06
ξ^2= 180°	3.486	2.272	2.199	6.48
Δ	**-0.005**	**-0.012**	**0.005**	**-0.58**
Group 2				
ξ^3= 180°	3.358	2.135	2.090	15.74
ξ^4= 0°	3.386	2.157	2.117	12.57
Δ	**0.028**	**0.022**	**0.027**	**-3.17**
$\xi^3(180)$-ξ(0)	-0.133	-0.149	-0.104	8.68

5.3.2. Effect of Step Size on the Conformation Profile of Fenamate Molecule

Given that conformational profiles depend on small correlated changes in atomic positions, I investigated the effect of step size on the conformational profiles. For FA, the conformation profile is unaffected by the small step size of 1 and 2° for the minima and maximum respectively (compare Figure 5.10 with Figure 5.11). Over all possible conformations of FA, the barrier to rotation is approximately 6 kJ mol^{-1} (see Figure 5.10 and Figure 5.11).

Although the phenyl ring of FA is symmetric, the rotation of this ring produces asymmetry conformational profiles around the maximum and minima points (Figure 5.10). The slight asymmetry observed for the conformational profile of FA (Chapter 4) is

therefore not due to the small step size. These changes can be substantial; for example, the two minima for FA scan at ξ=38.94° and 144.71° (Figure 5.10) correspond to conformations that differ by 0.61 Å in the 1-4 distance between the amide proton and C_{13} (or C_9), even though the minima only differ in energy by 1.88 x 10^{-3} kJ mol^{-1} (7.16 x 10^{-7} au). Even the C_8-C_9 and C_8-C_{13} aromatic bond lengths differ by ±0.0037 Å for ξ=0 or 180°, but only by ±0.0004 Å for ξ=80° or 100°, with larger changes in the bonded hydrogen positions for FA. It is therefore clear that the molecular structure of the phenyl ring is not perfectly symmetric (Figure 5.9) and that it changes with torsion.

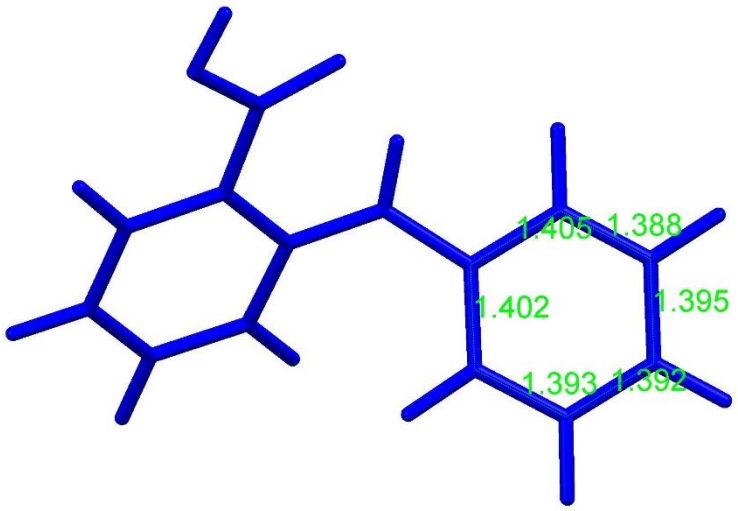


Figure 5.9: Molecular conformation of the relaxed gas phase at PBE0/6-31+G(d) showing the bond lengths of the phenyl C-C bond (ξ=180°) of FA.

For the definitive scans, full optimization was performed near each symmetry-independent potential minimum with a step size of Δξ=5°. The highly repulsive points for the substituted fenamates (TA and MA) were omitted due to the huge energy penalty as ξ approached 180° (section 4.3.1 and Figure 5.8).

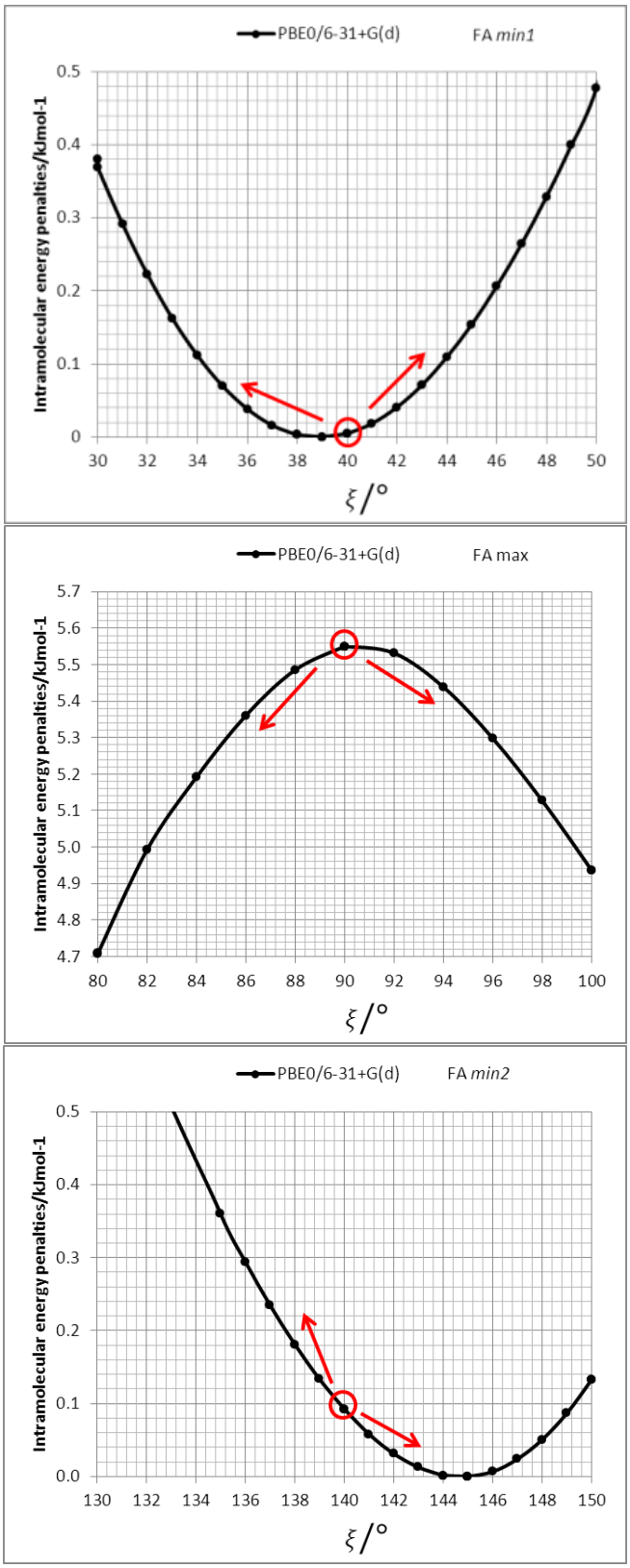


Figure 5.10: Relaxed scans of FA at PBE0/6-31+G(d) with step sizes 2 and 1° around the maxima and minima respectively. The red circles indicate the starting points of the scans, while the arrows indicate the directions of the scans.

5.3.3. Conformational Energy Profile of the Fenamates

The conformational profile for four fenamate molecules (Figure 5.11) shows that there are two distinct minima, which are only close in energy for the symmetric FA and R_2-substituted ClFA (Figure 5.1). The ClFA molecule has one global minimum at 148.38°, in addition to the local minimum at 35.86°, while FA has two global minima at 38.94° and 144.71° (Figure 5.11). There is one potential maximum that varies slightly between the molecules for the planar conformation (as drawn in Figure 5.1); there is a second maximum when the aromatic rings are approximately perpendicular; and there is a third where there is a significant steric clash for the R_1-substituted fenamates TA and MA as the other planar conformation is approached. An analysis of the observed values of this torsion angle in experimental crystal structures containing the fenamate fragment shows that the observed angles were clustered around the two minima (Chapter 4, Figure 4.12), a finding that is consistent with the expectation that most molecules adopt low-energy conformations in crystal structures[208]. Hence, the two low-energy barriers, although only ~5-9 kJ mol^{-1}, clearly have a major effect on the crystal packing. Indeed, this conformational energy is very significant; it is large compared with most measured polymorphic energy differences, including those of TA, which cover less than 2 kJ mol^{-1} (Chapter 4).

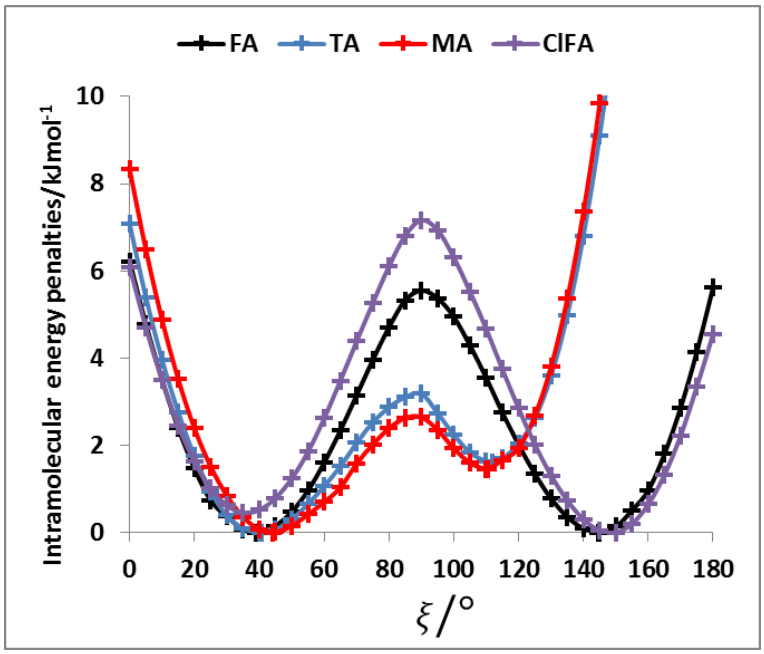


Figure 5.11: Relaxed conformational scans at the PBE0/6-31+G(d) level of theory for the fenamates. The minima were at ξ=38.94° and 144.71° for FA, 40.63°, and 111.86° for TA, 44.08° and 110.48° for MA, and 35.86° and 148.38° for ClFA.

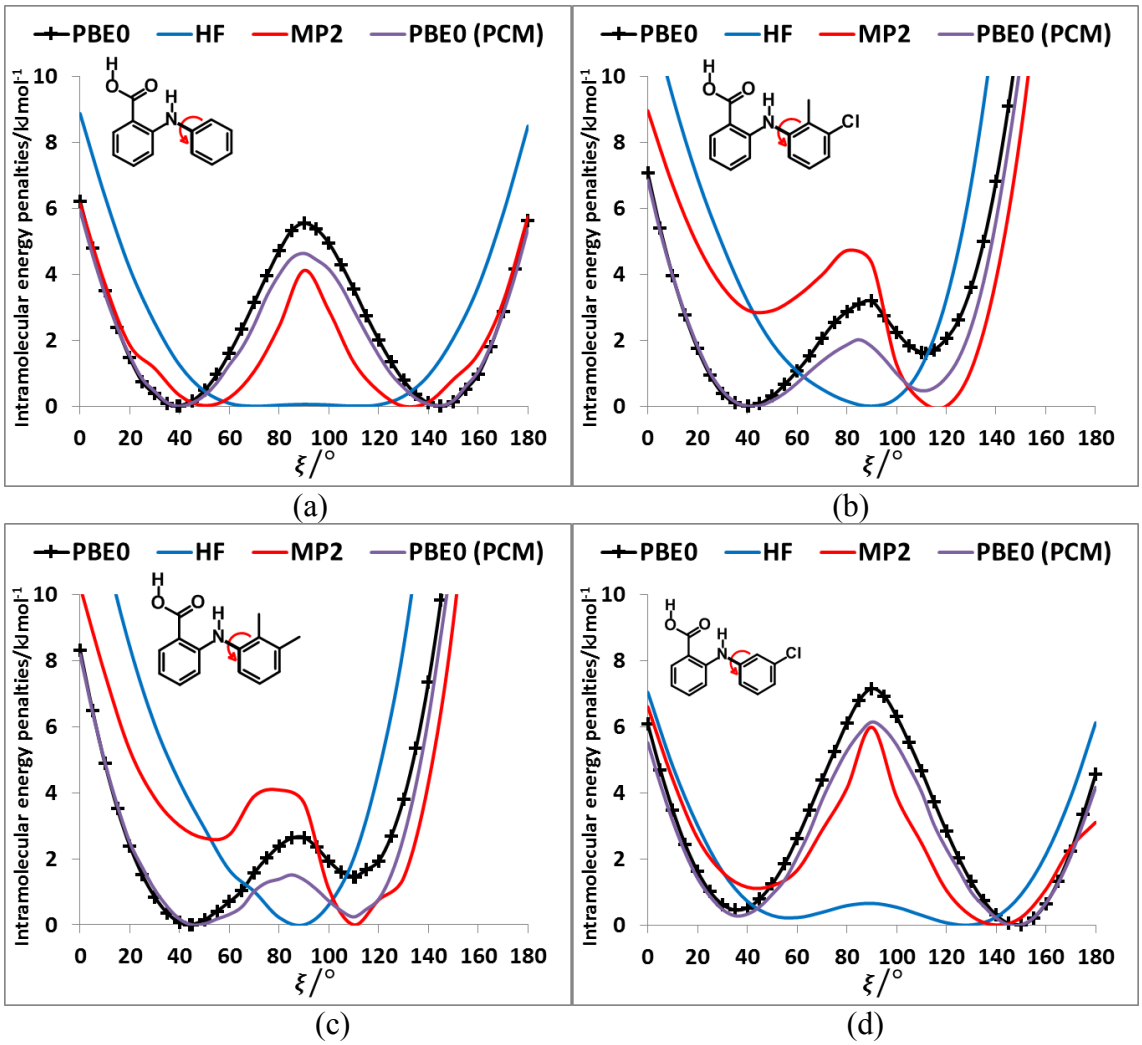


Figure 5.12: Relaxed conformational scans of fenamates (a) FA, (b) TA, (c) MA, and (d) ClFA at HF and PBE0 method with 6-31+G(d) basis set. These are contrasted with the single-point energies at the MP2/6-31+G(d) level and within a polarizable continuum model (PCM) with ε=3 for the PBE0/6-31+G(d) conformations.

By testing the sensitivity of this conformational profile to the choice of *ab initio* method (Figure 5.12), it becomes clear that even obtaining a conformational profile that is in qualitative agreement with the profile derived from experimental crystal structures is sensitive to method. The SCF scan has only one minimum, as seen in Figure 5.12, at a conformation that was not observed in the crystal structures of fenamates (section 4.3.3, Figure 4.12). If the CSP study of Chapter 4 were based on a SCF conformational profile, the crystal structures generated for FA and TA would have been qualitatively incorrect. Therefore, it is essential that the *ab initio* methods include some description of the electron correlation that produces a maximum at around 90°. Although there is moderate agreement on the barriers to conformational change for TA and MA, it is notable that single-point MP2 calculations swap the relative energy positions of the two minima (Figure 5.12). The PBE0 calculations are in better agreement with the analysis of the

crystalline conformations of fenamate-like fragments with a substituent at C_{13} (Chapter 4).

The scans carried out within PCM with ε=3 showed a reduction in energy penalty around the energy maxima (ξ∼90°) for all three fenamates (Figure 5.12). This reduction is more pronounced for TA and MA around the local minima (ξ∼110°) than it is for FA. This finding suggests that the effect of solution or the crystalline environment, the method used for generating the final energy landscapes of FA and TA in Figure 4.14 of section 4.3.6, does not qualitatively change the conformational profile. Hence, the clustering of observed torsion angles in fenamate crystal structures around the minima is not an artefact of the neglect of the environment (Figure 4.12 in section 4.3.3). The results shown in Figure 5.12 confirm that the definitive *ab initio* conformational profiles cannot be obtained. Nonetheless, the scans in Figure 5.11 are adequate for the purposes of this study.

5.3.4. Rigid Versus Relaxed Scans of Fenamates

To establish the importance of changes in the other torsion angles, bond angles, and bond lengths during the relaxed scans in Figure 5.11, the conformation scans were repeated for the three fenamates, starting with the fully optimized structure at the PBE0/6-31+G(d) level of theory. Only torsion angle ξ was allowed to vary, which meant that each point had an identical conformation for all three *ab initio* methods (HF, MP2 and PBE0). The difference between rigid (solid lines in Figure 5.13) and relaxed (dotted line in Figure 5.13) scans was very marked for all the fenamates, even in the lowest energy regions, where the changes in the benzoic acid ring and the intramolecular hydrogen bond were small.

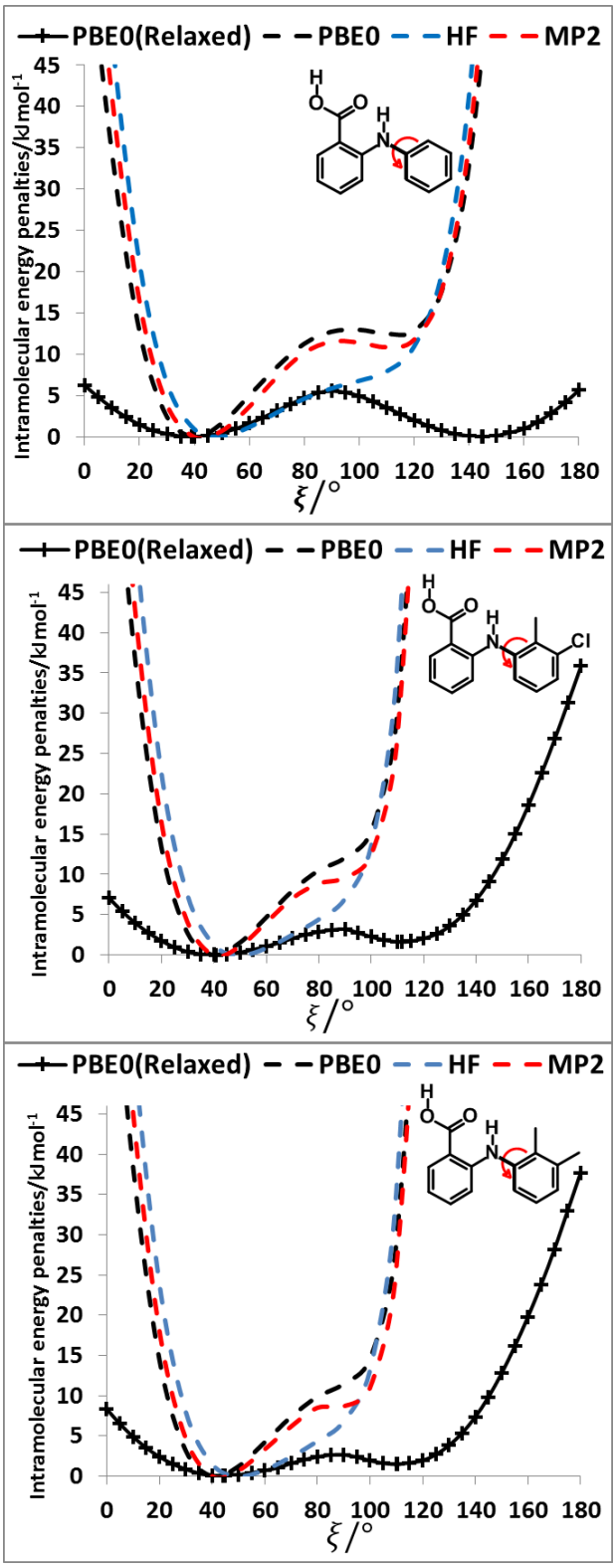


Figure 5.13: Comparison of relaxed (solid) and rigid (dotted line) scan of FA (top), TA (middle) and MA (right). For the rigid scans, calculations were performed using HF, MP2, and PBE0 methods with 6-31+G(d) basis set as a function of torsion angle ξ. The relaxed scans from Figure 5.11 were performed at the PBE0/6-31+G(d) level of theory.

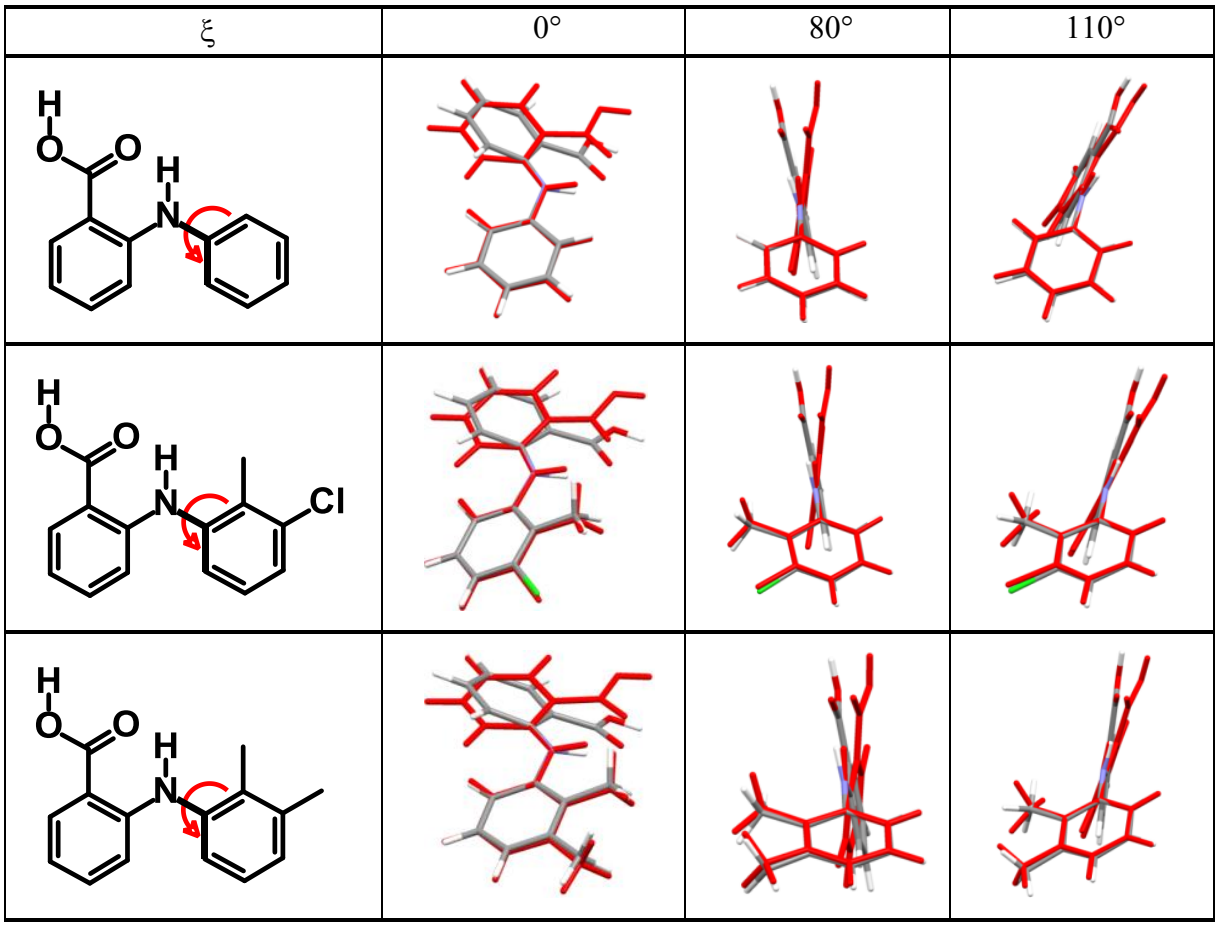


Figure 5.14: Overlay of relaxed (coloured by element, Figure 5.11) and rigid (red, Figure 5.13) conformations of FA (top), TA (middle) and MA (bottom); overlaying the atoms defining the ξ torsion angle at the PBE0/6-31+G(d) level of theory.

Figure 5.14 shows a comparison of the rigid and relaxed conformations at the repulsive barrier and the second minimum for FA, TA, and MA. For TA, the differences in the relaxed and rigid scan conformations are very marked around the energy maxima (Figure 5.14, ξ=80°), even though these maxima are less than 7 kJ mol^{-1} in the relaxed scans (Figure 5.13). For FA, there is a significant change in the relative orientation of the benzoic acid group, changing the internal hydrogen bond around all maxima. The methyl torsion is clearly playing a role in reducing the energy penalty around the 90° maximum for TA and MA (Figure 5.14). Indeed, even for FA, the changes in the other conformational variables are vital to produce a low energy maximum at 90°, to the extent that it is not surprising that even the relaxed scans differed in this region, producing the slight asymmetry between the two minima (Figure 5.11).

The contrast between the relaxed and rigid (Figure 5.13) scans and the corresponding conformations (Figure 5.14) confirm that the changes in the other conformational variables played a major role in lowering the conformational barriers, providing low values over a wide range of ξ angles, as implied by the crystallographic data (Chapter 4).

5.3.5. Splitting electronic from steric contributions to the barrier to rotation

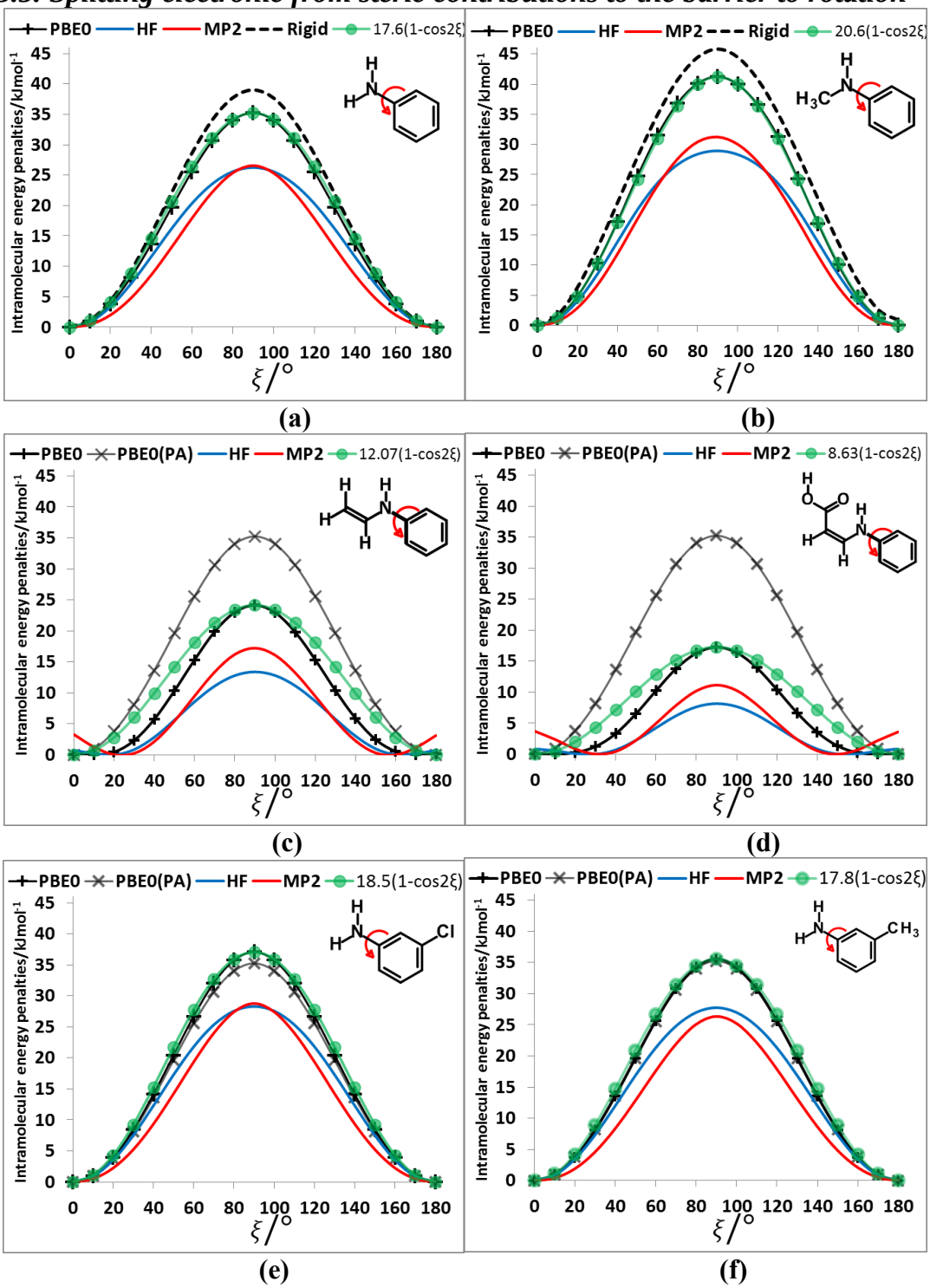


Figure 5.15: The solid black lines show the relaxed torsional scans of planar-N-constrained models for the phenyl rotation (Figure 5.6), where the benzoic acid group of FA has been replaced by (a) hydrogen atom (PA), (b) methyl, (c) vinyl and (d) prop-2-enoic acid, and the hydrogen in the meta position of PA has been replaced by (e) chlorine, and (f) methyl using HF, PBE0 and MP2 methods with the 6-31+G(d) basis set. Plots of $(h/2)(1-\cos(2\xi))$, where h is the height of the barrier of relaxed PBE0/6-31+G(d) scans, are shown in green. For (c)-(f), the PBE0/6-31+G(d) relaxed scans of PA from (a) are shown in grey for comparison. In (a) and (b) a rigid scan at the PBE0/6-31+G(d) level of theory is shown by a dotted line.

To establish the importance of the intramolecular steric clashes, the rigid and relaxed scans were repeated with model molecules that replaced the benzoic acid group of FA with smaller fragments and the bonds around the nitrogen constrained to be planar (Figure 5.6) in order to avoid the pyramidalization of the amine (Figure 5.7). The potential energy scan has a very large maximum at 90° for planar-N-constrained phenylamine (PA), approximately 6 times higher than the barrier in the fenamates (reflected in the scale change between Figure 5.15a and Figure 5.11). Replacing one constrained hydrogen with a methyl (Figure 5.6) produces a slightly larger barrier but very little asymmetry, strongly suggesting that this is an electronic effect of conjugation between the lone pair on the nitrogen and the aromatic ring.

Adding a double bond to PA reduces the barrier to rotation by 11.10 kJ mol^{-1} (Figure 5.15c). Adding a carboxylic acid that forms an intramolecular H-bond to the N-H group further reduces this barrier by almost 7 kJ mol^{-1} (Figure 5.15d). These observations imply that the conjugation of the benzoic acid ring of the fenamates contributed significantly to reducing the electronic barrier. The intramolecular hydrogen bond in this model molecule (Figure 5.15d) varies in length from 1.94 to 1.92 Å as ξ changes from 0 to 180°; by comparison, the FA hydrogen bond varies from 1.87 to 1.84 Å. Thus, it is evident that the intramolecular H-bonding in the fenamates reduces the electronic barrier. In contrast, substituting Cl and CH_3 at the meta position of PA (Figure 5.15e and f) shows little change to the barrier height, with half-height (*h/2*) values of 3-chloroaniline (18.5 kJ mol^{-1}) and 3-methylaniline (17.8 kJ mol^{-1}) representing only a small increase over that of PA (17.6 kJ mol^{-1}).

The scans in Figures 5.15a, b, e, and f have no maxima at 0 or 180°, confirming that the steric clash between the aromatic C_6-**H** and C_{13}-$\mathbf{R_1}$ (Figure 5.1) of the phenyl ring is responsible for this maximum. These curves are very well reproduced by $((h/2)(1-\cos(2\xi))$ where *h* is the potential maximum. As the nitrogen substituents (Figures 5.15c and d) get larger, there are signs of additional steric effects, particularly at the MP2 level, suggesting more changes in intramolecular dispersion. The difference between a rigid scan and a relaxed one is small (Figures 5.15a and b), and the differences in the curves according to the type of calculation are relatively minor compared with the qualitative differences between the HF and correlated methods for the fenamates (Figure 5.12). From contrasting the conformational scans of model molecules with minimal steric effects (Figure 5.15) with those of the fenamates (Figure 5.11), the overriding conclusion is that there is an electronic contribution to the torsional barrier at ξ=90°, which can be

represented by a $(h/2)(1-\cos(2\xi))$ term, where h is the maximum energy from the electronic contribution.

There is a dramatic change in the HOMO isosurfaces of the phenyl ring with torsion angle (Figure 5.16), which is remarkably similar between PA and the fenamates, FA and TA. This finding suggests that the electronic contribution observed for PA is fundamental to the fenamates. The destructive interference of the MOs at ξ=90° (Figure 5.16 and Figure 5.17b) is the origin of the huge torsional barrier that is seen in each of the model molecules in Figure 5.15. The smaller torsional barriers after adding a double bond (Figure 5.15c) and acidic (Figure 5.15d) substituents to PA (Figure 5.15a) was a result of increased electron delocalisation (compare Figure 5.17a with Figure 5.17b).

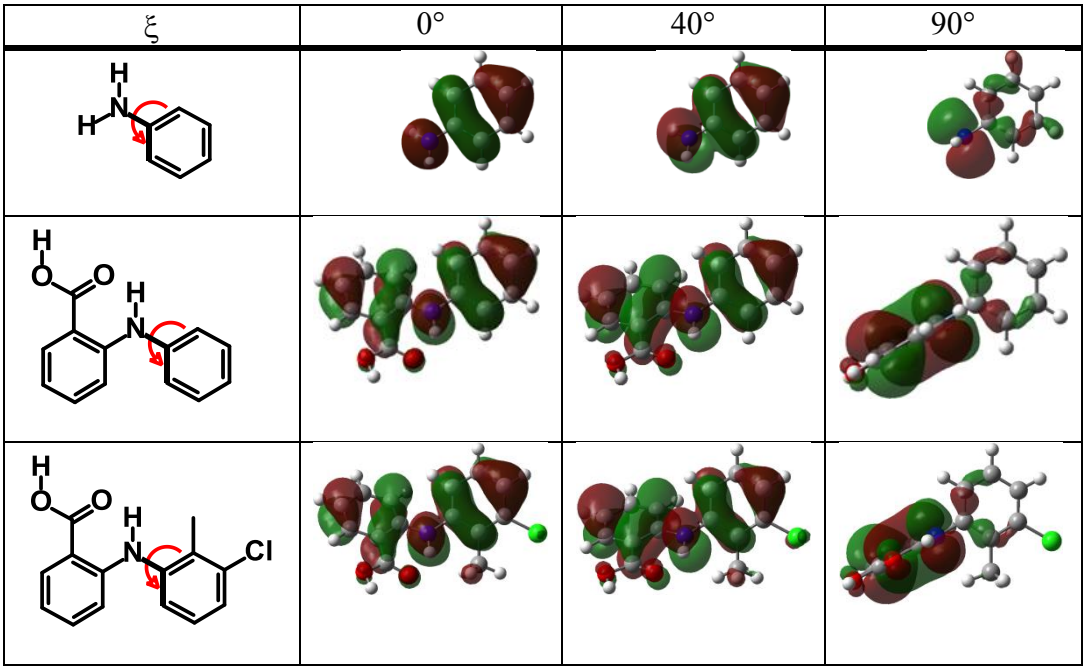
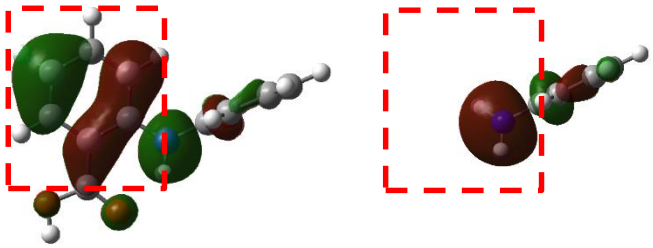


Figure 5.16: The electron density surface of HOMO frontier orbitals for different conformations of PA (top), FA (middle) and TA (bottom) obtained from the PBE0/6-31+G(d) level of theory. Contour map isovalues were drawn at a 0.02 au level.



(a) **(b)**

Figure 5.17: Alternative HOMO isosurface view of (a) FA and (b) PA molecule at ξ=90° from Figure 5.16.

My attempt at modelling delocalisation over the entire FA molecule even for the lowest occupied π orbital failed (Figure 5.18), which suggests that a completely flat model of

FA ($\xi = 0$) is unrealistic. This is because of the huge steric barrier (Figure 5.13) of a planar conformation and the linear combinations of atomic orbitals in many molecular orbitals are sensitive to conformation. The electronic effects are not well defined for TA i.e. no Figure 5.18 equivalent for TA due to the bulky substituent, hence excluded in the analysis.

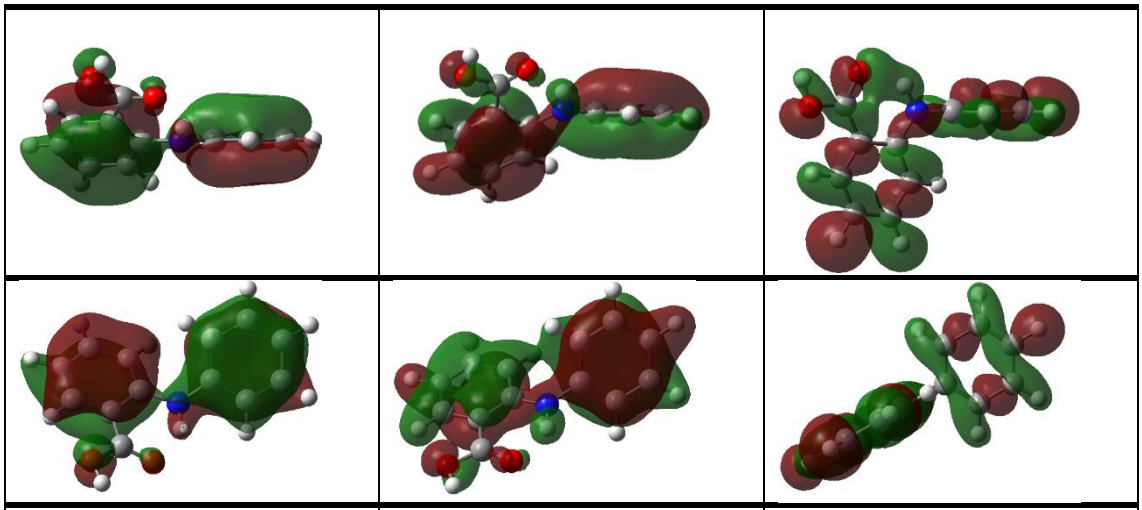


Figure 5.18: HOMO-10 (46 MOs) isosurface view of side (top) and front (bottom) FA molecule at ξ=0, 40 and 90°.

The MOs' picture can only give a qualitative guide to the electronic contribution because it is not possible to have an aromatic substitution at R_2 or a methyl group substitution on C_{13} (as in MA and TA) without steric interactions. At present, independently investigating how the electronic barrier varies among the fenamates is not possible. However, it is clear from Figure 5.15 that whilst the benzoic acid substituents will have a dramatic effect on lowering the barrier, the meta-substitution of chlorine and methyl on the phenyl ring has a much smaller effect on the barrier.

5.3.6. Estimating the Steric Contributions with an Atom-Atom Potential

The steric effects of FA and TA can be analysed by using Gavezzotti's parameters (Table 5.2) to produce a conformational profile (see Figure 5.19), which can then be subdivided into the repulsion $(A_{\iota\kappa}\exp(-B_{\iota\kappa}R_{ik}))$ and dispersion $(-C_{\iota\kappa}/R_{ik}^6)$ contributions. This subdivision (Figure 5.19) suggests that the repulsion terms dominate, producing maxima at 0 and 180° and minimum at 90°, a finding that is consistent with the supposition that these maxima occur because of steric clashes. However, the repulsions are severely overestimated. The dispersion contribution gives a maximum at $\xi \sim 90°$, which is of the same order of magnitude as the *ab initio* maxima, particularly for FA. This confirms the observation from Figure 5.12 that electron correlation, the origin of dispersion interactions, makes a significant contribution to this maximum, supplementing

the electronic effect (c.f. Figure 5.15). The difference between TA and FA in Figure 5.19 is consistent with the large dispersion interaction of the Cl substituent as well as the repulsion from the methyl substituent.

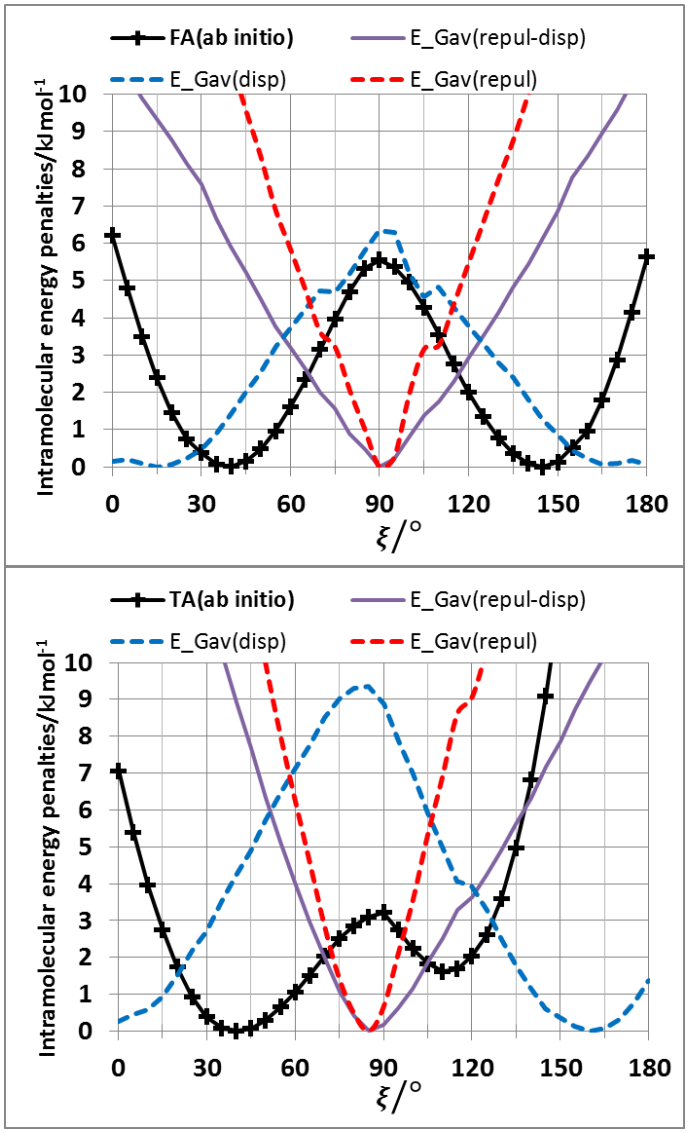


Figure 5.19: Comparison of the *ab initio* (black) intramolecular energy scans at the PBE0/6-31+G(d) level of theory for FA (top) and TA (bottom) with Gavezzotti's *exp*-6 intramolecular energy models summed over all intramolecular atom-atom distances from the third nearest neighbour (1-4) and above. The net effect (purple) is subdivided into the repulsive (red), and dispersion (blue) contributions.

5.3.7. Investigating the Important Interactions with Atom-Atom Potential

For FA and TA, the dominant interactions involve C, H, and HB (Table 5.6), while the dominant interactions involving N and Cl (Table 5.7) were exclusive to TA. Although FA and TA have the N atom, only TA has the methyl substituent ($N_1\cdots\mathbf{C_{14}}^{f}$ and $N_1\cdots\mathbf{H_{12}}^{f}$, Table 5.7). Of all the dominant interacting atoms for FA and TA, only Cl of TA interacts with atoms (methyl substituent) belonging to the same aromatic ring ($^{df}\mathbf{C_{14}}\cdots Cl_1$, $^{df}Cl_1\cdots\mathbf{H_{10}}$ and $^{df}Cl_1\cdots\mathbf{H_{11}}$, Table 5.7). These results confirm the observation that the

methyl substituent plays a role in reducing the energy penalty for TA (section 5.3.4, Figure 5.14).

Table 5.6: Atom-atom interactions of FA that dominate the intramolecular energies as estimated using Gavezzotti's exp-6 potential. The interaction is listed as dominant if the difference between the minimum and maximum total energy ($E_{repul\text{-}dis}$) is greater than or equal to 1 kJ mol^{-1}.

at_i···at_k	neigh	R_{ik}/kJ mol^{-1}		E_{repul}/kJ mol^{-1}		E_{disp}/kJ mol^{-1}		$E_{repul\text{-}disp}$/kJ mol^{-1}	
		min	*max(Δ)*	*min*	*max(Δ)*	*Min*	*max(Δ)*	*min*	*max(Δ)*
C_7···C_9	4	3.07	3.71**(0.64)**	0.57	5.33**(4.76)**	-2.89	-0.92**(1.96)**	-0.35	2.45**(2.80)**
C_7···C_{13}	4	3.07	3.71**(0.64)**	0.57	5.32**(4.75)**	-2.88	-0.92**(1.96)**	-0.35	2.44**(2.79)**
C_6···C_9	5	3.10	4.43**(1.33)**	0.05	4.75**(4.70)**	-2.70	-0.32**(2.39)**	-0.39	2.05**(2.43)**
C_6···C_{13}	5	3.11	4.44**(1.33)**	0.05	4.68**(4.63)**	-2.68	-0.32**(2.36)**	-0.39	2.00**(2.38)**
C_6···H_7	6	2.52	4.88**(2.36)**	0.00	3.98**(3.98)**	-1.86	-0.03**(1.82)**	-0.21	2.12**(2.33)**
C_8···H_5	5	2.54	2.88**(0.34)**	0.89	3.57**(2.68)**	-1.75	-0.83**(0.92)**	0.07	1.82**(1.76)**
C_9···H_5	6	2.57	4.16**(1.59)**	0.00	3.17**(3.16)**	-1.63	-0.09**(1.54)**	-0.20	1.54**(1.74)**
H_5···H_7	7	2.08	4.76**(2.68)**	0.00	5.78**(5.78)**	-1.35	-0.01**(1.34)**	-0.04	4.43**(4.47)**
C_{13}···HB_6	4	2.51	3.30**(0.78)**	0.16	4.03**(3.87)**	-1.87	-0.37**(1.51)**	-0.21	2.16**(2.36)**
HB_6···H_{11}	5	2.30	3.68**(1.38)**	0.01	2.40**(2.39)**	-0.74	-0.04**(0.70)**	-0.04	1.66**(1.70)**
C_6···H_8	4	2.89	3.10**(0.20)**	4.85	9.85**(5.00)**	-4.12	-2.74 **(1.38)**	2.12	5.73**(3.62)**
C_6···H_{11}	6	2.52	4.88**(2.37)**	0.00	3.96**(3.96)**	-1.85	-0.03**(1.82)**	-0.20	2.11**(2.31)**
C_{13}···H_5	6	2.57	4.17**(1.59)**	0.00	3.14**(3.14)**	-1.62	-0.09**(1.53)**	-0.20	1.52**(1.72)**
H_5···H_{11}	7	2.09	4.76**(2.67)**	0.00	5.48**(5.48)**	-1.30	-0.01**(1.29)**	-0.04	4.18**(4.22)**
C_9···HB_6	4	2.51	3.30**(0.79)**	0.16	4.06**(3.90)**	-1.88	-0.36**(1.52)**	-0.21	2.18**(2.38)**
HB_6···H_7	5	2.30	3.69**(1.39)**	0.01	2.38**(2.37)**	-0.74	-0.04**(0.69)**	-0.04	1.64**(1.69)**

Table 5.7: Atom-atom interactions of TA that dominate the intramolecular energies as estimated from Gavezzotti's exp-6 potential. The interaction is listed as dominant if the difference between the minimum and maximum total energy ($E_{repul\text{-}dis}$) is greater than or equal to 1 kJ mol^{-1}. ^{d}Interactions between atoms on the same aromatic ring. ^{f}Interactions involving the methyl substituent.

at_i···at_k	neigh	R_{ik}/kJ mol^{-1}		E_{repul}/kJ mol^{-1}		E_{disp}/kJ mol^{-1}		$E_{repul\text{-}disp}$/kJ mol^{-1}	
		min	*max(Δ)*	*min*	*max(Δ)*	*min*	*max(Δ)*	*min*	*max(Δ)*
C_7···C_9	4	3.02	3.61**(0.59)**	0.81	6.30**(5.49)**	-3.17	-1.09**(2.09)**	-0.28	3.13**(3.40)**
C_7···C_{13}	4	3.14	3.72**(0.58)**	0.56	4.17**(3.61)**	-2.52	-0.91**(1.61)**	-0.35	1.65**(2.01)**
C_6···C_9	5	3.08	4.12**(1.04)**	0.14	5.13**(4.99)**	-2.82	-0.50**(2.33)**	-0.39	2.31**(2.69)**
C_6···C_{13}	5	3.24	4.44**(1.19)**	0.05	2.92**(2.87)**	-2.07	-0.32**(1.76)**	-0.39	0.85**(1.23)**
C_6···H_7	6	2.48	4.45**(1.97)**	0.00	4.58**(4.58)**	-2.02	-0.06**(1.96)**	-0.20	2.56**(2.76)**
C_8···H_5	5	2.54	2.89**(0.35)**	0.86	3.63**(2.77)**	-1.76	-0.81**(0.95)**	0.05	1.87**(1.82)**
C_9···H_5	6	2.60	3.69**(1.09)**	0.03	2.86**(2.83)**	-1.54	-0.19**(1.35)**	-0.21	1.32**(1.53)**
H_5···H_7	7	2.14	4.18**(2.04)**	0.00	4.51**(4.50)**	-1.13	-0.02**(1.11)**	-0.04	3.37**(3.42)**
C_{13}···HB_6	4	2.54	3.22**(0.68)**	0.22	3.64**(3.41)**	-1.76	-0.43**(1.34)**	-0.20	1.87**(2.07)**
HB_6···$H_{11}{}^{f}$	6	2.19	4.50**(2.31)**	0.00	3.64**(3.64)**	-0.98	-0.01**(0.96)**	-0.04	2.66**(2.71)**
HB_6···$H_{12}{}^{f}$	6	2.28	3.15**(0.87)**	0.08	2.54**(2.46)**	-0.77	-0.11**(0.66)**	-0.03	1.77**(1.80)**
C_6···C_8	4	2.89	3.10**(0.21)**	4.85	9.95**(5.11)**	-4.14	-2.73**(1.41)**	2.11	5.81**(3.70)**
C_6···$C_{14}{}^{f}$	6	3.10	5.10**(2.00)**	0.00	4.82**(4.82)**	-2.73	-0.14**(2.59)**	-0.39	2.10**(2.48)**
C_7···$C_{14}{}^{f}$	5	3.12	4.14**(1.02)**	0.13	4.55**(4.42)**	-2.64	-0.48**(2.16)**	-0.39	1.91**(2.30)**
C_7···$H_{12}{}^{f}$	6	2.55	3.85**(1.30)**	0.02	3.44**(3.42)**	-1.71	-0.14**(1.56)**	-0.20	1.73**(1.94)**
N_1···$C_{14}{}^{f}$	4	2.78	2.98**(0.21)**	4.87	10.81**(5.94)**	-6.07	-3.95**(2.12)**	0.93	4.74**(3.82)**
N_1···$H_{12}{}^{f}$	5	2.43	2.62**(0.19)**	1.62	3.85**(2.22)**	-2.43	-1.54**(0.88)**	0.08	1.42**(1.34)**
$^{df}C_{14}$···Cl_1	4	3.04	3.11**(0.07)**	6.88	8.75**(1.87)**	-4.88	-4.27**(0.61)**	2.61	3.87**(1.26)**
$^{f}C_{14}$···HB_6	5	2.41	3.71**(1.30)**	0.03	6.18**(6.15)**	-2.41	-0.18**(2.23)**	-0.21	3.77**(3.97)**
$^{d,f}Cl_1$···H_{10}	5	2.57	3.24**(0.68)**	0.51	8.13**(7.62)**	-4.10	-1.00**(3.09)**	-0.49	4.04**(4.53)**
$^{df}Cl_1$···H_{11}	5	2.69	3.74**(1.05)**	0.07	4.89**(4.83)**	-3.08	-0.42**(2.66)**	-0.50	1.81**(2.31)**

Analysis of which particular atom-atom interactions vary most with conformation (Table 5.6 and Table 5.7) reveals that intramolecular interactions involving the N-H proton, type HB, as well as certain aromatic and TA methyl protons, type H, vary strongly with conformation, as expected based on Figure 5.14. Therefore, the poorly determined parameters for HB and H strongly affect the energy profiles of Figure 5.19.

5.4. Discussion: The Physical Origins of Torsional Potentials

The torsional barriers for fenamates apart from $\xi > 140°$ are smaller than for PA and ethane. For the fenamates, the difference in energy penalty arises from the high degree of conjugation between the two aromatic rings of the fenamates via $\pi\cdots\pi$ overlap involving the lone pair of electrons of the N atom (Figure 5.16)[234]. This conjugation effect is absent for ethane and reduced for PA, because these have just one aromatic ring. It is clear that compared to PA, the barrier to rotation reduces as the model molecules tend toward two aromatic rings. These observations confirm that the delocalisation of electrons plays a role in the origin of the torsional barrier for fenamate molecules FA and TA.

The maximum at $\xi = 90°$ (Figure 5.15) is due to dispersion effects. The investigation of different methods for producing conformational profiles demonstrates the importance of accounting for correlation effects. For example, the HF method is hopeless for reproducing the correct conformation profiles around the minima for fenamate molecules FA, TA, MA and ClFA (Figure 5.12) because the method excludes the correction effects. Therefore, it is vital that methods of examining the conformational profiles include the correlation effects. This is most important in CSP, where accurate ranking of generated crystal structures depends upon accurately evaluating the energy penalty contribution to the total lattice energy (equation 5.1).

Although the conformational profiles are dependent on specific atoms that define flexible torsion (Figure 5.8), the comparison of the rigid and relaxed scans shows the important role that the atomic positions of other atoms play (Figure 5.13 and Figure 5.14) during relaxation. For example, during the relaxation of TA, methyl rotation plays a role in reducing the energy penalty and barrier to rotation (section 5.3.4). The enormous difference between the relaxed and rigid conformational profiles (Figure 5.13) and their conformers (Figure 5.14) suggests that the atomic positions of most atoms in a molecule during relaxation are important. Consequently, we can conclude that the origin of the torsional barrier involves most atoms in a molecule.

By analysing the low-energy torsional barriers in fenamates (Figure 5.11), it has become clear that larger organic molecules retain the contributions as identified for small-

model molecules such as ethane. There is both an electronic and a steric component to these torsional barriers in fenamates. However, as the model molecules become larger, the effects of small correlated changes in the other bond angles and the dispersion contributions become very significant (Figure 5.15c and d). As the size of the molecule increases in both the number of atoms and the electron density, there is an increasing contribution to the torsional profile from the intra-molecular equivalent of the intermolecular dispersion.

Separating the "through space" intramolecular dispersion from the other electron correlation effects—specifically those that contribute to the $(h/2)(1-\cos(2\xi))$ electronic barrier from "conjugation" or delocalization between the two aromatic rings—is unlikely to be meaningful when using an *ab initio* method that is still approaching the quantitative accuracy needed.

5.5. Summary

I could not use the definitive *ab initio* torsional potential for each fenamate, as the torsional potentials of organic molecules containing aromatic rings are very demanding of the electronic structure method used. However, PBE0 method and 6-31+G(d) basis set were sufficient to investigate the origin of the barrier to rotation for FA and TA.

The analysis of the fenamates' torsional potential reveals that, like the classical small-molecule torsional potentials of ethane, there is both an *electronic effect* (loosely described as changes in MOs involving $\pi\cdots\pi$ between the aromatic rings and the nitrogen lone pair destabilizing the non-planar conformations) and a *steric effect* from the variation in overlap of non-bonded atoms. For larger molecules like the fenamate molecules, the intramolecular *dispersion* and small changes in the other conformational variables also make significant contributions to the torsional profile.

When the aromatic rings are perpendicular to each other, the destructive interference of MOs creates an intrinsic barrier. The two rings in fenamate molecules stabilise the expected high-energy barrier. A $\cos2\xi$ term accurately represents the torsional barrier of model molecules with minimal steric effects that have been constrained to prevent pyramidalization. Finally, this chapter showed that the torsional barriers for FA and TA depend on small correlated changes in degrees of freedom other than the most flexible torsion. We observed this by contrasting the conformations and their corresponding energies of rigid scans against relaxed scans for FA and TA.

The origin of the torsional barrier of fenamates is therefore a combination of intrinsic intramolecular *dispersion*, *electronic* contribution and the steric *repulsion*. A new

analytical form for intramolecular energies (Chapter 6) of fenamates that is physically justified must incorporate these origins and account for small correlated changes among different conformations.

Chapter 6. Toward New Force fields for Fenamate Molecules

"*Avoid complexities. Make everything as simple as possible*."

— Henry Maudslay

6.1. Introduction

6.1.1. Background

Force field methods (section 2.5) for calculating a system's energy ignore electronic motions and instead calculate the system's energy as a function of the nuclear positions alone[63]. Such force field methods are less computationally demanding than quantum mechanical calculations, but they cannot predict properties that depend upon a molecule's electronic distribution. In addition, the limitations of electronic structure calculations, as discussed in Chapters 2 (section 2.7) and 4 (sections 4.3.8 and 4.3.9), suggest the need for a new analytical model that can combine the accuracy of electronic methods and the speed of force field methods. This chapter proposes a solution to this problem.

In general, force field methods are based on two important components: (i) the functional form of the potential energy functions; and (ii) the parameterisation of these functions[235,236]. In Chapter 5, I investigated the first component, the origin of the torsional barrier for fenamate molecules. For small molecules that were constrained to have minimal steric clashes, the origin is mainly an *electronic* effect and can be expressed as a $cos2\xi$ term. However, for larger molecules like fenamic (FA) and tolfenamic acids (TA), there are additional correlated changes, namely the *repulsion* and *dispersion* effects. Chapter 5, by investigating these contributions with an *exp*-6 term, dealt with the first component required for building a force field.

The present chapter deals with the second component of force fields: *re*-parameterisation. In this chapter, I will derive an analytical model for FA and TA (Figure 6.1) by fitting the conformational profiles (Figure 5.11) to analytical expressions discussed in Chapter 5. One of the key conclusions of Chapter 5 was that definitive *ab initio* conformational energy scans could not be obtained. However, the *ab initio* scans at the PBE0/6-31+G(d) level of theory (Figure 5.11) are adequate for the purposes of creating a new analytical model.

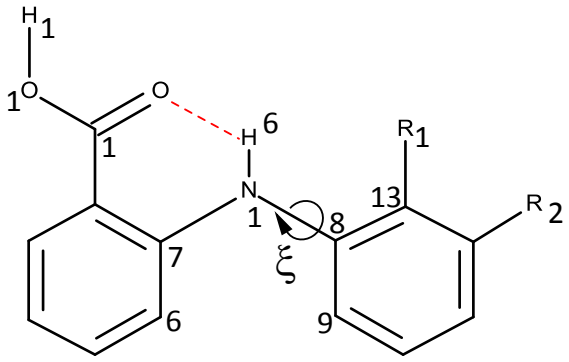


Figure 6.1: The fenamate family, showing the low-barrier torsion angle (ξ=C_7-N_1-C_8-C_9) and atomic numbering. ξ=0 when the aromatic rings are coplanar as drawn. The fenamates mentioned in this chapter are fenamic acid (FA) R_1= R_2= H, tolfenamic acid (TA) R_1= CH_3, R_2= Cl, mefenamic acid (MA) R_1= R_2= CH_3, flufenamic acid (FFA) R_1= H, R_2= CF_3 and clorofenamic acid (ClFA) R_1=H, R_2=Cl. The dotted line represents an intramolecular hydrogen bond.

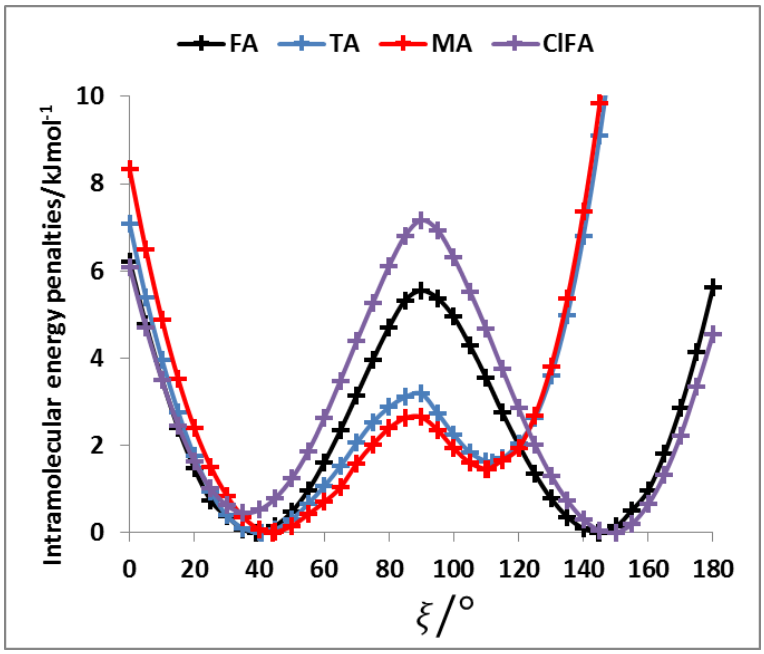


Figure 6.2: Relaxed conformational scans at the PBE0/6-31+G(d) level of theory for the fenamates from section 5.3.3. Fenamic acid (FA), tolfenamic acid (TA), mefenamic acid (MA), and chlorofenamic acid (ClFA).

6.1.2. A Case for Modelling Decoupled Intramolecular Energy

For challenging applications involving molecular recognition of pharmaceuticals, such as computer-aided drug design, the importance of an accurate balance of intermolecular and intramolecular forces cannot be overemphasized[93]. The success of CSP studies[17,127] (Chapters 3 and 4) and the limitations of current force field (section 2.5.4) and electronic structure methods (sections 2.7 and 4.3.9) suggest decoupling the models for the intermolecular forces from those for the intramolecular forces (i.e. conformational

changes) as a means of more accurately modelling force fields for pharmaceutical molecules should be attempted.

6.1.3. Analytical Models of Intramolecular Energies

Traditional force field modelling includes an explicit torsional term[237] of the general form:

$$E(\xi) = \sum \frac{V_n}{2}[1 + \cos(n\xi - \gamma)] \qquad 6.1$$

where γ defines the phase shift, and n the periodicity. For FA, ξ as defined in Figure 6.1 requires that $\gamma = 0$, and the near-symmetry dictates that n is an even integer. For substituted fenamates, odd values of n could contribute, although the conformational profile should be symmetric about $\xi = 0$.

Non-bonded atom–atom interactions should be able to represent the intramolecular energies well if the origin of the conformational profile is predominantly steric (caused by the varying repulsion between the overlapping charge distributions of 1-4 atoms with ξ). An *exp*-6 atom–atom model (Chapter 5) would be expected to give a better representation than a Lennard-Jones (R^{-12}) model of the variation of the repulsion with distance, given the success of the overlap model in parameterizing intermolecular repulsion potentials[238,239]. Thus, a crude starting point for my investigation of the "non-bonded" contribution to the torsion potential is the *exp*-6 atom–atom model potential with a parameterization that was developed for modelling the intermolecular forces between organic molecules in crystal structures[42,231]:

$$E(\xi) = \sum_{i,k} A_{\iota\kappa} \exp(-B_{\iota\kappa} R_{ik}) - C_{\iota\kappa} R_{ik}^{-6} \qquad 6.2$$

where atoms i and k of atomic types ι and κ are separated by intramolecular distances R_{ik}, calculated from the molecular conformation with torsion angle ξ. This functional form, with no assumed relationship between the like ($\iota\iota$ and $\kappa\kappa$) and unlike ($\iota\kappa$) interactions, means that the parameters ($A_{\iota\kappa}$, $B_{\iota\kappa}$ and $C_{\iota\kappa}$) have at least partially absorbed the electrostatic effects in fitting. This chapter seeks to *re*-parameterize the equations of the origin of torsional potential (equations 6.1 and 6.2) in order to derive analytical models that accurately represent the conformational profiles of fenamic (FA) and tolfenamic acids (TA).

6.1.4. Overview

The goal of this chapter is to develop a generalizable method that uses a FORTRAN code to derive analytical models for the conformational profiles of FA and TA. This

chapter will investigate three analytical forms: the cosine series, rescaling repulsion and combined physical models. Each of these new analytical forms will have different parameters for intramolecular and intermolecular contributions. I will discuss the methodology of fitting analytical models to conformational profiles, and test the transferability properties of the combined physical models for FA and TA to other fenamate molecules (Figure 6.1).

This chapter will show that the cosine series and rescaled repulsion models are inadequate for modelling the conformational energy profiles of fenamates. More importantly, this chapter will show that combining a cos2ξ term to represent the *electronic* barrier with an intramolecular atom–atom *exp-6 repulsion-dispersion* term for all atom pairs separated by three or more bonds (bond-paths ≥ 1-4) allows a very effective representation of the *ab initio* conformational energies.

6.2. Methodology: Fitting Analytical Models to Conformational Profiles

The only weighting applied to the conformational profiles of all the fenamate molecules **(**Figure 5.11 at the PBE0/6-31+G(d) level of theory**)** in this chapter was the restriction of conformational energies to 10 kJ mol^{-1} or below. This weighting was appropriate because this chapter does not seek to represent the steric barrier above 145° for TA (and MA) accurately. Hence, this chapter considers $N_p = 37$ points for the FA and $N_p = 30$ points for TA (and MA). Because the definition of the *ab initio* energies, E_{ab}, are relative to the lowest energy conformation for each molecule, the analytical model's energies, E_{model}, were adjusted by a constant, E_{base}, such that the lowest energy conformation of both the *ab initio* and analytical model were zero. The algorithms (Figure 6.3 and Figure 6.4) in this chapter fit a constant, c, that matched the overall conformational curves rather than the minima. As such, c differs from $-E_{base}$ by a fraction of a kJ mol^{-1}. The standard error, σ_{intra}, in E_{model} was:

$$\sigma_{intra} = \sqrt{\frac{\sum \left(E_{ab} - E_{model} \right)^2}{\left(N_p - N_k \right)}} \qquad 6.3$$

where N_p and N_k are the number of data points and fitted coefficients respectively. However, there were only N_k -1 physical parameters, as one ($c \sim -E_{base}$) served only to adjust the baseline.

6.2.1. Torsional Potential Model

The explicit torsional term (equation 6.1) can be simplified as the following cosine series:

$$E(\xi) = \sum_{k=0}^{N_k} \alpha_k \cos([k-1]\xi);\ N_k = 5, \cdots, N_p - 1 \qquad 6.4$$

where N_k and N_p are the number of fitted coefficients and data points respectively. The intercept, c, is α_0. Using the linear least squares method (section 2.6), the cosine series (equation 6.4) were fitted to the conformational profiles of FA and TA (Figure 5.11) to determine the unknown coefficients, α_k. Below is a flowchart of the cosine series algorithm (Figure 6.3).

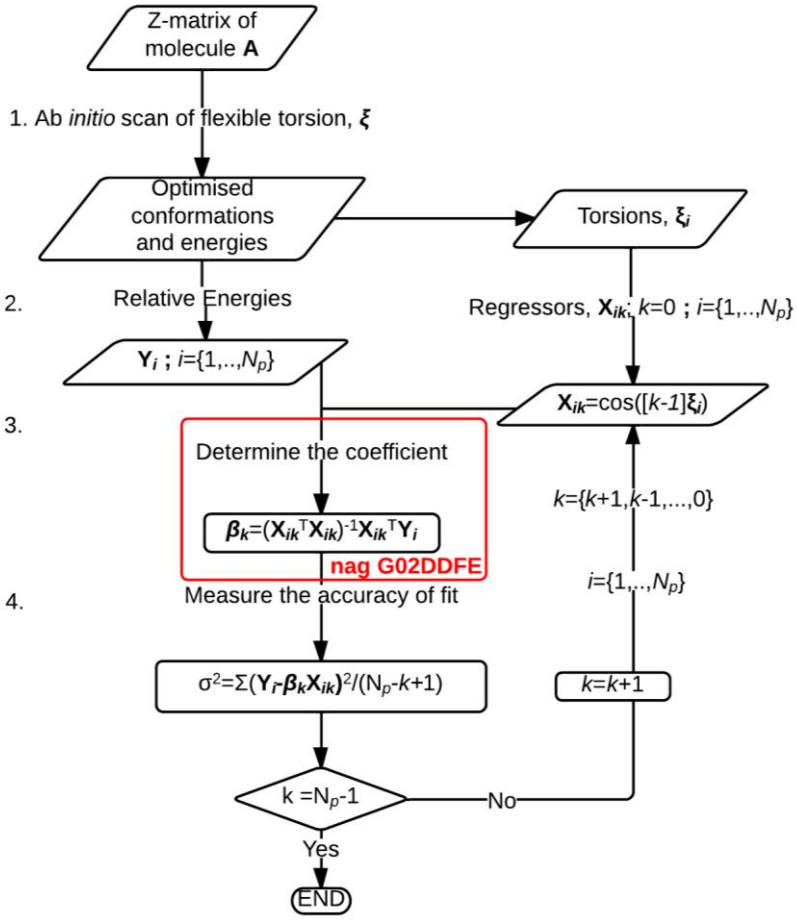


Figure 6.3: An algorithm showing the overview of modelling *ab initio* intramolecular energy using a cosine series potential. The source codes are included in the CD ROM attached to this thesis (see Appendix A).

In the initialization step (**step 1**), the algorithm required the *ab initio* intramolecular energies and the inputs of their equivalent torsion angles, ξ (Figure 5.11). These inputs (generated from the *ab initio* scans reported in Chapter 5) were used to construct a regression model (**step 2**). The G02DDFE[97] NAG subroutine (section 2.6.2) solved the resulting linear regression problem; that is, the unknown coefficients, α_k, were determined. Finally, the standard error, σ_{intra}, of the modelled intramolecular energy, E_{model} (equation 6.3) was calculated.

6.2.2. Atom–Atom Model

Due to the considerable correlation between the atom–atom coefficients, particularly the two repulsion parameters ($A_{\iota\kappa}$ and $B_{\iota\kappa}$), it is difficult to vary both repulsion parameters simultaneously[240,241]. This section discusses the derivation of analytical atom–atom models for conformational profiles of FA and TA, as expressed in the algorithm shown in Figure 6.4. In particular, this section discusses the rescaling of selected repulsion (and dispersion) coefficients, which gives a linear model that contains (an appropriate cosine term and) a baseline constant *c*:

$$E(\xi) = \alpha\cos(2\xi) + \sum_{i,k} \beta_{\iota\kappa} A_{\iota\kappa} \exp(-B_{\iota\kappa} R_{ik}) - \gamma_{\iota\kappa} C_{\iota\kappa} R_{ik}^{-6} + c \qquad 6.5$$

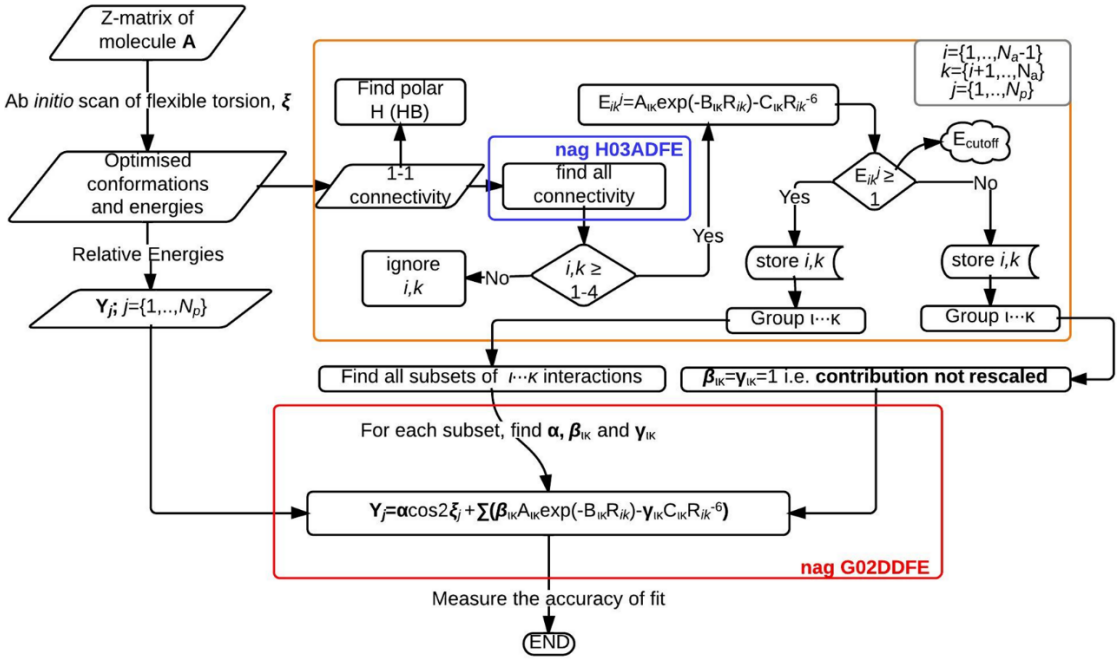


Figure 6.4: An algorithm showing the modelling of *ab initio* intramolecular energy by a physical potential. The source codes are included in the CD ROM attached to this thesis (see Appendix A).

This algorithm (Figure 6.4), which uses NAG[129] library routines, shows the systematic comparison of how well various selections of linear parameters (α, $\beta_{\iota\kappa}$, $\gamma_{\iota\kappa}$, and *c*) can represent the conformational profiles of fenamates. The algorithm first extracted the geometries and the corresponding intramolecular energies of the fenamate molecules, which were generated from *ab initio* calculations. These geometries were converted to an atomic type-friendly mol2 format (Sybyl typing[230]) using the Babel[242] chemical toolbox. The other inputs are the starting values of Gavezzotti's atom–atom *exp*-6 potential parameters[41,231,232] (Table 5.2 of Chapter 5). The covalent 1-1 connectivity (from the mol2 file) helped to distinguish the polar hydrogen, HB, from the non-polar hydrogen, H, atoms. The remaining atom–atom connectivity (i.e. nearest neighbour or bond-path) and

distances, R_{ik}, within the molecule were determined using the H03ADF subroutine and the 1-1 connectivity. H03ADF subroutine evaluates the nearest neighbour by finding the shortest path between two atoms.

Next, the atom-atom interactions were summed over all 1-4 atom-atom distances and higher bond-paths in the entire molecule (equation 6.5). The maximum bond-path was 1-11 for all three fenamates, and most pharmaceutical molecules were sufficiently small that there was no need to define a summation limit in terms of intramolecular distance or bond-path length for a single molecule. The use of 1-4 distances as the shortest intramolecular interactions included in the atom–atom summation was traditional. Although, it should be noted that 1-3 interactions were used elsewhere, e.g. in the CSP code of GRACE where they have sometimes been found to be problematic, including the recent case of a bulky side group attached to an aromatic ring, which required specific scaling down[13]. The atom–atom formulation resulted in many virtually constant terms, such as the H···H, C···C, and H···C contributions from within the aromatic rings. These terms contributed to the baseline energy, E_{base}, defined as the minimum energy found in the scan with a specific parameterized model.

The atom–atom model involved far too many coefficients for all to be fitted independently to the *ab initio* data. In addition, many atom–atom distances did not change significantly with the torsion angle, as observed from investigating the atom–atom interaction that dominated the intermolecular contribution in Chapter 5. There were only 16 and 21 dominant interactions out of 351 and 435 total atom–atom interactions for FA and TA. Therefore, only interactions that were classed as dominant were rescaled. The algorithm in Figure 6.4 determined the dominant interaction by using a cut-off value of one, i.e. E_{cutoff} = 1 kJ mol^{-1}. The dominant atom–atom interactions for FA and TA were reported in Chapter 5 (Table 6.6 and Table 5.7), which are summarised in Table 6.1. If atom–atom interaction of a given type (ι···κ) varied less than the E_{cutoff} (for example Cl···O and O···HB), then these parameters were not rescaled but may vary the contribution to E_{base}. One notable observation was the significant interaction between atoms of the same aromatic ring for TA (Table 6.1[d]), which was absent in FA.

Table 6.1: Contrasting the atom–atom interactions of FA and TA that dominated the intramolecular energies, formed by combining Table 5.6 and Table 5.7 from Chapter 5. Interactions unique to FA and TA are coloured blue and red respectively. [d]Interactions between atoms of the same aromatic ring. [f]Interactions involving the methyl substituent.

at_i···at_k	neigh	FA R_{ik}/kJ mol^{-1}		TA R_{ik}/kJ mol^{-1}		FA $E_{repul-disp}$/kJ mol^{-1}		TA $E_{repul-disp}$/kJ mol^{-1}	
		min	*max(Δ)*	*min*	*max(Δ)*	*min*	*max(Δ)*	*min*	*max(Δ)*
					C···C				
C_7···C_9	4	3.07	3.71(**0.64**)	3.02	3.61(**0.59**)	-0.35	2.45(**2.80**)	-0.28	3.13(**3.40**)
C_7···C_{13}	4	3.07	3.71(**0.64**)	3.14	3.72(**0.58**)	-0.35	2.44(**2.79**)	-0.35	1.65(**2.01**)
C_6···C_9	5	3.10	4.43(**1.33**)	3.08	4.12(**1.04**)	-0.39	2.05(**2.43**)	-0.39	2.31(**2.69**)
C_6···C_8	4			2.89	3.10(**0.21**)			2.11	5.81(**3.70**)
C_6···C_{13}	5	3.11	4.44(**1.33**)	3.24	4.44(**1.19**)	-0.39	2.00(**2.38**)	-0.39	0.85(**1.23**)
[f]C_6···C_{14}	6			3.10	5.10(**2.00**)			-0.39	2.10(**2.48**)
[f]C_7···C_{14}	5			3.12	4.14(**1.02**)			-0.39	1.91(**2.30**)
					C···H				
[f]C_7···H_{12}	6			2.55	3.85(**1.30**)			-0.20	1.73(**1.94**)
C_6···H_7	6	2.52	4.88(**2.36**)	2.48	4.45(**1.97**)	-0.21	2.12(**2.33**)	-0.20	2.56(**2.76**)
C_6···H_8	4	2.89	3.10(**0.20**)			2.12	5.73(**3.62**)		
C_6···H_{11}	6	2.52	4.88(**2.37**)			-0.20	2.11(**2.31**)		
C_8···H_5	5	2.54	2.88(**0.34**)	2.54	2.89(**0.35**)	0.07	1.82(**1.76**)	0.05	1.87(**1.82**)
C_9···H_5	6	2.57	4.16(**1.59**)	2.60	3.69(**1.09**)	-0.20	1.54(**1.74**)	-0.21	1.32(**1.53**)
C_{13}···H_5	6	2.57	4.17(**1.59**)			-0.20	1.52(**1.72**)		
					H···H				
H_5···H_7	7	2.08	4.76(**2.68**)	2.14	4.18(**2.04**)	-0.04	4.43(**4.47**)	-0.04	3.37(**3.42**)
H_5···H_{11}	7	2.09	4.76(**2.67**)			-0.04	4.18(**4.22**)		
					C···HB				
C_{13}···HB_6	4	2.51	3.30(**0.78**)	2.54	3.22(**0.68**)	-0.21	2.16(**2.36**)	-0.20	1.87(**2.07**)
C_9···HB_6	4	2.51	3.30(**0.79**)			-0.21	2.18(**2.38**)		
C_{14}···HB_6	5			2.41	3.71(**1.30**)			-0.21	3.77(**3.97**)
					HB···H				
HB_6···H_{11}	5	2.30	3.68(**1.38**)			-0.04	1.66(**1.70**)		
HB_6···H_7	5	2.30	3.69(**1.39**)			-0.04	1.64(**1.69**)		
[f]HB_6···H_{11}	6			2.19	4.50(**2.31**)			-0.04	2.66(**2.71**)
[f]HB_6···H_{12}	6			2.28	3.15(**0.87**)			-0.03	1.77(**1.80**)
					N···C				
[f]N_1···C_{14}	4			2.78	2.98(**0.21**)			0.93	4.74(**3.82**)
					N···H				
[f]N_1···H_{12}	5			2.43	2.62(**0.19**)			0.08	1.42(**1.34**)
					C···Cl				
[f]C_{14}···Cl_1	4			3.04	3.11(**0.07**)			2.61	3.87(**1.26**)
					Cl···H				
[df]Cl_1···H_{10}	5			2.57	3.24(**0.68**)			-0.49	4.04(**4.53**)
[df]Cl_1···H_{11}	5			2.69	3.74(**1.05**)			-0.50	1.81(**2.31**)

These atom–atom interactions were grouped into unique atom-type···atom-type sets (ι···κ). The algorithm then systematically fit different subsets of the ι···κ rescaling parameters using combination mathematics[243]. E.g., for interactions C···H, C···HB, and H···HB, the following seven subsets of interactions were rescaled:

1. C···H *or* C···HB *or* H···HB
2. C···H and C···HB *or* C···H and H···HB *or* C···HB and H···HB
3. C···H, C···HB and H···HB

For each fit, the subroutine G02DDFE[97] (section 2.6.2) calculated the N_k coefficients (α, $\beta_{\iota\kappa}$, $\gamma_{\iota\kappa}$, c) being fitted, and minimized the residual sum of squares (RSS) using a general linear regression model[244]. Within the routine, a matrix decomposition process established whether a solution was unique, or whether the parameters were linearly dependent and therefore had no unique solution. Unless otherwise noted, all fits reported in this chapter had a unique solution. Finally, the standard error, σ_{intra}, in the model's intramolecular energy was calculated after adjusting the minimum to 0 kJ mol^{-1} (equation 6.3).

This chapter used the method discussed above to investigate two analytical models:

(1) Rescaling only the *repulsion* contribution of the atom–atom *exp*-6 potential, i.e. equation 6.5 with α and γ set at 0 and 1 respectively:

$$E(\xi) = \sum_{i,k} \beta_{\iota\kappa} A_{\iota\kappa} \exp(-B_{\iota\kappa} R_{ik}) - C_{\iota\kappa} R_{ik}^{-6} + c \qquad 6.6$$

(2) Rescaling all the parameters i.e. equation 6.5.

6.3. Results

6.3.1. Cosine Series Model

The traditional cosine series expansion for a torsional potential (equation 6.4) must include cos4ξ to have the correct number of minima for the fenamates. Alone, this term gave a poor position of the minima for FA, and it was qualitatively wrong for TA (Figure 6.5). A least squares fit that included the lower cosine terms (N_k=5, equation 6.4) gave a qualitatively reasonable representation (Figure 6.5), but further improvement is slowly converging (Figure 6.6). For FA, the fitting steadily improved in quality, particularly upon the addition of cos4$n\xi$ terms, where n are integers. For TA, the fitting became unstable after 21 terms even though there was no attempt to represent the large steric clash beyond 10 kJ mol^{-1}.

This result demonstrated that the cosine series was a fitting exercise and did not reflect the physics. It was effectively an exercise in modelling the relaxed scans by a functional

form that assumed the scan was rigid (i.e. only the torsion angle changed). This assumption is a crude approximation since rigid scans are different from relaxed scans (see Figure 5.13). Poor convergence of the cosine model expansion were also reported for biphenyl[245], a dye which required 7 terms, while polynorbornene[246] required 6 and 15 terms for the meso and racemic dimer respectively.

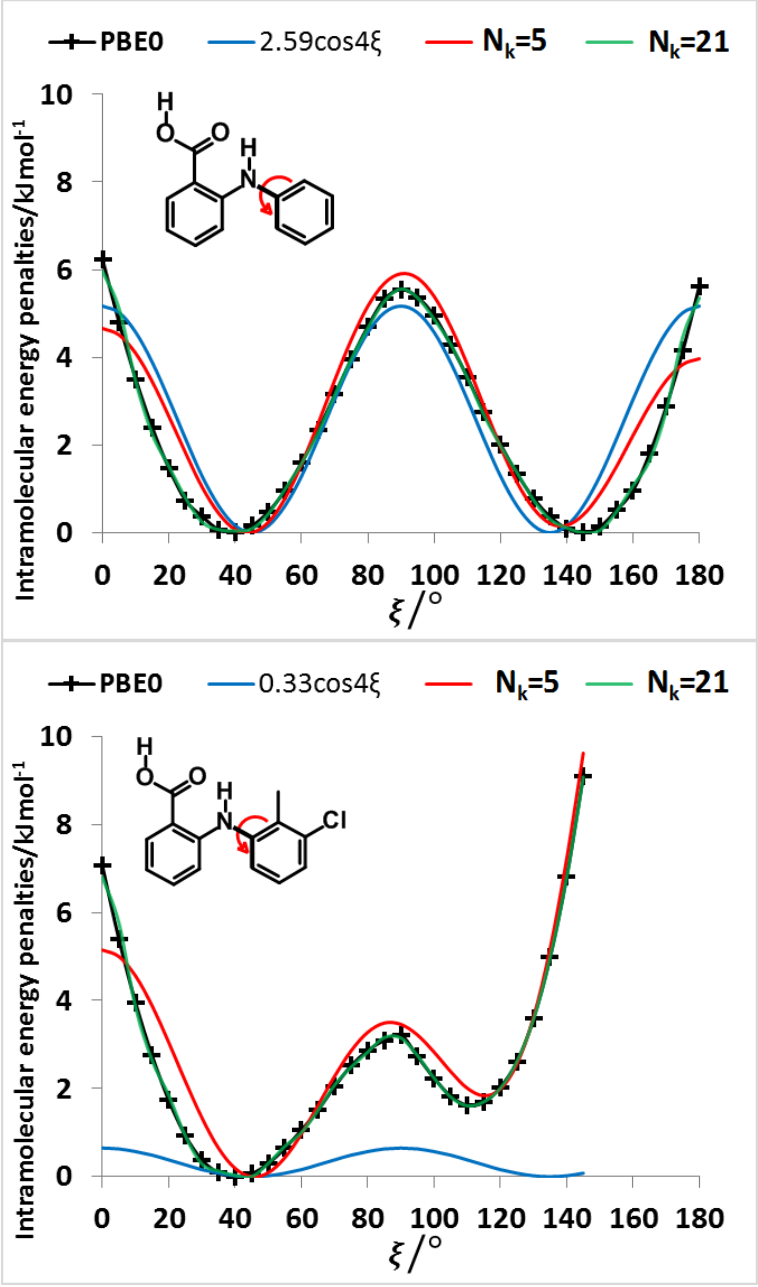


Figure 6.5: Comparison of the *ab initio* relaxed scan at the PBE0/6-31+G(d) level of theory with linearly fitted least square cosine series model for FA (top) and TA (bottom), with N_k=5 (red lines), the optimal cos4ξ terms (blue lines), and N_k=21 (green lines, σ_{intra} =0.17 kJ mol^{-1} for FA and TA).

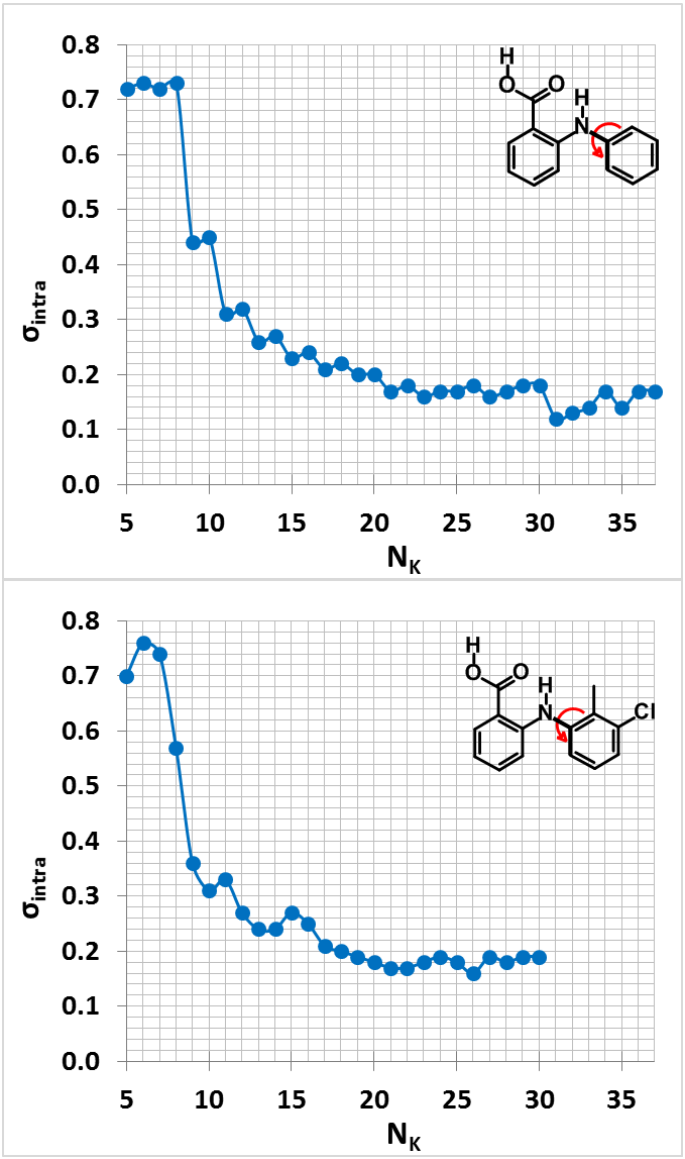


Figure 6.6: Plot of the standard error of the intramolecular energies against the number of cosine terms, N_k, for FA (top) and TA (bottom). The fit is *ill-determined*, i.e. there is no unique solution, for TA after N_k=21.

6.3.2. Rescaled Repulsion Model

From Chapter 5, it is clear that the repulsion contributions were severely overestimated for FA and TA with Gavezzotti's *exp*-6 potential. Therefore, an alternative model is to rescale just the repulsion terms (i.e. fit selected β_{ik} in equation 6.6) as the intermolecular atom–atom model clearly overestimated the intramolecular steric effects (Figure 5.19). Rescaling just a few of Gavezzotti's repulsion parameters, for example those involving the poorly parameterized H···H interactions, resulted in a qualitatively good fit (Figure 6.7), as shown by the results for fitting all combinations of repulsion parameters for FA (Table 6.2**)** and the summary in Table 6.3 of the **511** such fits for TA. Rescaling just two (C···H and H···H) repulsion contributions for FA and five (C···H, C···HB, C···N, H···H and H···HB) for TA produced a model that reproduced the torsional profile well (Figure 6.7).

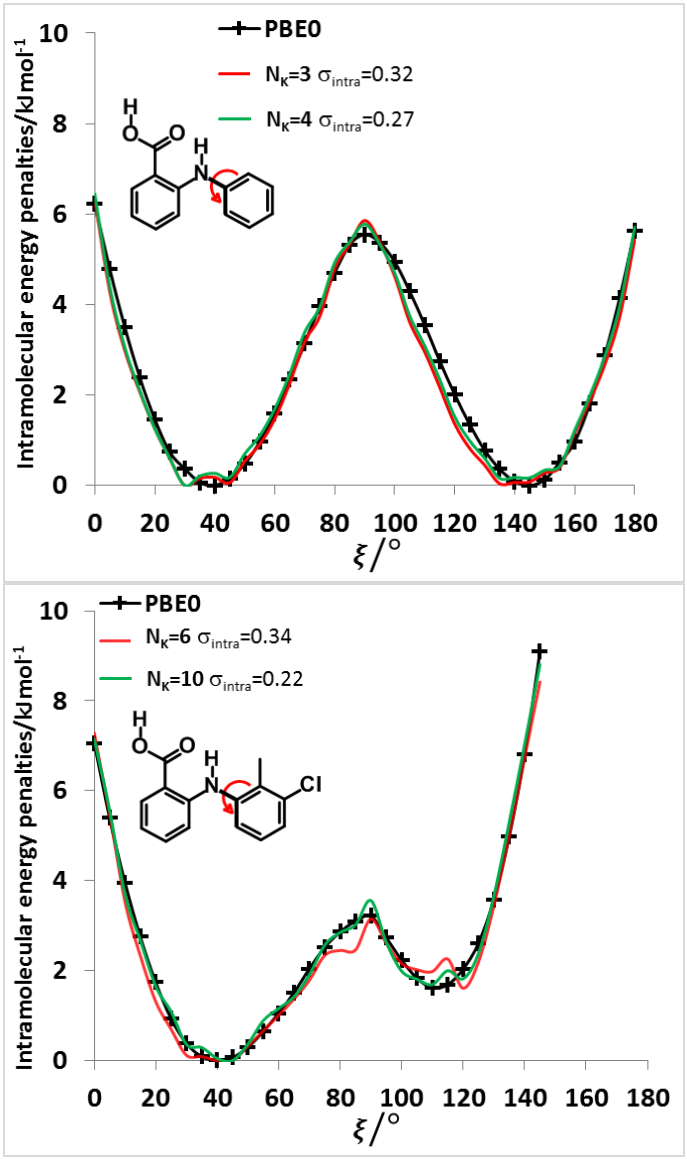


Figure 6.7: Comparison of the relative energies from *ab initio* calculations (solid black lines) of FA (top) and TA (bottom) with selected best-fit *exp*-6 rescaled repulsion models. For FA, the selected N_k=3 σ_{intra}=0.32 kJ mol^{-1} model had $\beta_{C\text{-}H}$=-4.53, $\beta_{H\text{-}H}$=2.70, and E_{base}=-5.06, while the N_k =4 σ_{intra}=0.27 kJ mol^{-1} model had β_{CH}=-4.46, $\beta_{H\text{-}H}$=2.10, $\beta_{H\text{-}HB}$=2.57, and E_{base}=-8.75. For TA, the selected N_k=6 (σ_{intra}=0.34 kJ mol^{-1}) and 10 (σ_{intra}=0.22 kJ mol^{-1}) parameter fits are shown bold in Table 6.3.

The rescaling of the C···N interaction seemed particularly critical for TA (Table 6.3), which can be attributed to the importance of the C(methyl)···N interaction in describing the steric clash involving the methyl group (Figure 5.14). However, many of the fitted rescaling parameters were negative ($\beta_{i\kappa}$ <0), which suggested that the exponential steric repulsion had become attractive. Thus, again, this appeared to be a fitting exercise, so that exponential terms were reducing the over-estimate of the barrier at 90° by the intermolecular dispersion parameters (Figure 5.19).

Table 6.2: Complete set of results after rescaling N_k (repulsion contributions, $\beta_{\iota\kappa}$, and baseline energy, E_{base}) parameters of various interacting atomic types for FA.

	$\beta_{\iota\kappa}$					E_{base}	σ_{intra}
N_k	C -C	C -H	C -HB	H -H	H -HB	kJ mol^{-1}	
2	-	-2.26	-	-	-	4.37	1.68
2	-	-	-1.48	-	-	10.76	2.84
2	-	-	-	-0.88	-	22.15	2.87
2	-	-	-	-	-4.23	35.79	2.99
2	-3.32	-	-	-	-	-399.48	4.51
3	-	-4.53	-	2.69	-	-5.06	0.32
3	-	-4.28	-	-	5.39	-15.11	0.32
3	-2.10	-2.17	-	-	-	-310.11	0.36
3	-	-5.23	3.65	-	-	2.78	0.4
3	-5.23	-	-1.73	-	-	-626.22	0.45
3	-6.21	-	-	-1.24	-	-713.56	0.65
3	-6.66	-	-	-	-5.49	-743.17	0.66
3	-	-	-9.91	-	20.82	-75.37	1.27
3	-	-	-9.09	7.22	-	-22.78	1.74
3	-	-	-	-7.38	19.72	-25.77	2.47
4	-	-4.46	-	2.10	2.57	-8.75	0.27
4	-	-4.67	1.44	2.43	-	-3.88	0.3
4	0.79	-4.38	-	2.58	-	-25.57	0.32
4	0.41	-3.90	-	-	4.59	-71.01	0.32
4	-	-4.32	1.09	-	5.25	-14.56	0.33
4	-1.73	-2.54	1.33	-	-	-272.52	0.37
4	-4.58	-	-2.96	-	3.94	-572.20	0.4
4	-5.07	-	-2.18	1.37	-	-611.33	0.45
4	-6.39	-	-	-0.35	-1.57	-725.71	0.64
4	-	-	-9.58	0.06	22.75	-78.93	1.25
5	1.30	-4.67	-	2.18	2.78	20.69	0.28
5	-	-4.48	1.04	2.10	2.51	-8.49	0.28
5	2.36	-5.91	2.18	2.70	-	132.31	0.29
5	-0.16	-3.27	0.37	-	4.81	-129.63	0.31
5	-4.56	-	-2.84	0.61	4.79	-572.09	0.39
6	2.08	-5.50	1.70	2.37	2.25	100.68	0.28

Table 6.3: Summary table of the best results after rescaling N_k (repulsion contributions, $\beta_{\iota\kappa}$, and baseline energy, E_{base}) parameters of various interacting atomic types for TA.

	$\beta_{\iota\kappa}$									E_{base}	σ_{intra}
N_k	C-C	C-Cl	C-H	C-HB	Cl-H	C-N	H-H	H-HB	H-N	kJ mol^{-1}	
2	-	9.09	-	-	-	-	-	-	-	112.81	1.30
3	-0.08	-	-	-	-	-2.99	-	-	-	-135.47	0.93
4	-	-	-	-	7.51	-2.69	-	-	14.54	117.47	0.87
5	-	-	-2.79	6.22	-	-4.64	-	-5.49	-	-54.29	0.51
6	**-**	**-**	**-3.06**	**4.72**	**-**	**-4.47**	**4.26**	**-5.29**	**-**	**-52.79**	**0.34**
7	-	-	-3.92	4.72	-	-4.87	5.32	-6.11	-0.54	-76.90	0.25
8	-	5.03	-3.59	3.87	2.65	-3.40	5.81	-5.70	-	3.00	0.22
9	-	5.31	-3.43	3.55	3.18	-3.27	5.96	-5.39	1.90	17.48	0.21
10	**0.84**	**5.41**	**-3.19**	**3.30**	**3.47**	**-3.31**	**6.11**	**-5.27**	**2.19**	**6.81**	**0.22**

6.3.3. Combined Physical Model

Assuming a model that both describes the electronic effects and allows rescaling of the atom–atom interactions:

$$E(\xi) = \alpha\cos(2\xi) + \sum_{i,k} \beta_{\iota\kappa} A_{\iota\kappa} \exp(-B_{\iota\kappa} R_{ik}) - \gamma_{\iota\kappa} C_{\iota\kappa} R_{ik}^{-6} \qquad 6.7$$

Then there is a huge number of ways for finding a satisfactory fitting of the data. The rescaling of the three contributions in equation 6.7 is physically justified since:

(1) The α term represents the electronic barrier that arises from the *conjugation* of benzoic acid rings (section 5.3.5).

(2) The β rescale term is necessary due to the overestimation of the *repulsion* contribution (section 5.3.6).

(3) The γ rescale term accounts for electron correlation i.e. *dispersion*. Section 2.7.5 emphasised the importance of the *dispersion* term, while section 5.3.3 showed the consequence of ignoring the electron correlation e.g. HF method.

Cases where β and $\gamma > 0$ retained the repulsive and attractive (dispersion) nature i.e. repulsion did not become attraction and vice versa. It was also required that α was less than zero ($\alpha < 0$) to give a maximum around ξ=90°, corresponding to conjugation (Figure 5.15). Table 6.4 contains all the fits to FA, which have the physically correct sign for all parameters and the full set of results for fitting up to 12 parameters (including the baseline estimate, E_{base}, as given in Table 6.5). This shows that fitting α and only rescaling a few atom–atom interactions gives a qualitatively accurate fit (Figure 6.9). It is notable that the parameters involving H and HB usually needed the greatest rescaling, a finding that is consistent with the fact that HB parameters were not fitted in the intermolecular potential model. In addition, the intermolecular parameterization of H···H is a notable exception in that its minimum is very shallow (Figure 6.8) and is not sharply defined, being repulsive for a wide range of intermolecular H···H distances[42]. Furthermore, there is a huge difference between other polar H (HB) interactions and H···H interactions; therefore, it is not surprising that the interactions that involve HB atoms required the greatest rescaling.

Table 6.4: All fits obtained for FA with the physically correct sign for the repulsion, $\beta_{\iota\kappa}$, and dispersion, $\gamma_{\iota\kappa}$, and electronic α parameters.

	$\beta_{\iota\kappa}$					$\gamma_{\iota\kappa}$					α	E_{base}	σ_{intra}
N_k	C-C	C-H	C-HB	H-H	H-HB	C-C	C-H	C-HB	H-H	H-HB	kJ mol^{-1}		
6	-	-	8.01	10.74	-	-	-	16.54	58.00	-	-3.39	-123.83	0.26
6	-	0.28	-	-	11.09	-	7.83	-	-	34.79	-1.17	-75.49	0.28
4	-	-	-	11.08	-	-	-	-	50.03	-	-3.91	-59.26	0.33
4	-	6.19	-	-	-	-	20.82	-	-	-	-1.57	-180.49	0.36
4	-	-	-	-	34.67	-	-	-	-	136.98	-5.30	-12.46	0.57
4	-	-	36.30	-	-	-	-	110.38	-	-	-9.45	-334.66	0.76

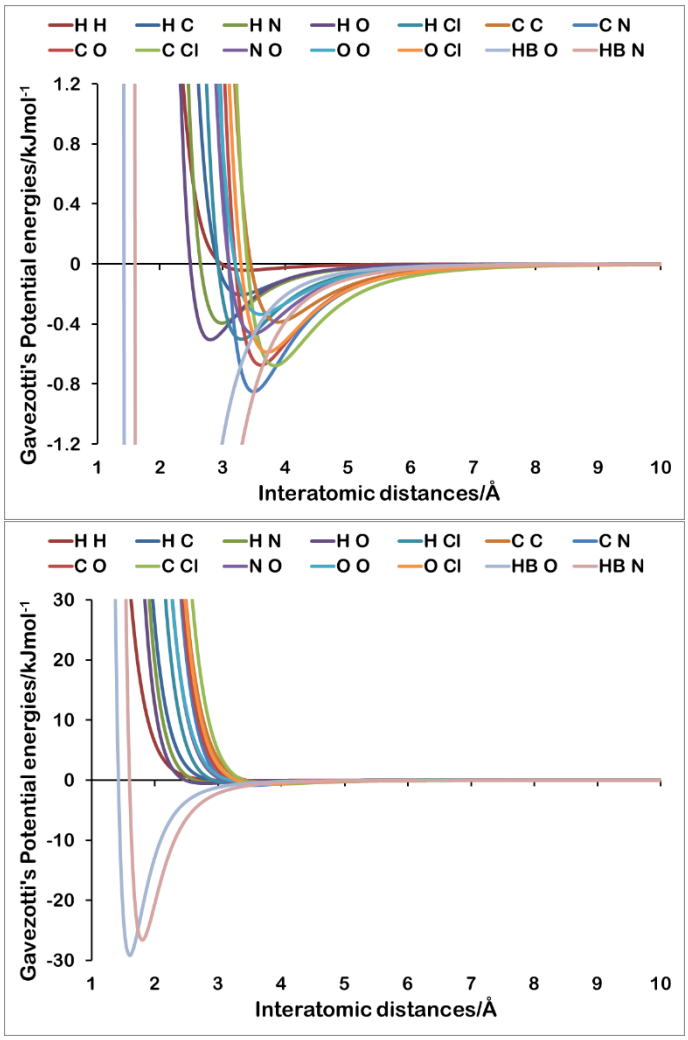


Figure 6.8: Gavezzotti *exp*-6 potential for all atom–atom interactions in fenamates, displayed in two different energy scales. The parameters are from Chapter 5 (Table 5.2)

Although virtually perfect fits can be obtained (Table 6.5), the variation in the fitted parameters is significant given the exponential sensitivity of the repulsion to changes in atom–atom distances. These could be substantial: for example, the lack of symmetry for the FA torsional scan where the two minima were found at ξ=38.94° and 144.71°, though only differing in energy by 1.88 x 10^{-3} kJ mol^{-1} (7.16 x 10^{-7} au), corresponded to a change of 0.61 Å in the 1-4 interaction between the amide proton and C_{13} or C_9. Even the C_8-C_9 and C_8-C_{13} aromatic bond lengths differed by ±0.0037 Å for ξ=0 or 180°, but only by ±0.0004 Å for ξ=80° or 100°, with larger changes in the bonded hydrogen positions for FA. Hence, the systematic consideration of all combinations of sets of parameters that could be meaningfully fitted, and the resulting coefficients, only provided a qualitative guide to which of the atom–atom interactions dominated the intramolecular energy (Chapter 5, Table 6.6) and were most critical.

Table 6.5: The rescale parameters for the repulsion, β_{ik}, dispersion, γ_{ik}, and electronic, α, contributions of the physical *exp*-6 and cos2ξ model for all possible combinations of atom type interactions C···C, C···H, C···HB, H···H and H···HB for FA. N_k are the number of fitted rescale parameters, β_{ik} and γ_{ik}, and the baseline energy, E_{base}.

	β_{ik}					γ_{ik}					α	E_{base}	σ_{intra}
N_k	C-C	C-H	C-HB	H-H	H-HB	C-C	C-H	C-HB	H-H	H-HB		kJ mol^{-1}	
4	-	-	-	11.08	-	-	-	-	50.03	-	-3.91	-59.26	0.33
4	-	6.19	-	-	-	-	20.82	-	-	-	-1.57	-180.49	0.36
4	-7.97	-	-	-	-	-15.08	-	-	-	-	-2.72	-105.30	0.45
4	-	-	-	-	34.67	-	-	-	-	136.98	-5.30	-12.46	0.57
4	-	-	36.30	-	-	-	-	110.38	-	-	-9.45	-334.66	0.76
6	-3.66	4.61	-	-	-	-10.53	14.77	-	-	-	-0.80	-37.68	0.10
6	-4.36	-	-	6.10	-	-7.64	-	-	29.66	-	-1.14	-158.33	0.12
6	-8.52	-	-1.88	-	-	-12.27	-	-2.29	-	-	0.55	-306.21	0.17
6	-6.00	-	-	-	11.35	-8.62	-	-	-	54.15	-1.43	-235.13	0.18
6	-	-1.03	14.76	-	-	-	5.74	40.73	-	-	-2.20	-173.85	0.23
6	-	-	-	12.08	-1.82	-	-	-	65.37	-29.18	-2.08	-95.76	0.24
6	-	-2.55	-	6.19	-	-	0.13	-	20.48	-	-2.16	-15.13	0.25
6	-	-	8.01	10.74	-	-	-	16.54	58.00	-	-3.39	-123.83	0.26
6	-	0.28	-	-	11.09	-	7.83	-	-	34.79	-1.17	-75.49	0.28
6	-	-	-13.26	-	65.19	-	-	-69.60	-	328.79	-5.39	244.22	0.36
8	-2.75	3.51	-	-	2.71	-9.35	13.44	-	-	4.12	-0.75	5.34	0.08
8	-2.36	4.14	-	-1.57	-	-10.78	18.76	-	-17.19	-	-0.76	93.74	0.09
8	-3.73	-	-	6.26	3.34	-6.81	-	-	34.75	2.22	-0.66	-152.08	0.09
8	-4.93	5.51	-1.89	-	-	-12.18	15.30	-6.42	-	-	-0.47	-64.99	0.10
8	-5.30	-	-7.37	7.25	-	-10.54	-	-24.17	32.67	-	-0.44	-33.45	0.10
8	-6.98	-	-16.55	-	30.47	-12.63	-	-60.95	-	147.74	0.12	58.42	0.13
8	-	-4.19	5.20	11.95	-	-	-8.46	6.91	60.39	-	-2.46	-9.95	0.19
8	-	-	18.43	11.85	-16.77	-	-	58.81	72.51	-111.25	-1.96	-303.53	0.19
8	-	-0.89	26.97	-	-31.96	-	9.52	96.22	-	-180.54	-0.82	-404.33	0.21
8	-	-4.45	-	15.41	-10.16	-	-9.70	-	77.19	-63.85	-1.66	-7.70	0.21
10	-4.45	6.86	-5.55	-5.41	-	-15.77	26.64	-14.89	-41.79	-	0.17	137.52	0.05
10	-5.24	4.69	-7.40	-	9.53	-12.47	12.57	-24.81	-	38.80	-0.22	-0.75	0.07
10	-3.20	4.65	-	-0.26	5.93	-9.77	15.32	-	-3.49	17.63	-0.64	-34.03	0.08
10	-5.98	-	-14.29	4.55	18.04	-11.88	-	-51.12	18.25	86.95	-0.25	69.86	0.09
10	-	1.66	27.17	6.71	-29.44	-	7.88	93.84	43.50	-178.54	-1.13	-447.69	0.14
12	-3.83	7.09	-1.55	-5.70	-4.47	-14.98	28.35	-0.35	-42.60	-29.03	0.18	97.08	0.06

Table 6.6: The rescale parameters for the repulsion, β_{ik}, dispersion, γ_{ik}, and electronic, α, contributions of the physical *exp*-6 and cos2ξ model for all possible combinations of atom type interactions C···C, C···Cl, C···H, C···HB, Cl···H, C···N, H···H and H···HB for TA. N_k are the number of fitted rescale parameters and E_{base} is the baseline energy. [g,h]The parameters used in Figure 6.9b and [i]for testing transferability in Figure 6.10b and Figure 6.11.

	β_{ik}									γ_{ik}									α	E_{base}	σ_{intra}
N_k	C-C	C-Cl	C-H	C-HB	Cl-H	C-N	H-H	H-HB	H-N	C-C	C-Cl	C-H	C-HB	Cl-H	C-N	H-H	H-HB	H-N		kJ mol^{-1}	
4	-	10.19	-	-	-	-	-	-	-	-	5.93	-	-	-	-	-	-	-	-1.35	86.90	1.08
4	-	-	-	-	-	9.30	-	-	-	-	-	-	-	-	29.01	-	-	-	-5.30	-248.91	1.11
4	-	-	-	-	-	-	-	-	18.39	-	-	-	-	-	-	-	-	51.01	-2.93	-141.15	1.49
4	-	-	-	-	-	-	-	8.61	-	-	-	-	-	-	-	-	38.84	-	-2.54	20.58	1.65
4	-	-	-	3.03	-	-	-	-	-	-	-	-	11.02	-	-	-	-	-	-1.67	-11.33	1.68
4	1.81	-	-	-	-	-	-	-	-	0.95	-	-	-	-	-	-	-	-	-7.61	132.28	1.93
6	-	-	-	-	-	22.37	1.73	-	-	-	-	-	-	-	65.17	18.10	-	-	-1.35	-636.04	0.49
6	-	4.84	-	-	-	19.74	-	-	-	-	41.73	-	-	-	71.86	-	-	-	-8.48	-994.38	0.67
6	-	-	15.60	-	-	-	18.56	-	-	-	-	8.17	-	-	-	119.67	-	-	-13.11	-86.78	0.77
6	-	-	-	-	9.67	21.71	-	-	-	-	-	-	-	44.31	69.77	-	-	-	-1.67	-891.72	0.78
6	6.59	-	-	-	-	44.09	-	-	-	9.55	-	-	-	-	123.97	-	-	-	-16.02	-993.38	1.02
6	-	-	-	-	-	-	6.04	-	22.88	-	-	-	-	-	-	20.12	-	75.06	-6.74	-276.19	1.20
[g]8	-	19.33	-	5.67	-	21.65	-	-	-	-	72.89	-	22.07	-	69.76	-	-	-	-1.02	-1141.80	0.15
8	-	-	-	-	19.57	30.68	-	4.98	-	-	-	-	-	81.30	96.34	-	26.47	-	-0.96	-1337.67	0.16
8	-	16.18	-	-	-	28.74	-	6.76	-	-	62.63	-	-	-	89.52	-	37.54	-	-3.61	-1196.16	0.23
8	-	0.38	-	-	-	24.51	1.33	-	-	-	11.39	-	-	-	75.74	15.28	-	-	-3.20	-836.84	0.28
8	-	-	10.15	8.01	-	-	5.05	-	-	-	-	0.83	4.37	-	-	111.23	-	-	-9.62	-41.65	0.37
8	-	-	-	-	6.34	26.50	0.27	-	-	-	-	-	-	20.75	79.27	11.06	-	-	-3.13	-862.12	0.41
8	-	-	-	-	-	22.11	1.27	3.67	-	-	-	-	-	-	63.26	18.98	12.48	-	-0.09	-623.50	0.47
8	-	-	1.56	-	-	23.56	1.05	-	-	-	-	1.89	-	-	68.73	15.61	-	-	-1.78	-675.59	0.51
8	-	-	14.22	-	-	-	11.94	12.25	-	-	-	11.18	-	-	-	86.25	53.31	-	-6.21	-143.14	0.63
8	-	-	7.68	-	-	-	14.97	-	13.53	-	-	2.21	-	-	-	81.54	-	27.30	-4.36	-93.01	0.71
[h]10	-	10.49	-	-	26.00	27.31	-	6.79	-	-	3.66	-	-	91.02	81.40	-	38.89	-	-4.26	-1122.47	0.08
10	-	-	1.35	-	8.53	21.73	-	4.26	-	-	-	4.04	-	27.24	68.67	-	24.76	-	-1.93	-837.10	0.09
10	-	8.53	-	-	-	23.86	2.27	5.83	-	-	37.42	-	-	-	74.31	12.38	29.41	-	-2.22	-946.94	0.10
10	-	-	-	-	14.65	27.85	0.89	5.42	-	-	-	-	-	57.88	85.69	7.20	25.68	-	-1.07	-1121.06	0.11
10	-	11.43	1.94	4.57	-	20.86	-	-	-	-	48.33	3.41	18.29	-	67.62	-	-	-	-0.96	-1026.46	0.16
10	0.50	-	-	-	17.63	26.92	-	4.96	-	0.85	-	-	-	70.84	86.15	-	27.44	-	-0.32	-1234.24	0.18
[i]10	**-**	**-**	**1.80**	**-**	**-**	**20.79**	**-**	**6.04**	**0.89**	**-**	**-**	**6.27**	**-**	**-**	**61.76**	**-**	**35.44**	**8.83**	**-0.51**	-719.71	**0.26**
10	-	-	4.13	5.06	-	13.34	0.18	-	-	-	-	4.73	11.46	-	40.10	25.32	-	-	-0.67	-483.08	0.31
10	3.14	-	9.45	11.22	-	-	5.35	-	-	5.31	-	1.10	18.33	-	-	100.29	-	-	-9.66	-95.28	0.32
10	-	11.60	7.43	9.25	-	-	3.65	-	-	-	27.57	0.57	15.32	-	-	80.17	-	-	-4.87	-171.28	0.36
12	-	-	-	0.93	13.21	28.31	3.14	-	15.19	-	-	-	5.69	41.86	82.38	9.41	-	38.12	-4.14	-1109.66	0.07
12	-	7.09	-	2.53	17.21	26.53	-	-	9.03	-	11.98	-	10.91	55.30	79.56	-	-	17.92	-3.50	-1111.02	0.08
12	-	-	2.37	1.37	10.32	23.50	-	-	7.00	-	-	4.32	7.02	27.19	72.55	-	-	10.37	-3.51	-885.68	0.08
12	-	-	4.25	-	8.43	22.74	0.65	6.56	-	-	-	5.88	-	24.15	69.38	13.56	31.01	-	-4.40	-840.12	0.08
12	-	-	0.46	-	-	20.81	5.36	2.74	13.91	-	-	3.44	-	-	57.22	13.60	23.19	50.07	-2.45	-780.93	0.27
14	-	-	3.84	2.14	8.86	20.10	1.90	-	8.05	-	-	4.56	6.90	20.56	61.36	16.35	-	14.62	-5.32	-764.72	0.08
14	-	-	2.49	0.12	9.54	24.42	-	3.16	4.10	-	-	4.57	2.14	24.39	75.63	-	12.79	2.09	-3.60	-861.83	0.09

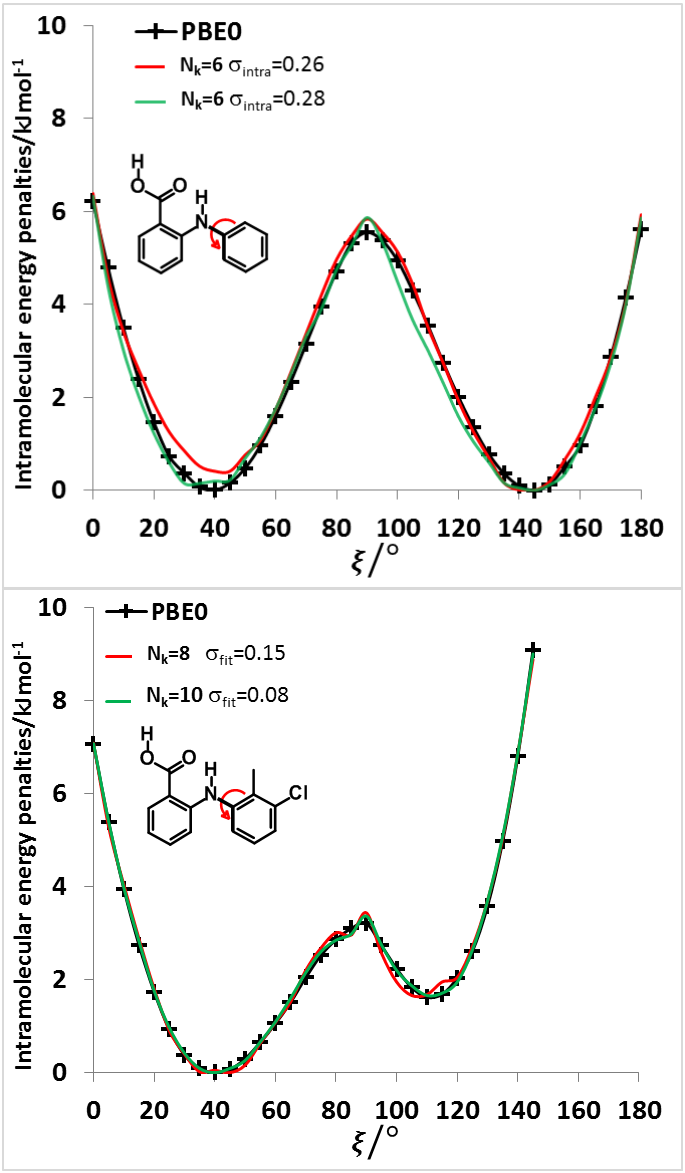


Figure 6.9: Comparison of the relative energies from *ab initio* calculations (solid black lines) of FA (top) and TA (bottom) with selected physical models. For FA, the selected N_k=6 σ_{intra}=0.26 kJ mol^{-1} model has $\beta_{C\text{-}HB}$ =8.01, $\gamma_{C\text{-}HB}$ =16.54, $\beta_{H\text{-}H}$ =10.74, $\gamma_{H\text{-}H}$ =58, α=-3.39, and E_{base}=-123.83, whilst the N_k =6 σ_{intra}=0.28 kJ mol^{-1} model has $\beta_{C\text{-}H}$ =0.28, $\gamma_{C\text{-}H}$ =7.83, $\beta_{H\text{-}HB}$ = 11.09, $\gamma_{H\text{-}HB}$ =34.79, α=-1.17, and E_{base}=-75.49. For TA, parameters for the selected N_k=8 (σ_{intra}=0.15 kJ mol^{-1}) and 10 (σ_{intra}=0.08 kJ mol^{-1}) are in [g,h]Table 6.6.

In contrast to FA, TA (Table 6.6) gave a qualitatively accurate fit only when at least three types of atom–atom parameters were rescaled (N_k=8), including C···N, as shown in Figure 6.9. The relaxed scans were sensitive to a large number of atom–atom distances, so that the 5° change in ξ did not necessarily give a smooth curve between the analytical values. This was because of the differences in the molecular relaxation and resulting atom–atom distances, not the step size, because Figure 5.10 of Chapter 5 showed that smaller step size for FA produced a smooth but asymmetric *ab initio* energy-curve. All the fits with the physically correct parameters were *unique* solutions (section 2.6.2).

Depending on which atomic types were being fitted, there were some variations in the parameters for a given atom–atom interaction. I noticed that for TA (Table 6.6), the C···N parameters consistently needed to be drastically rescaled, reflecting the importance of the methyl carbon (R_1=CH_3, Figure 6.1) in the conformational profile of TA. Indeed, TA differed from FA in having many more intramolecular atom–atom distances that changed significantly within the range ξ= 0 to 145°, including some within the same aromatic ring, such as the methyl–chloro interactions (Table 6.1 *or compare* Table 5.6 with Table 5.7 in Chapter 5). Day[6] also observed that pyramidalizations of atoms (e.g. the nitrogen atom in an amine group) significantly affect the intramolecular parameters.

By following the procedure described in the methodology (section 6.2.2), I obtained a variety of analytical models of the form of equation 6.7 that could reproduce the torsional profiles of FA and TA with a high degree of accuracy, despite the significant variation in many atom–atom distances during the relaxed torsional scans. However, the analytical model had a maximum number of parameters for FA and TA (12 and 22, respectively), with 5 (plus 1) atomic types: C, H, N, O, HB (plus Cl), when considering atomic-type···atomic-type interaction that contributed significantly to energy changes with conformation. Fitting too many parameters led to models where some of the fitted parameters had the wrong sign for their physical interpretation, and some of these fits for TA showed a linear dependence between the parameters (i.e. the fitted parameters were not unique, section 2.6.2). Nonetheless, the ease with which this *physically justified model* (equation 6.7) could reproduce the conformational dependence of FA and TA shows that this approach to modelling the intramolecular energies of fenamates is promising.

6.3.4. Transferability

A further investigation of the physical reasonableness of the atom–atom plus electronic functional form (equation 6.7) is to test whether the models transfer to related molecules. The difference between the energy scans when the substituents are far from the varying torsion angle, for example, the change of H to Cl atom i.e. from FA to ClFA or between TA and MA (Figure 5.11), should not change the main steric interactions, and should only have a small effect on the electronic term (Figure 5.15e and f).

Transferring a set of parameters fitted to FA, and adding Gavezzotti's Cl parameters did indeed (Figure 6.10) give some of the asymmetry in the well depths seen for ClFA and the higher central barrier, as compared to FA (Figure 5.11), when the electronic term, α, was fitted. Similarly, a set of parameters fitted to TA could reproduce the lower barrier at ξ = 90° and higher barrier at ξ = 0° in MA (Figure 6.10b), despite the fact that the

conformational relaxation of the methyl–methyl interaction was somewhat different from that of the methyl-chloro geometry (**Figure 5.13**).

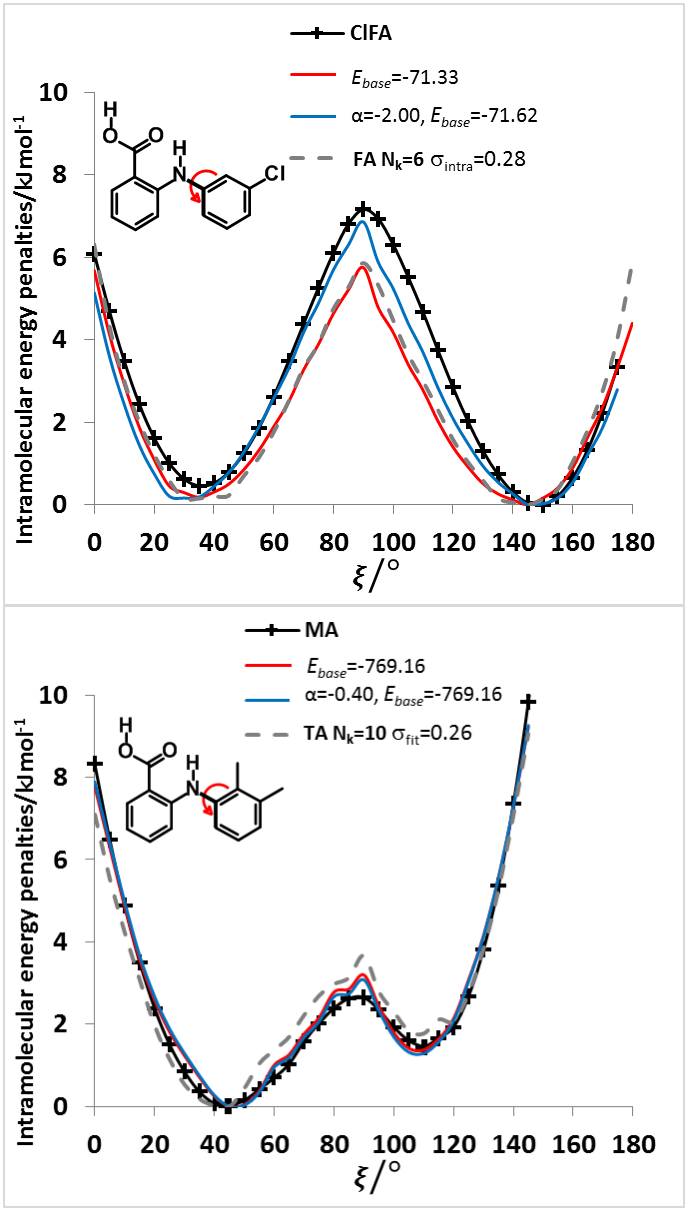


Figure 6.10: Comparison of the relative energies from *ab initio* calculations (solid black lines) with models using transferred β_{tk} and γ_{tk} parameters from FA and TA (dotted grey plots) to ClFA (top) and MA (bottom) respectively. The red curve has only had the baseline adjusted, whereas the blue curve had the α parameter refitted to ClFA or MA respectively. The transferred FA parameters are those N_k=6, σ_{fit}=0.28 kJ mol^{-1} in Figure 6.9, while those of TA N_k= 10 σ_{fit}= =0.26 kJ mol^{-1} are shown in *i*Table 6.6.

A more stringent test would be to apply the same TA parameters to FA and ClFA; which gave qualitatively correct results around the minima (Figure 6.11) and the steric barriers at 0 and 180°. However, the barrier height around 90° would remain similar to that in TA, and this underestimation would be only partially corrected by refitting the αcos(2ξ) term. Given the limitations in the parameter optimization, the qualitative

accuracy is promising. In particular, the poorer transferability may be attributable to the fact that the fitting of parameters for C and H interactions in TA was modelling both the aromatic and methyl atoms, and the methyl···N interaction was critical in fitting the parameters to the TA profile.

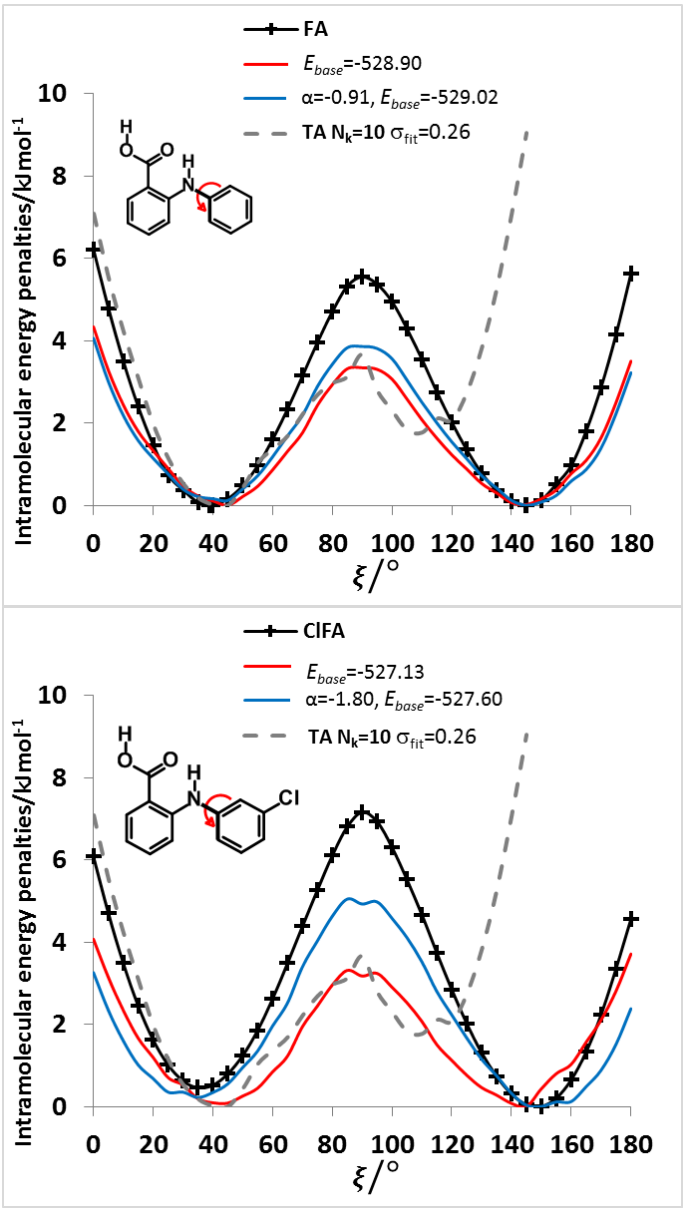


Figure 6.11: Comparison of the relative energies from *ab initio* calculations (solid black lines) with models using β_{tk} and γ_{tk} parameters transferred from TA (dotted grey plots) to (a) FA and (b) ClFA. The red curve has only had the baseline adjusted, whereas the blue curve has the α parameter refitted to FA or ClFA respectively. The transferred TA parameters are those of N_k= 10 σ_{fit}= 0.26 kJ mol^{-1}, as shown in *i*Table 6.6.

6.4. Summary

Neither the cosine terms nor the analytical models that rescaled overestimated repulsion contributions were adequate for modelling the conformational profile of fenamates. However, a physically justified model—one that combined a cos2ξ term with an intramolecular *exp*-6 atom–atom term for all atom pairs separated by three or more bonds (1-4 and above)—allowed for a very effective representation of the *ab initio* conformational energies. In this chapter, I successfully fitted conformational profiles of FA and TA to the cos2ξ plus *exp*-6 model. This success suggests that the separation of intramolecular and intermolecular forces is a promising new method for discovering force fields that are more accurate, which could in turn pave the way for modelling the condensed phases of pharmaceutical-like organic molecules. Considering a wide range of molecules and bigger dataset could improve the transferability of this physical model.

Chapter 7. Conclusions and Future Work

"*It is good to have an end to journey toward; but it is the journey that matters, in the end.*"

— Ernest Hemingway

7.1. Position of CSP in the Context of this Thesis

The CSP results of Chapters 3 and 4 successfully validated the ability of *CrystalPredictor* to generate crystal structures of flexible molecules. In Chapter 3, I successfully predicted the crystal structure of BT, which has an unusual and unexpected packing. The success required the use of distributed multipoles and *ab initio* calculations of intramolecular energy. This success inspires hope that the use of multipolar electrostatic models (as being implemented in the second-generation force fields such as AMOEBA[91]), with careful evaluation of the accuracy of the intramolecular forces, will provide a route to greater reliability for computer-aided drug design with uncommon ligands. However, the failure of a conventional point charge electrostatic model used in the intermediate step has considerable implications for other studies that rely on modelling the intermolecular interactions of 1-substituted *1H*-tetrazoles.

Chapter 4 successfully investigated the concept of *polymorphophore* via CSP studies of FA and TA in an attempt to understand why polymorphism occurs. I discovered that the fenamate fragment did qualify as a *polymorphophore* in that the dimeric unit could pack in a variety of ways with different torsion angles of the phenyl groups due to the torsional flexibility of FA and TA. Once the aromatic rings had interdigitated with their nearest neighbours, there was a large barrier to rearrangement to a very different packing, and the polymorphs were long-lived, with the torsion angles varying slightly to optimize the packing for a given structure, as we saw for TA. It was, however, the comparison between the FA and TA crystal energy landscapes that showed that it was the substituents that played the major role in determining the possible packing and relative energies, and hence, the polymorphic behaviours of FA and TA. These observations clearly demonstrate the links among *conformational flexibility*, *polymorphism*, and *polymorphophore*.

The considerable similarity in large clusters of the known and hypothetical structures of TA, and the closeness of their energies, suggested that crystal growth errors were very likely. More importantly, though, this finding guided my colleague in the discovery of a new polymorph of TA (form VI), which was nucleated by solution evaporation after

dissolving TA (form I) in ethanol and seeding it with mefenamic acid (form I). The major component of this new form (form VI is disordered: 65% and 35%) was not among the crystal structures on the final energy landscape despite generating the corresponding hypothetical structure during an intermediate step (*CrystalPredictor*). Although this observation emphasised the success of the *CrystalPredictor* algorithm, it also showed the need for larger numbers of crystal structures to go through *CrystalOptimizer*. This need for a larger dataset in the subsequent *CrystalOptimizer* step raises the question, "Is there a clear cut-off in energy for flexible molecules?" This question is even more relevant when you consider that desolvated solvates can have much higher energy than organic crystal structures[247]. In fact, it might even be desirable to go even higher in energy due to likely high conformation energy of the desolvated solvates. The *CrystalOptimizer* algorithm does allow the build-up of a database known as Local Approximate Models (LAMs), but it would be preferable to be able to check many more structures. If we could better fit intramolecular energy, which requires the bulk of the computing time, to a force field, we would be able to compute the energies of many more structures of a target molecule. In addition, computing the energies of the higher-energy structures would be much cheaper.

Chapters 3 and 4 used methods suitable for studying flexible molecules[17], as the *ab initio* charge density for the specific conformation was used in modelling the intermolecular interactions, thus representing the interplay between molecular conformation and intermolecular interactions which have been shown to be important for TA[211]. Conformational changes can optimise the directionality of strong interactions, such as H-bonding, and increase the packing density, and hence the stabilization of the structure by dispersion forces. Although, CSP methodologies have advanced to aid solid form screening in pharmaceutical development[247-250], it is nevertheless desirable to calculate the relative free energies of organic polymorphs via molecular dynamics methods[251-253]. Although there has been a great push to include the kinetics and their ability to control the polymorphs formed[1], a vital first step is to develop a sound understanding of thermodynamics.

7.2. Towards more accurate force-fields for pharmaceutical molecules

The challenge of extending the reliability and accuracy of quantum mechanical methods to larger molecules, which are more typical of pharmaceuticals and realistic biological molecules in isolation or condensed phases, is the subject of active research. This research ranges from empirical corrections to DFT methods to the use of better

functionals[254], as well as implementing explicitly correlated wavefunction methods more effectively.

Chapter 5 emphasized the challenges with using a sufficiently realistic method for the *ab initio* determination of conformational energies for larger molecules. For example, the HF method provides a qualitatively wrong and misleading conformational profile of FA and TA due to the absence of the *dispersion* (correlation effect) contribution. *Dispersion* also plays an important role in the barrier to rotation for the fenamates, especially when the aromatic rings are perpendicular to one another. The analysis of the fenamates' torsional potential revealed that the origin of the torsional barrier of the fenamates is a combination of intrinsic intramolecular *dispersion*, *electronic* contribution and the steric *repulsion*. In addition, small changes to other conformational angles made significant contributions to the torsional profile of the fenamate molecules.

In Chapter 6, a physically motivated analytical model—cos2ξ term combined with an intramolecular *exp*-6 atom–atom term for all atom pairs separated by three or more bonds—was derived using knowledge gained from Chapter 5. This model accurately fits the intramolecular energy profiles of FA and TA. This approach of modelling the conformational energies of the fenamates could be extended to pharmaceutical-like flexible organics.

7.3. Future Work

One can envisage a general scheme for determining the physically motivated potentials (Chapter 6) for a given molecule following a CSP study by an approach that is being developed[127] and applied to pharmaceutical molecules[247-249,255]. The refinement of the molecular conformations simultaneously with the crystallographic cell using *CrystalOptimizer*[72] involves the calculation and storage of a large database (LAMs) of *ab initio* conformational energies and forces for the molecule. For example, the CSP studies for FA and TA contain the results of hundreds of molecular wavefunction calculations with varying torsion and bond angles.

The database generated in a CSP study will cover most of the range of conformations that are likely to be sampled in a Molecular Dynamics study of the molecules in condensed phases, with a strong bias towards the conformations that occur in low-energy crystal structures, both observed and potential polymorphs[1]. The LAMs database could be used to improve the parameterization of the analytical conformational energy model, adapting the potential fitting routines (Appendix A) used in Chapter 6. An advantage of using an analytical functional form rather than the LAMs to interpolate between the *ab*

initio points in the CSP study is that the extrapolation to other conformations will be more realistic. The great benefit of this scheme of building up from a CSP study for a specific molecule is that the model intramolecular force field could be validated by ensuring that the model reproduces the crystal energy landscape, i.e. the energies of different packing, H-bonding and stacking modes are correctly balanced with the accompanying conformational changes.

The application of separate analytical potentials for both intra- and intermolecular terms will require adaptation of molecular modelling codes. However, the coding for the energies, forces, and second derivatives of the proposed intramolecular force field (equation 6.7) is already present in most codes. After inserting the new parameters, the major adaptation that will be required for an existing code—one that uses traditional force fields such as DL_POLY[251,256] for Molecular Dynamics simulations—is separating out the calculation of inter- and intramolecular terms. Accurate intramolecular forces are most needed in combination with the accurate anisotropic atom-atom intermolecular potentials for organic molecules, where the electrostatic term is represented by atomic multipoles, for example the static organic crystal structure modelling code, DMACRYS[34], or the rigid molecule molecular dynamics code, DL_MULTI[257]. Furthermore, the adaptation of molecular modelling code would require an adequate representation for the conformation dependence of the distributed multipoles for specific molecules, but the dependency could be represented by an analytical model[258].

Although this change in approach to pharmaceutical force fields involves envisaging a specific model fitted for each molecule, the physical basis of the current model and the results from Chapter 6 suggest that a reasonably transferable set of atom-atom intramolecular exp-6 potentials could be fitted for families of molecules. Deriving a transferable model would require a very large dataset of *ab initio* conformational profiles of many molecules calculated at an appropriate accuracy. The transferability of the electronic term (α coefficients) would need investigating. However, the use of separate atom-atom models for the forces within and between molecules appears to provide a significant improvement in the realism of current force fields. In conclusion, this approach could potentially provide an affordable route for Molecular Dynamics simulations of flexible pharmaceuticals that have the accuracy of CSP.

Appendix A

Contents of enclosed CD-ROM

The CD-ROM attached to this thesis contains FORTRAN source codes of the analytical models presented in Chapter 6 as well as some of the input files. The input data for FA, TA, MFA, and ClFA that were used in conjunction with the source codes are in directory **A1**. The inputs files included for each molecule are as follows:

a) The summary.*r* contains a summary of the flexible degree of freedom, ξ, and the intramolecular energy relative to the most stable conformation, ΔE_{intra}.
b) The geometries of all the optimised conformation in *.mol2* file format.

The source codes and example input files are in directories **A2** (cosine model), **A3** (rescaled repulsion model), and **A4** (physical model). The **A2** directory contains

a) an executable file (cos_model.*exe*),
b) FORTRAN source code (cos_model.*f90*),
c) input (summary.*r*) file,
d) output (dominant.*r*) file.

The **A3** and **A4** directories contain

a) an executable file (.*exe*),
b) FORTRAN source code (.*f90*),
c) output (dominant.*r*) file,
d) an **input** subdirectory that contains
    i. Gavezzotti's exp-6 atom-atom parameters from Table 5.2
    ii. h03adfe subroutine computes the nearest neighbour of two atoms, and
    iii. summary.*r* file.

The **A5** directory contains a utility file (extract_mol2.*sh*) that extracts the input data (summary.*r* and all the optimised conformations in *.mol2*) from the GAUSSIAN03[53] (.*log*) file.

Bibliography

1. Price, S. L. Why don't we find more polymorphs? *Acta Crystallogr. , Sect. B.* **2013,** *69*, 313-328.

2. Price, S. L. Predicting crystal structures of organic compounds. *Chem. Soc. Rev.* **2014,** *43* (7), 2098-2111.

3. Svärd, M. Structural, Kinetic and Thermodynamic Aspects of the Crystal Polymorphism of Substituted Monocyclic Aromatic Compounds. 2011.

4. Svärd, M.; Nordström, F. L.; Jasnobulka, T.; Rasmuson, Å. C. Thermodynamics and Nucleation Kinetics of m-Aminobenzoic Acid Polymorphs. *Cryst. Growth Des.* **2009,** *10* (1), 195-204.

5. Chemburkar, S. R.; Bauer, J.; Deming, K.; Spiwek, H.; Patel, K.; Morris, J.; Henry, R.; Spanton, S.; Dziki, W.; Porter, W.; Quick, J.; Bauer, P.; Donaubauer, J.; Narayanan, B. A.; Soldani, M.; Riley, D.; McFarland, K. Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development. *Org. Process Res. Dev.* **2000,** *4* (5), 413-417.

6. Day, G. M. Current approaches to predicting molecular organic crystal structures. *Crystallography Reviews* **2011,** *17* (1), 3-52.

7. Byrn, S. R.; Pfeiffer, R.; Ganey, M.; Hoiberg, C.; Poochikian, G. Pharmaceutical Solids - A Strategic Approach to Regulatory Considerations. *Pharmaceut. Res.* **1995,** *12* (7), 945-954.

8. Price, S. L. Lattice energy, nailed? *Science* **2014,** *345* (6197), 619-620.

9. Newman, A. Specialized Solid Form Screening Techniques. *Org. Process Res. Dev.* **2013,** *17* (3), 457-471.

10. Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A.; Schweizer, B. Significant progress in predicting the crystal structures of small organic molecules - a report on the fourth blind test. *Acta Crystallogr. , Sect. B* **2009,** *65* (2), 107-125.

11. Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S. X.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; la Valle, R. G.; Desiraju, G. R.; van Eijck, B. P.; Facelli, J. C.; Ferraro, M. B.; Grillo, D.; Habgood, M.; Hofmann, D. W.; Hofmann, F.; Jose, K., V; Karamertzanis, P. G.; Kazantsev, A. V.; Kendrick, J.; Kuleshova, L. N.; Leusen, F. J.; Maleev, A. V.; Misquitta, A. J.; Mohamed, S.; Needs, R. J.; Neumann, M. A.; Nikylov, D.; Orendt, A. M.; Pal, R.; Pantelides, C. C.; Pickard, C. J.; Price, L. S.; Price, S. L.; Scheraga, H. A.; van de

Streek, J.; Thakur, T. S.; Tiwari, S.; Venuti, E.; Zhitkov, I. K. Towards crystal structure prediction of complex organic compounds - a report on the fifth blind test. *Acta Crystallogr. , Sect. B* **2011,** *67* (6), 535-551.

12. Price, S. L. Developing Methods of Crystal Structure and Polymorph Prediction. *Implications of Molecular and Materials Structures for New Technologies* **1999,** 315-320.

13. Neumann, M. A. Tailor-made force fields for crystal-structure prediction. *J. Phys. Chem. B* **2008,** *112* (32), 9810-9829.

14. Price, S. L. Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism. *Accounts Chem. Res.* **2009,** *42* (1), 117-126.

15. Groom, C. R.; Reilly, A. M. Sixth blind test of organic crystal-structure prediction methods. *Acta Crystallogr. , Sect. B* **2014,** *70* (4), 776-777.

16. Beyer, T.; Lewis, T.; Price, S. L. Which organic crystal structures are predictable by lattice energy minimisation? *CrystEngComm* **2001,** *3* (44), 178-212.

17. Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L.; Galek, P. T.; Day, G. M.; Cruz-Cabeza, A. J. Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction. *Int. J. Pharm.* **2011,** *418* (2), 168-178.

18. Lopez-Mejias, V.; Kampf, J. W.; Matzger, A. J. Nonamorphism in Flufenamic Acid and a New Record for a Polymorphic Compound with Solved Structures. *J. Am. Chem. Soc.* **2012,** *134* (24), 9872-9875.

19. Lopez-Mejias, V.; Kampf, J. W.; Matzger, A. J. Polymer-Induced Heteronucleation of Tolfenamic Acid: Structural Investigation of a Pentamorph. *J. Am. Chem. Soc.* **2009,** *131* (13), 4554-4555.

20. Yu, L. Polymorphism in Molecular Solids: An Extraordinary System of Red, Orange, and Yellow Crystals. *Accounts Chem. Res.* **2010,** *43* (9), 1257-1266.

21. Chen, S. A.; Xi, H. M.; Yu, L. Cross-nucleation between ROY polymorphs. *J. Am. Chem. Soc.* **2005,** *127* (49), 17439-17444.

22. Murthy, H. K.; Bhat, T. N.; Vijayan, M. Structural studies of analgesics and their interactions. 9. Structure of a new crystal form of 2-([3-(trifluoromethyl) phenyl] amino) benzoic acid (flufenamic acid). *Acta Crystallogr. , Sect. B* **1982,** *38* (JAN), 315-317.

23. Lutker, K. M.; Tolstyka, Z. P.; Matzger, A. J. Investigation of a privileged polymorphic motif: A dimeric ROY derivative. *Cryst. Growth Des.* **2008,** *8* (1), 136-139.

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

25. Bernstein, J. *Polymorphism in Molecular Crystals;* Clarendon Press: Oxford, 2002.

26. Stone, A. J. *The Theory of Intermolecular Forces;* 2 ed.; Oxford University Press: Oxford, 2013.

27. Buckingham, A. D.; Utting, B. D. Intermolecular Forces. *Annu. Rev. Phys. Chem.* **1970,** *21* (1), 287-316.

28. Maitland, G. C.; Rigby, M.; Smith, E. B.; Wakeham, W. A. *Intermolecular Forces. Their Origin and Determination.;* Clarendon Press: Oxford, 1981.

29. Israelachvili, J. *Intermolecular and Surface Forces, Third Edition;* Academic Press: 2010.

30. London, F. The general theory of molecular forces. *Trans. Faraday Soc.* **1937,** *33*, 8b-26.

31. Spencer, N. D.; Moore, J. H. *Encyclopedia of Chemical Physics and Physical Chemistry: Fundamentals;* 1 ed.; Taylor & Francis: 2001.pp. 1004.

32. Welch, G. W. A.; Karamertzanis, P. G.; Misquitta, A. J.; Stone, A. J.; Price, S. L. Is the induction energy important for modeling organic crystals? *J. Chem. Theory Comput.* **2008,** *4* (3), 522-532.

33. Price, S. L. From crystal structure prediction to polymorph prediction: interpreting the crystal energy landscape. *Phys. Chem. Chem. Phys.* **2008,** *10* (15), 1996-2009.

34. Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling Organic Crystal Structures using Distributed Multipole and Polarizability-Based Model Intermolecular Potentials. *Phys. Chem. Chem. Phys.* **2010,** *12* (30), 8478-8490.

35. Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. Evaluation of charge penetration between distributed multipolar expansions. *J. Chem. Phys.* **2000,** *112* (17), 7300-7306.

36. Lund, A. M.; Orendt, A. M.; Pagola, G. I.; Ferraro, M. B.; Facelli, J. C. Optimization of Crystal Structures of Archetypical Pharmaceutical Compounds: A Plane-Wave DFT-D Study Using Quantum Espresso. *Crystal Growth & Design* **2013,** *13* (5), 2181-2189.

37. Neumann, M. A.; Perrin, M. A. Energy ranking of molecular crystals using density functional theory calculations and an empirical van der Waals correction. *J. Phys. Chem. B* **2005,** *109* (32), 15531-15541.

38. Cox, S. R.; Hsu, L. Y.; Williams, D. E. Nonbonded Potential Function Models for Crystalline Oxohydrocarbons. *Acta Crystallogr. , Sect A.* **1981,** *37* (MAY), 293-301.

39. Williams, D. E.; Cox, S. R. Nonbonded Potentials For Azahydrocarbons: the Importance of the Coulombic Interaction. *Acta Crystallogr. , Sect. B* **1984,** *40* (8), 404-417.

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

41. Gavezzotti, A. *Theoretical Aspects and Computer Modeling of the Molecular Solid State;* John Wiley: Chichester, 1997; Vol. 1.

42. Filippini, G.; Gavezzotti, A. Empirical Intermolecular Potentials For Organic-Crystals: the '6-Exp' Approximation Revisited. *Acta Crystallogr. , Sect. B* **1993,** *B49*, 868-880.

43. Williams, D. E.; Weller, R. R. Lone-Pair Electronic Effects On the Calculated Abinitio Scf-Mo Electric-Potential and the Crystal-Structures of Azabenzenes. *J. Am. Chem. Soc.* **1983,** *105* (13), 4143-4148.

44. Karamertzanis, P. G.; Pantelides, C. C. Optimal site charge models for molecular electrostatic potentials. *Mol. Simulat.* **2004,** *30* (7), 413-436.

45. Spackman, M. A. Molecular Electric Moments from X-ray Diffraction Data. *Chem. Rev.* **1992,** *92*, 1769-1797.

46. Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles From Molecular Electrostatic Potentials - the Need For High Sampling Density in Formamide Conformational-Analysis. *J. Comput. Chem.* **1990,** *11* (3), 361-373.

47. Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. Charges fit to electrostatic potentials. II. Can atomic charges be unambiguously fit to electrostatic potentials? *J. Comput. Chem.* **1996,** *17* (3), 367-383.

48. Stone, A. J.; Alderton, M. Distributed Multipole Analysis - Methods and Applications. *Molecular Physics* **1985,** *56* (5), 1047-1064.

49. Welch, G. W. A. The induction contribution to the lattice energy of organic crystals. PhD 2010.

50. Stone, A. J. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. *Chemical Physics Letters* **1981,** *83* (2), 233-239.

51. Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005,** *1* (6), 1128-1132.

52. *GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System,* version 2.2; University of Cambridge: Cambridge, United Kingdom, 2010

53. *Gaussian 03,* Gaussian Inc.: Wallingford CT, 2004

54. *GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System,* version 1.0; University of Cambridge: Cambridge, United Kingdom, 1999

55. Price, S. L. Computational prediction of organic crystal structures and polymorphism. *Int. Rev. Phys. Chem.* **2008,** *27* (3), 541-568.

56. Kazantsev, A. V. Molecular Flexibility in Crystal Structure Prediction. PhD Imperial College, London, 2011.

57. Smith, W. Point multipoles in the Ewald summation. *CCP5 Quarterly* **1982,** *4*, 13-25.

58. Cruz-Cabeza, A. J.; Bernstein, J. Conformational Polymorphism. *Chem. Rev.* **2014,** *114* (4), 2170-2191.

59. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force-Field For the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules (Vol 117, Pg 5179, 1995). *J. Am. Chem. Soc.* **1996,** *118* (9), 2309.

60. van Mourik, T.; Karamertzanis, P. G.; Price, S. L. Molecular conformations and relative stabilities can be as demanding of the electronic structure method as intermolecular calculations. *J. Phys. Chem. A* **2006,** *110* (1), 8-12.

61. van Mourik, T. Assessment of Density Functionals for Intramolecular Dispersion-Rich Interactions. *J. Chem. Theory Comput.* **2008,** *4* (10), 1610-1619.

62. Comba, P.; Hambley, T. W.; Martin, B. *Molecular modeling of inorganic compounds;* John Wiley & Sons: 2009.

63. Leach, A. R. *Molecular Modelling. Principles and Applications;* Longman: Harlow, 1996.

64. Goodman, J. M. *Chemical Applications of Molecular Modelling;* Royal Society of Chemistry: Cambridge, U.K., 1998.

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

66. Lommerse, J. P. M.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Mooij, W. T. M.; Price, S. L.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. A test of crystal structure prediction of small organic molecules. *Acta Crystallogr. , Sect. B* **2000,** *56*, 697-714.

67. Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Lommerse, J. P. M.; Mooij, W. T. M.; Price, S. L.; Scheraga, H.; Schweizer, B.; Schmidt, M. U.; van Eijck,

B. P.; Verwer, P.; Williams, D. E. Crystal structure prediction of small organic molecules: a second blind test. *Acta Crystallogr. , Sect. B* **2002,** *58*, 647-661.

68. Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Protein Simulations* **2003,** *66*, 27-86.

69. Cornell, W. D.; Howard, A. E.; Kollman, P. Molecular mechanical potential functions and their application to study molecular systems: Current Opinion in Structural Biology 1991, 201-212. *Curr. Opin. Struc. Biol.* **1991,** *1* (2), 201-212.

70. Comba, P.; Hambley, T. W.; Martin, B. Molecular modeling of inorganic compounds. **2009,** *3rd completely rev. and enlarged ed*.

71. Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C. CrystalOptimizer. An efficient Algorithm for Lattice Energy Minimisation of Organic Crystal using Isolated-Molecule Quantum Mechanical Calculations. In *Molecular System Engineering*, Adjiman, C. S., Galindo, A., Eds.; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2010; Vol. 6, pp 1-42.

72. Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C. Efficient Handling of Molecular Flexibility in Lattice Energy Minimization of Organic Crystals. *J. Chem. Theory Comput.* **2011,** *7* (6), 1998-2016.

73. Jones, J. E. On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proceedings of the Royal Society of London. Series A* **1924,** *106* (738), 463-477.

74. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J. Am. Chem. Soc.* **1984,** *106* (3), 765-784.

75. Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J. Phys. Chem.* **1975,** *79* (22), 2361-2381.

76. Nemethy, G.; Pottle, M. S.; Scheraga, H. A. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. *J. Phys. Chem.* **1983,** *87* (11), 1883-1887.

77. Hagler, A. T.; Lifson, S. Energy functions for peptides and proteins. II. Amide hydrogen bond and calculation of amide crystal properties. *J. Am. Chem. Soc.* **1974,** *96* (17), 5327-5335.

78. Hagler, A. T.; Huler, E.; Lifson, S. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J. Am. Chem. Soc.* **1974,** *96* (17), 5319-5327.

79. Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* **1981,** *103* (2), 335-340.

80. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.

81. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization and Dynamics Calculations. *J. Comput. Chem.* **1983,** *4* (2), 187-217.

82. Becker, O. M. *Computational biochemistry and biophysics;* M. Dekker: New York, 2001.

83. MacKerell, A. D.; Bashford, D.; Bellott; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998,** *102* (18), 3586-3616.

84. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010,** *31* (4), 671-690.

85. Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *P. Natl. Acad. Sci. USA* **2005,** *102* (19), 6665-6670.

86. Allinger, N. L.; Chen, K.; Lii, J. H. An improved force field (MM4) for saturated hydrocarbons. *J. Comput. Chem.* **1996,** *17* (5-6), 642-668.

87. Halgren, T. A. Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996,** *17* (5-6), 490-519.

88. Allinger, N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.* **1977,** *99* (25), 8127-8134.

89. Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2011,** *8* (1), 61-74.

90. Stouch, T. R. The errors of our ways: taking account of error in computer-aided drug design to build confidence intervals for our next 25 years. *J. Comput. -Aided Mol. Des.* **2012,** *26* (1), 125-134.

91. Ponder, J. W.; Wu, C. J.; Ren, P. Y.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010,** *114* (8), 2549-2564.

92. Day, G. M.; Motherwell, W. D. S.; Jones, W. Beyond the isotropic atom model in crystal structure prediction of rigid molecules: Atomic multipoles versus point charges. *Cryst. Growth Des.* **2005,** *5* (3), 1023-1033.

93. Marshall, G. R. Limiting assumptions in molecular modeling: electrostatics. *J. Comput. -Aided Mol. Des.* **2013,** *27* (2), 107-114.

94. Payne, R. S.; Rowe, R. C.; Roberts, R. J.; Charlton, M. H.; Docherty, R. Potential polymorphs of aspirin. *J. Comput. Chem.* **1999,** *20* (2), 262-273.

95. Gourlay, M. D.; Kendrick, J.; Leusen, F. J. J. Rationalization of racemate resolution: Predicting spontaneous resolution through crystal structure prediction. *Cryst. Growth Des.* **2007,** *7* (1), 56-63.

96. Uzoh, O. G. Analytically Modelling Intramolecular Energies of Fenamic and Tolfenamic Acids. Unpublished Work, 2013.

97. *NAG Fortran Library Routine Document G02DDF*, Oxford, United Kingdom: 1989

98. Beran, G. J.; Hirata, S. Fragment and localized orbital methods in electronic structure theory. *Physical Chemistry Chemical Physics* **2012,** *14* (21), 7559-7561.

99. Orio, M.; Pantazis, D. A.; Neese, F. Density functional theory. *Photosynthesis Research* **2009,** *102* (2-3), 443-453.

100. Sherrill, C. D. An introduction to Hartree-Fock molecular orbital theory. *School of Chemistry and Biochemistry Georgia Institute of Technology* **2000**.

101. Shaik, S. S.; Hiberty, P. C. *A chemist's guide to valence bond theory;* John Wiley & Sons: 2007.

102. Wu, W.; Su, P. F.; Shaik, S.; Hiberty, P. C. Classical Valence Bond Approach by Modern Methods. *Chem. Rev.* **2011,** *111* (11), 7557-7593.

103. Gasteiger, J.; Engel, T. *Chemoinformatics: a textbook;* John Wiley & Sons: 2006.

104. Fukui, K.; Yonezawa, T.; Shingu, H. A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *J. Chem. Phys.* **1952,** *20* (4), 722-725.

105. Bell, S.; Dines, T. J.; Chowdhry, B. Z.; Withnall, R. Computational chemistry using modern electronic structure methods. *J. Chem. Educ.* **2007,** *84* (8), 1364-1370.

106. Moller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934,** *46*, 618-622.

107. Valdes, H.; Klusak, V.; Pitonak, M.; Exner, O.; Stary, I.; Hobza, P.; Rulisek, L. Evaluation of the intramolecular basis set superposition error in the calculations of larger molecules: [n]Helicenes and Phe-Gly-Phe tripeptide. *J. Comput. Chem.* **2008,** *29* (6), 861-870.

108. Simon, S.; Duran, M.; Dannenberg, J. J. How does basis set superposition error change the potential surfaces for hydrogen bonded dimers? *Journal of Chemical Physics* **1996,** *105* (24), 11024-11031.

109. Szczesniak, M. M.; Scheiner, S. Correction of the Basis Set Superposition Error in SCF and MP2 Interaction Energies - the Water Dimer. *J. Chem. Phys.* **1986,** *84* (11), 6328-6335.

110. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab-Initio Calculation of Vibrational Absorption and Circular-Dichroism Spectra Using Density-Functional Force-Fields. *J. Phys. Chem.* **1994,** *98* (45), 11623-11627.

111. Car, R. Introduction to density-functional theory and ab-initio molecular dynamics. *Quant. Struct. -Act. Rel.* **2002,** *21* (2), 97-104.

112. Grimme, S.; Antony, J.; Schwabe, T.; Muck-Lichtenfeld, C. Density functional theory with dispersion corrections for supramolecular structures, aggregates, and complexes of (bio)organic molecules. *Organic & Biomolecular Chemistry* **2007,** *5* (5), 741-758.

113. Kohn, W.; Becke, A. D.; Parr, R. G. Density functional theory of electronic structure. *J. Phys. Chem.* **1996,** *100* (31), 12974-12980.

114. Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review B* **1964,** *136* (3B), B864-B871.

115. Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965,** *140* (4A), A1133.

116. Becke, A. D. Density-Functional Thermochemistry .3. the Role of Exact Exchange. *J. Chem. Phys.* **1993,** *98* (7), 5648-5652.

117. Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula Into A Functional of the Electron-Density. *Phys. Rev. B* **1988,** *37* (2), 785-789.

118. Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999,** *110* (13), 6158-6170.

119. Sherrill, C. D. Introduction to Electron Correlation. *School of Chemistry and Biochemistry Georgia Institute of Technology* **2008**.

120. Grafenstein, J.; Cremer, D. An efficient algorithm for the density-functional theory treatment of dispersion interactions. *J. Chem. Phys.* **2009,** *130* (12).

121. Grimme, S. Density functional theory with London dispersion corrections. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2011,** *1* (2), 211-228.

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

123. Brown, T. L. *Chemistry: The Central Science;* Pearson Education: 2009.

124. Gavezzotti, A.; Filippini, G. Polymorphic Forms of Organic-Crystals at Room Conditions - Thermodynamic and Structural Implications. *J. Am. Chem. Soc.* **1995,** *117* (49), 12299-12305.

125. Karamertzanis, P. G. Prediction of crystal structure of molecular solids. PhD Imperial College, London, 2004.

126. Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure prediction. II. Flexible molecules. *Mol. Phys.* **2007,** *105* (2-3), 273-291.

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

128. Sobol', I. M. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comp. Math. Math. Phys.* **1967,** *7*, 86-112.

129. *The NAG Fortran Library,* Oxford, United Kingdom, 2013

130. Karamertzanis, P. G.; Price, S. L. Energy Minimization of Crystal Structures Containing Flexible Molecules. *J. Chem. Theory Comput.* **2006,** *2* (4), 1184-1199.

131. Cooper, T. G.; Hejczyk, K. E.; Jones, W.; Day, G. M. Molecular Polarization Effects on the Relative Energies of the Real and Putative Crystal Structures of Valine. *J. Chem. Theory Comput.* **2008,** *4* (10), 1795-1805.

132. Mennucci, B.; Tomasi, J. Continuum solvation models: A new approach to the problem of solute's charge distribution and cavity boundaries. *J. Chem. Phys.* **1997,** *106* (12), 5151-5158.

133. Habgood, M.; Price, S. L.; Portalone, G.; Irrera, S. Testing a Variety of Electronic-Structure-Based Methods for the Relative Energies of 5-Formyluracil Crystals. *J. Chem. Theory Comput.* **2011,** *7* (9), 2685-2688.

134. Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure prediction - I. Rigid molecules. *Journal of Computational Chemistry* **2005,** *26* (3), 304-324.

135. Williams, D. E. Improved intermolecular force field for molecules containing H, C, N, and O atoms, with application to nucleoside and peptide crystals. *J. Comput. Chem.* **2001,** *22* (11), 1154-1166.

136. Day, G. M.; Motherwell, W. D. S.; Jones, W. A strategy for predicting the crystal structures of flexible molecules: the polymorphism of phenobarbital. *Physical Chemistry Chemical Physics* **2007,** *9* (14), 1693-1704.

137. Kendrick, J.; Leusen, F. J.; Neumann, M. A.; van de Streek, J. Progress in Crystal Structure Prediction. *Chemistry - A European Journal* **2011,** *17* (38), 10735-10743.

138. Braun, D. E.; Karamertzanis, P. G.; Price, S. L. Which, if any, hydrates will crystallise? Predicting hydrate formation of two dihydroxybenzoic acids. *Chem. Commun.* **2011,** *47* (19), 5443-5445.

139. Cruz-Cabeza, A. J.; Karki, S.; Fabian, L.; Friscic, T.; Day, G. M.; Jones, W. Predicting stoichiometry and structure of solvates. *Chem. Commun.* **2010,** *46* (13), 2224-2226.

140. Karamertzanis, P. G.; Kazantsev, A. V.; Issa, N.; Welch, G. W. A.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L. Can the Formation of Pharmaceutical Co-Crystals Be Computationally Predicted? II. Crystal Structure Prediction. *J. Chem. Theory Comput.* **2009,** *5* (5), 1432-1448.

141. Issa, N.; Barnett, S. A.; Mohamed, S.; Braun, D. E.; Copley, R. C. B.; Tocher, D. A.; Price, S. L. Screening for cocrystals of succinic acid and 4-aminobenzoic acid. *CrystEngComm* **2012,** *14* (7), 2454-2464.

142. Cruz-Cabeza, A. J.; Day, G. M.; Jones, W. Structure prediction, disorder and dynamics in a DMSO solvate of carbamazepine. *Phys. Chem. Chem. Phys.* **2011,** *13* (28), 12808-12816.

143. Habgood, M. Form II Caffeine: A Case Study for Confirming and Predicting Disorder in Organic Crystals. *Cryst. Growth Des.* **2011,** *11* (8), 3600-3608.

144. Torrisi, A.; Leech, C. K.; Shankland, K.; David, W. I. F.; Ibberson, R. M.; Benet-Buchholz, J.; Boese, R.; Leslie, M.; Catlow, C. R. A.; Price, S. L. The solid phases of cyclopentane: a combined experimental and simulation study. *J. Phys. Chem. B* **2008,** *112* (12), 3746-3758.

145. Braun, D. E.; Karamertzanis, P. G.; Arlin, J. B.; Florence, A. J.; Kahlenberg, V.; Tocher, D. A.; Griesser, U. J.; Price, S. L. Solid-State Forms of beta-Resorcylic Acid: How Exhaustive Should a Polymorph Screen Be? *Cryst. Growth Des.* **2011,** *11* (1), 210-220.

146. Llinas, A.; Goodman, J. M. Polymorph Control: past, present and future. *Drug Discover. Today* **2008,** *13* (5/6), 198-210.

147. Arlin, J. B.; Price, L. S.; Price, S. L.; Florence, A. J. A strategy for producing predicted polymorphs: catemeric carbamazepine form V. *Chem. Commun.* **2011,** *47* (25), 7074-7076.

148. Maddox, J. Crystals from first Principles. *Nature* **1988,** *335* (6187), 201.

149. Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwer, P. A third blind test of crystal structure prediction. *Acta Crystallogr. , Sect. B* **2005,** *61* (5), 511-527.

150. Collins, A.; Wilson, C. C.; Gilmore, C. J. Comparing entire crystal structures using cluster analysis and fingerprint plots. *CrystEngComm* **2010,** *12* (3), 801-809.

151. Chisholm, J. A.; Motherwell, S. COMPACK: a program for identifying crystal structure similarity using distances. *J. Appl. Crystallogr.* **2005,** *38*, 228-231.

152. Asmadi, A.; Kendrick, J.; Leusen, F. J. J. Crystal Structure Prediction and Isostructurality of Three Small Molecules. *Chem. Eur. J* **2010,** *16* (42), 12701-12709.

153. Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. Mercury CSD 2.0 - new features for the visualization and investigation of crystal structures. *J. Appl. Crystallogr.* **2008,** *41*, 466-470.

154. Chisholm, J. A.; Motherwell, S. A new algorithm for performing three-dimensional searches of the Cambridge Structural Database. *J. Appl. Crystallogr.* **2004,** *37*, 331-334.

155. Gelbrich, T.; Hursthouse, M. B. Systematic investigation of the relationships between 25 crystal structures containing the carbamazepine molecule or a close analogue: a case study of the XPac method. *CrystEngComm* **2006,** *8* (6), 448-460.

156. Gelbrich, T.; Hursthouse, M. B. A versatile procedure for the identification, description and quantification of structural similarity in molecular crystals. *CrystEngComm* **2005,** *7*, 324-336.

157. McKinnon, J. J.; Spackman, M. A.; Mitchell, A. S. Novel tools for visualizing and exploring intermolecular interactions in molecular crystals. *Acta Crystallogr. , Sect. B* **2004,** *60*, 627-668.

158. Spackman, M. A.; McKinnon, J. J. Fingerprinting intermolecular interactions in molecular crystals. *CrystEngComm* **2002,** 378-392.

159. Parkin, A.; Barr, G.; Dong, W.; Gilmore, C. G.; Jayatilaka, D.; McKinnon, J. J.; Spackman, M. A.; Wilson, C. C. Comparing entire crystal structures: structural genetic fingerprinting. *CrystEngComm* **2007,** *9*, 648-652.

160. McKinnon, J. J.; Mitchell, A. S.; Spackman, M. A. Hirshfeld surfaces: A new tool for visualising and exploring molecular crystals. *Chem. Eur. J* **1998,** *4* (11), 2136-2141.

161. Etter, M. C.; MacDonald, J. C.; Bernstein, J. Graph-Set Analysis of Hydrogen-Bond Patterns in Organic- Crystals. *Acta Crystallogr. , Sect. B* **1990,** *46*, 256-262.

162. Grell, J.; Bernstein, J.; Tinhofer, G. Graph-set analysis of hydrogen-bond patterns: some mathematical concepts. *Acta Crystallogr. , Sect. B* **1999,** *55*, 1030-1043.

163. Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical methods: the art of scientific computing (Fortran version);* Cambridge University Press: Cambridge, 1989.

164. Galek, P. T. A. Novel comparison of crystal packing by moments of inertia. *CrystEngComm* **2011,** *13* (3), 841-849.

165. Bonnet, P. A.; Streek, J. v. d.; Trask, A. V.; Motherwell, W. D. S.; Jones, W. The crystal structure of an unstable polymorph of beta-D-allose. *CrystEngComm* **2004,** *6* (86), 535-539.

166. Potter, B. S.; Palmer, R. A.; Withnall, R.; Chowdhry, B. Z.; Price, S. L. Aza analogues of nucleic acid bases: experimental determination and computational prediction of the crystal structure of anhydrous 5-azauracil. *J. Mol. Struct.* **1999,** *486*, 349-361.

167. Nowell, H.; Frampton, C. S.; Waite, J.; Price, S. L. Blind crystal structure prediction of a novel second polymorph of 1-hydroxy-7-azabenzotriazole. *Acta Crystallogr. , Sect. B* **2006,** *62* (4), 642-650.

168. D'Oria, E.; Karamertzanis, P. G.; Price, S. L. Spontaneous Resolution of Enantiomers by Crystallization: Insights from Computed Crystal Energy Landscapes. *Cryst. Growth Des.* **2010,** *10* (4), 1749-1756.

169. Thornber, C. W. Isosterism and molecular modification in drug design. *Chem. Soc. Rev.* **1979,** *8* (4), 563-580.

170. Singh, H.; Singh Chawla, A.; Kapoor, V. K.; Paul, D.; Malhotra, R. K. 4 Medicinal Chemistry of Tetrazoles. In *Progress in Medicinal Chemistry*, Volume 17 ed.; Ellis and, G. P., Ed.; Elsevier: 1980; pp 151-183.

171. Lipinski, C. A. Chapter 27. Bioisosterism in Drug Design. In *Annual Reports in Medicinal Chemistry*, Volume 21 ed.; Denis, M. B., Ed.; Academic Press: 1986; pp 283-291.

172. *SHELXL97*, University of Göttingen: Göttingen, Germany, 1997

173. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. , Sect. A.* **2008,** *64* (1), 112-122.

174. Farrugia, L. WinGX suite for small-molecule single-crystal crystallography. *J. Appl. Crystallogr.* **1999,** *32* (4), 837-838.

175. Ladd, M. F. C.; Palmer, R. A. *Structure Determination by X-Ray Crystallography;* Springer: 2003.

176. Khan, S. A.; Reshak, A. H. First Principle Study of Electronic Structure, Chemical Bonding and Optical Properties of 5-azido-1H-tetrazole. *International Journal of Electrochemical Science* **2013,** *8* (7), 9459-9473.

177. Desiraju, G. R. Hydrogen bridges in crystal engineering: Interactions without borders. *Accounts Chem. Res.* **2002,** *35* (7), 565-573.

178. Herrebout, W. A.; Suhm, M. A. Weak hydrogen bonds - strong effects? *Phys. Chem. Chem. Phys.* **2011,** *13* (31).

179. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure Systems. *J. Comput. Chem.* **1993,** *14*, 1347-1363.

180. Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *J. Chem. Phys.* **2002,** *117* (1), 43-45.

181. An email message sent Febrary 2012 to Rex Palmer stated the following "we would be delighted if it were 2nd_23.res P21 a = 7.9 ... or 1st_120.res P21/c a = 10.0 ... as our two lowest energy structures".

182. Allen, F. H. The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallogr. , Sect. B* **2002,** *58* (3), 380-388.

183. Thakur, T. S.; Desiraju, G. R. Crystal Structure Prediction of a Co-Crystal Using a Supramolecular Synthon Approach: 2-Methylbenzoic Acid-2-Amino-4-methylpyrimidine. *Cryst. Growth Des.* **2008,** *8* (11), 4031-4044.

184. Surov, A. O.; Perlovich, G. L. Thermodynamics of sublimation, thermophysical and structural aspects of the molecular crystals of fenamates. *J. Struct. Chem.* **2010,** *51* (2), 308-315.

185. Li, H.; Stowell, J. G.; Borchardt, T. B.; Byrn, S. R. Synthesis, Conformational Polymorphism, and Construction of a GT Diagram of 2-[(2-Nitrophenyl)amino]-3-thiophenecarbonitrile. *Cryst. Growth Des.* **2006,** *6* (11), 2469-2474.

186. Pagola, S.; Stephens, P. W.; He, X.; Byrn, S. R. Crystal structure solution of the dark red, light red and orange pollymorphs of 5-methyl-2-[(2-nitro-4-methylphenyl)amino]-3-thiophenecarbonitrile by high resolution X-ray powder diffraction and simulated annealing techniques. Delhez, R., Mittemeijer, E. J., Eds.; 2001; pp 789-794.

187. Zhou, T.; Li, F.; Fan, Y.; Song, W.; Mu, X.; Zhang, H.; Wang, Y. Hydrogen-bonded dimer stacking induced emission of aminobenzoic acid compounds. *Chem. Commun.* **2009,** (22), 3199-3201.

188. Mei, X. F.; August, A. T.; Wolf, C. Regioselective copper-catalyzed amination of chlorobenzoic acids: Synthesis and solid-state structures of N-aryl anthranilic acid derivatives. *Journal of Organic Chemistry* **2006,** *71* (1), 142-149.

189. Alvarez, A. J.; Singh, A.; Myerson, A. S. Polymorph Screening: Comparing a Semi-Automated Approach with a High Throughput Method. *Cryst. Growth Des.* **2009,** *9* (9), 4181-4188.

190. SeethaLekshmi, S.; Guru Row, T. N. Conformational polymorphism in a non-steroidal anti-inflammatory drug, mefenamic acid. *Cryst. Growth Des.* **2012,** *12* (8), 4283-4289.

191. Dokorou, V.; Kovala-Demertzi, D.; Jasinski, J. P.; Galani, A.; Demertzis, M. A. Synthesis, Spectroscopic Studies, and Crystal Structures of Phenylorganotin Derivatives with [Bis(2,6-dimethylphenyl)amino]benzoic Acid: Novel Antituberculosis Agents. *Helv. Chim. Acta* **2004,** *87* (8), 1940-1950.

192. Bernstein, J. Polymorphism - A Perspective. *Cryst. Growth Des.* **2011,** *11* (3), 632-650.

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

194. Yang, S. S.; Guillory, J. K. Polymorphism in sulfonamides. *J. Pharm. Sci.* **1972,** *61* (1), 26-40.

195. Rossi, D.; Gelbrich, T.; Kahlenberg, V.; Griesser, U. J. Supramolecular constructs and thermodynamic stability of four polymorphs and a co-crystal of pentobarbital (nembutal). *CrystEngComm* **2012,** *14* (7), 2494-2506.

196. Lutker, K. M.; Matzger, A. J. Crystal polymorphism in a carbamazepine derivative: Oxcarbazepine. *J. Pharm. Sci.* **2010,** *99* (2), 794-803.

197. Costantino, L.; Barlocco, D. Privileged Structures as Leads in Medicinal Chemistry. *Current Medicinal Chemistry* **2006,** *13* (1), 65-85.

198. Keinänen S.; Similä S.; Kouvalainen K Oral antipyretic therapy: Evaluation of the N-aryl-anthranilic acid derivatives mefenamic acid, tolfenamic acid and flufenamic acid. *Eur J Clin Pharmacol* **1978,** *13* (5), 331-334.

199. Andersen, K. V.; Larsen, S.; Alhede, B.; Gelting, N.; Buchardt, O. Characterization of two polymorphic forms of tolfenamic acid, N-(2-methyl-3-chlorophenyl)anthranilic acid: their crystal structures and relative stabilities. *J. Chem. Soc. Perkin Trans 2* **1989,** (10), 1443-1447.

200. Mattei, A.; Li, T. Polymorph Formation and Nucleation Mechanism of Tolfenamic Acid in Solution: An Investigation of Pre-nucleation Solute Association. *Pharmaceut. Res.* **2012,** *29* (2), 460-470.

201. Surov, A. O.; Szterner, P.; Zielenkiewicz, W.; Perlovich, G. L. Thermodynamic and structural study of tolfenamic acid polymorphs. *J. Pharm. Biomed. Anal.* **2009,** *50* (5), 831-840.

202. Brittain, H. G. *Polymorphism in Pharmaceutical Solids;* Informa Healthcare: New York, London, 2009; Vol. 192.

203. van de Streek, J.; Motherwell, S. Searching the Cambridge Structural Database for polymorphs. *Acta Crystallogr. , Sect. B* **2005,** *61*, 504-510.

204. Desiraju, G. R. Polymorphism: The same and not quite the same. *Cryst. Growth Des.* **2008,** *8* (1), 3-5.

205. Bhatt, P. M.; Desiraju, G. R. Tautomeric polymorphism in omeprazole. *Chem. Commun.* **2007,** (20), 2057-2059.

206. de Gelder, R.; Wehrens, R.; Hageman, J. A. A generalized expression for the similarity of spectra: Application to powder diffraction pattern classification. *J. Comput. Chem.* **2001,** *22* (3), 273-289.

207. Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. New software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallogr. , Sect. B* **2002,** *58*, 389-397.

208. Allen, F. H.; Taylor, R. Research applications of the Cambridge Structural Database (CSD). *Chem. Soc. Rev.* **2004,** *33* (8), 463-475.

209. Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L.; Galek, P. T. A.; Day, G. M.; Cruz-Cabeza, A. J. Successful Prediction of a Model Pharmaceutical in the Fifth Blind Test of Crystal Structure Prediction. *Int. J. Pharm.* **2011,** *418* (2), 168-178.

210. Macrae, C. F.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Shields, G. P.; Taylor, R.; Towler, M.; van de Streek, J. Mercury: visualization and analysis of crystal structures. *J. Appl. Crystallogr.* **2006,** *39*, 453-457.

211. Mattei, A.; Li, T. L. Interplay between molecular conformation and intermolecular interactions in conformational polymorphism: A molecular perspective from electronic calculations of tolfenamic acid. *Int. J. Pharm.* **2011,** *418* (2), 179-186.

212. Habgood, M.; Grau-Crespo, R.; Price, S. L. Substitutional and orientational disorder in organic crystals: a symmetry-adapted ensemble model. *Phys. Chem. Chem. Phys.* **2011,** *13* (20), 9590-9600.

213. Lancaster, R. W.; Karamertzanis, P. G.; Hulme, A. T.; Tocher, D. A.; Lewis, T. C.; Price, S. L. The Polymorphism of Progesterone: Stabilization of a 'Disappearing' Polymorph by Co-Crystallization. *J. Pharm. Sci.* **2007,** *96* (12), 3419-3431.

214. Gavezzotti, A. The Calculation of Molecular Volumes and the Use of Volume Analysis in the Investigation of Structured Media and of Solid-State Organic-Reactivity. *J. Am. Chem. Soc.* **1983,** *105* (16), 5220-5225.

215. The Free Dictionary defines herringbone as a way of assembling, in a diagonal zigzag manner.

216. Desiraju, G. R. On the presence of multiple molecules in the crystal asymmetric unit (Z ' > 1). *CrystEngComm* **2007,** *9* (1), 91-92.

217. Anderson, K. M.; Steed, J. W. Comment on "On the presence of multiple molecules in the crystal asymmetric unit (Z ' > 1)" by Gautam R. Desiraju, CrystEngComm, 2007, 9, 91. *CrystEngComm* **2007,** *9* (4), 328-330.

218. Delaney, S. P.; Smith, T. M.; Pan, D.; Yin, S.; Korter, T. M. Cryogenic Phase Transition in Crystalline Aripiprazole Leads to an Eighth Polymorph. *Cryst. Growth Des.* **2014**.

219. From private communication with Rona Watson at University College London.

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

221. Mo, Y.; Gao, J. Theoretical Analysis of the Rotational Barrier of Ethane. *Accounts Chem. Res.* **2007,** *40* (2), 113-119.

222. Pophristic, V.; Goodman, L. Hyperconjugation not steric repulsion leads to the staggered structure of ethane. *Nature* **2001,** *411* (6837), 565-568.

223. Halpern, A. M.; Glendening, E. D. An intrinsic reaction coordinate calculation of the torsional potential in ethane: Comparison of the computationally and experimentally derived torsional transitions and the rotational barrier. *Journal of Chemical Physics* **2003,** *119* (21), 11186-11191.

224. Mulliken, R. S. Intensities of Electronic Transitions in Molecular Spectra IV. Cyclic Dienes and Hyperconjugation. **2004,** *7* (5), 339-352.

225. Mo, Y.; Song, L.; Lin, Y. Block-Localized Wavefunction (BLW) Method at the Density Functional Theory (DFT) Level. *J. Phys. Chem. A* **2007,** *111* (34), 8291-8301.

226. Brunck, T. K.; Weinhold, F. Quantum-Mechanical Studies on the Origin of Barriers to Internal-Rotation About Single Bonds. *J. Am. Chem. Soc.* **1979,** *101* (7), 1700-1709.

227. Pendas, A.; Blanco, M.; Francisco, E. Steric Repulsions, Rotation Barriers, and Stereoelectronic Effects: A Real Space Perspective. *J. Comput. Chem.* **2009,** *30* (1), 98-109.

228. Surov, A. O.; Perlovich, G. L.; Emel'yanenko, V. N.; Verevkin, S. P. Thermochemistry of Drugs. Experimental and First-Principles Study of Fenamates. *J. Chem. Eng. Data* **2011,** *56* (12), 4325-4332.

229. *GaussView*, version 5; 2009

230. Clark, M.; Cramer, R. D.; Van Opdenbosch, N. Validation of the general purpose tripos 5.2 force field. *J. Comput. Chem.* **1989,** *10* (8), 982-1012.

231. Gavezzotti, A.; Filippini, G. Geometry of the Intermolecular X-H...Y (X, Y=N, O) Hydrogen- Bond and the Calibration of Empirical Hydrogen-Bond Potentials. *J. Phys. Chem.* **1994,** *98* (18), 4831-4837.

232. Gavezzotti, A. Are Crystal-Structures Predictable? *Accounts Chem. Res.* **1994,** *27* (10), 309-314.

233. *Mercury,* version 3.1; 2013

234. Lozano, J. J.; Pouplana, R.; López, M.; Ruiz, J. Conformational analysis of the antiinflammatory fenamates: a molecular mechanics and semiempirical molecular orbital study. *Journal of Molecular Structure: THEOCHEM* **1995,** *335* (1–3), 215-227.

235. Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. Uff, a Full Periodic-Table Force-Field For Molecular Mechanics and Molecular-Dynamics Simulations. *J. Am. Chem. Soc.* **1992,** *114* (25), 10024-10035.

236. Jorgensen, W. L.; Tirado-Rives, J. The OPLS potential functions for proteins, energy minimizations and crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988,** *110* (6), 1657-1666.

237. Hopkins, C. W.; Roitberg, A. E. Fitting of Dihedral Terms in Classical Force Fields as an Analytic Linear Least-Squares Problem. *J. Chem. Inf. Model.* **2014,** *54* (7), 1978-1986.

238. Wheatley, R. J.; Price, S. L. An Overlap Model For Estimating the Anisotropy of Repulsion. *Mol. Phys.* **1990,** *69* (3), 507-533.

239. Totton, T. S.; Misquitta, A. J.; Kraft, M. A First Principles Development of a General Anisotropic Potential for Polycyclic Aromatic Hydrocarbons. *J. Chem. Theory Comput.* **2010,** *6* (3), 683-695.

240. Sorescu, D. C.; Rice, B. M.; Thompson, D. L. Intermolecular potential for the hexahydro-1,3,5-trinitro-1,3,5-s- triazine crystal (RDX): A crystal packing, Monte Carlo, and molecular dynamics study. *J. Phys. Chem. B* **1997,** *101* (5), 798-808.

241. Williams, D. E.; Houpt, D. J. Fluorine Nonbonded Potential Parameters Derived From Crystalline Perfluorocarbons. *Acta Crystallogr. , Sect. B* **1986,** *42* (JUN), 286-295.

242. O'Boyle, N.; Banck, M.; James, C.; Morley, C.; Vandermeersch, T.; Hutchison, G. Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **2011,** *3* (1), 33.

243. Combinations - Rosetta Code.

244. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition;* 2nd ed. 2009. Corr. 7th printing 2013 edition ed.; Springer: New York, 2011.pp. 745.

245. Schmidt, M. U.; Dinnebier, R. E.; Kalkhof, H. Crystal engineering on industrial diaryl pigments using lattice energy minimizations and X-ray powder diffraction. *J. Phys. Chem. B* **2007,** *111* (33), 9722-9732.

246. Ahmed, S.; Bidstrup, S. A.; Kohl, P. A.; Ludovice, P. J. Development of a new force field for polynorbornene. *J. Phys. Chem. B* **1998,** *102* (49), 9783-9790.

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

248. Ismail, S. Z.; Anderton, C. L.; Copley, R. C.; Price, L. S.; Price, S. L. Evaluating a Crystal Energy Landscape in the Context of Industrial Polymorph Screening. *Cryst. Growth Des.* **2013,** *13* (6), 2396-2406.

249. Price, L. S.; McMahon, J. A.; Lingireddy, S. R.; Lau, S.-F.; Diseroad, B. A.; Price, S. L.; Reutzel-Edens, S. M. A molecular picture of the problems in ensuring structural purity of tazofelone. *J. Mol. Struct.* **2014,** *ASAP*.

250. Kendrick, J.; Stephenson, G. A.; Neumann, M. A.; Leusen, F. J. Crystal Structure Prediction of a Flexible Molecule of Pharmaceutical Interest with Unusual Polymorphic Behavior. *Cryst. Growth Des.* **2013,** *13* (2), 581-589.

251. Price, S. L.; Hamad, S.; Torrisi, A.; Karamertzanis, P. G.; Leslie, M.; Catlow, C. R. A. Applications of DL_POLY and DL_MULTI to organic molecular crystals. *Mol. Simulat.* **2006,** *32* (12-13), 985-997.

252. Abramov, Y. A. QTAIM Application in Drug Development: Prediction of Relative Stability of Drug Polymorphs from Experimental Crystal Structures. *J. Phys. Chem. A* **2011,** *115* (45), 12809-12817.

253. Campeta, A. M.; Chekal, B. P.; Abramov, Y. A.; Meenan, P. A.; Henson, M. J.; Shi, B.; Singer, R. A.; Horspool, K. R. Development of a Targeted Polymorph Screening Approach for a Complex Polymorphic and Highly Solvating API. *J. Pharm. Sci.* **2010,** *99* (9), 3874-3886.

254. Klimes, J.; Michaelides, A. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **2012,** *137* (12), 120901.

255. Vasileiadis, M.; Pantelides, C. C.; Adjiman, C. S. Prediction of the crystal structures of axitinib, a polymorphic pharmaceutical molecule. *Chem. Eng. Sci.* **2015,** *121*, 60-76.

256. Smith, W.; Todorov, I. T.; Leslie, M. The DL_POLY molecular dynamics package. *Z. Kristallogr.* **2005,** *220* (5-6), 563-566.

257. Leslie, M. DL_MULTI - A molecular dynamics program to use distributed multipole electrostatic models to simulate the dynamics of organic crystals. *Mol. Phys.* **2008,** *106* (12-13), 1567-1578.

258. Koch, U.; Popelier, P. L. A.; Stone, A. J. Conformational Dependence of Atomic Multipole Moments. *Chem. Phys. Lett.* **1995,** *238* (4-6), 253-260.