Non-empirical Force-Field Development for Weakly-Bound Organic Molecules

Alexander Aina

Molecular Modelling and Materials Science EngD, Department of Chemistry, University College London

May 2020

Supervisor: Professor Sarah L Price

I, Alexander Aina, confirm that the following work presented in this thesis is entirely my own. I can confirm that any information that has been derived from other sources has been appropriately referenced and indicated within this thesis.

Alexander A A Aina

Abstract

This thesis pioneers the development of non-empirical anisotropic atom-atom force-fields for organic molecules, and their use as state-of-the-art intermolecular potentials for modelling the solid-state. The long-range electrostatic, polarization and dispersion terms have been derived directly from the molecular charge density, while the short-range terms are obtained through fitting to the symmetry-adapted perturbation theory (SAPT(DFT)) intermolecular interaction energies of a large number of different dimer configurations. This study aims to establish how far this approach, previously used for small molecules, could be applied to specialty molecules, and whether these potentials improve on the current empirical force-fields FIT and WILLIAMS01. The scaling of the underlying electronic structure calculations with system size means many adaptions have been made. This project aims to generate force-fields suitable for use in Crystal Structure Prediction (CSP) and for modelling possible polymorphs, particularly high-pressure polymorphs. By accurately modelling the repulsive wall of the potential energy surface, the high pressure/temperature conditions typically sampled by explosive materials could be studied reliably, as shown in a CSP study of pyridine using a non-empirical potential. This thesis also investigates the transferability of these potentials from the gas to condensed-phase, as well as the transferability and importance of the intermolecular interactions of flexible functional groups, in particular NO_2 groups. The charge distribution was found to be strongly influenced by variations in the observed NO_2 torsion angle and the conformation of the rest of the molecule. This conformation dependence coupled with the novelty of the methods and size of the molecules has made developing non-empirical models for flexible nitro-energetic materials very challenging. The thesis culminates in the development of a bespoke non-empirical force-field for rigid trinitrobenzene and its use in a CSP study.

Impact statement

Computer simulations of the properties of molecular organic materials requires an accurate modelling of the intermolecular interactions (interactions between neighbouring molecules) in the crystal. The potential energy surface of the molecules needs to be concisely modelled by a force-field that can be used in Molecular dynamics, Crystal structure prediction and other computer simulations, which correctly describes the interactions between molecules. This thesis is of greatest relevance in the development of computational studies of pharmaceutical and energetic materials. For energetic materials, it is critical that the repulsive wall is well represented by the force-field, in order to model the effects of high pressures and temperatures, which is a region that cannot be reliably extrapolated to or determined by force-fields fit to experimental data.

This thesis uses a comprehensive theory of intermolecular forces, as implemented in the code CamCASP, to derive state-of-the-art non-empirical force-fields, in an anisotropic atom-atom analytical form, from the charge density of the isolated molecules. These force-fields are tested for their ability to predict the gas-phase two-molecule interactions of small molecules and the solid-state behaviour of these molecules in their crystal structures.

Thus this thesis:

i. identifies a suitable workflow (software and parameters) for the generation of accurate atomistic force-fields,

ii. uses these accurate atomistic force-fields for the prediction of the crystal structures of pyridine and trinitrobenzene,

iii. validates the force-field and modelling by comparing the predictions to the known solid-state structures and properties of pyridine and trinitrobenzene,

iv. investigates the effects of the changes in conformation of nitro-groups on the charge density and intermolecular forces of a range of small energetic molecules.

This thesis shows that these non-empirical force-fields describe the experimental crystal structures as accurately as empirically fitted force-fields. In both cases, the differences are comparable to the changes in crystal structure with temperature, pressure or neglect of quantum effects. Crystal structure prediction studies generate the observed polymorphs and hypothetical thermodynamically competitive crystal structures. The non-empirical potential was able to predict a high-pressure phase of pyridine.

It is hoped that this approach to developing force-fields will be applied to other small organic molecules, and used to predict their properties in condensed phases.

Acknowledgements

I would like to thank my supervisors Professor Sarah (Sally) Price and Dr. Alston Misquitta for their insight, guidance and supervision through what has been a very long 5 years. I'd like to thank Sally for being like my second mother, guiding me and keeping me safe whenever I was losing sight of my goals or facing a roadblock. To quote her "*All PhDs are not made equal*", and I could not agree more. I'd like to thank Alston for being an amazing life advisor (older brother one could say), for re-enforcing the importance of a solid theoretical foundation, and for his positivity and enthusiasm towards the field of quantum physics. I am grateful for the many motivational discussions we have had, and for a proper introduction to the theory of intermolecular forces and how to develop these innovative non-empirical force-fields.

I thank the rest of the Price group, Dave, Gerit, Louise, Luca, Rebecca & Rui, for their support. Especially Louise for her help and warmth throughout this PhD. A special thanks to Dr. Maurice Leslie for developing DMACRYS to use anisotropic repulsion terms and minimizations including polarization. I'd also like to thank Profs Adjiman & Pantelides for the use of CrystalPredictor. I thank Prof Simon Parsons for the challenge of identifying the pyridine $d5$-III structure, and Dr. Jonas Nyman for the scripts AutoFREE and AutoLD for carrying out the lattice dynamics calculations.

I would like to thank Matthew Habgood, Andrew Willets, Max Phipps & Mark Storr of A.W.E. for supervising me at various points in this PhD. I also thank A.W.E. for the financial support through the EngDoc studentship from Molecular Modelling and Materials Science Centre for Doctoral Training.

I'd like to thank my homies (Dan, Ed, John, Loi, Omar & Raunaq) for their endless support and love, and most importantly I would like to dedicate this thesis to my family. My mum, my dad and my sister. Thank you for supporting me through this PhD and helping me become the first in the family to do one!

Table of Contents

Abbreviations

AC, Asymptotic correction;

AI, artificial intelligence;

anarot, multipole moments analytically rotated from the optimized structure’s local axis into the local axis of the experimental conformation;

BSSE, basis set superposition error;

BS-ISA, Basis-space Iterated Stockholder Atoms;

CCDC, Cambridge Crystallographic Data Centre;

CCSD(T), Coupled Cluster Theory;

CSD, Cambridge Structural Database;

CSP, Crystal Structure Prediction;

CHELP, CHarges from Electrostatic Potentials using a Grid-based method;

DC, dimer centred basis set;

DFT, Density Functional Theory;

DFT-D, DFT including a dispersion correction;

DFT-TB, DFT Tight-binding;

DIFF, Distributed intermolecular force-field;

DMA, Distributed Multipole Analysis;

FOX-7, 1,1-diamino-2,2-dinitroethene (DADNE);

GDMA, Gaussian distributed multipole analysis;

GTO, Gaussian-type orbital;

$h_{50\%}$, impact sensitivity;

HF, Hartree-Fock;

HI, Hirshfeld-iterative stockholder approach;

HNB, hexanitrobenezene;

HNIW, Hexanitrohexaazaisowurtzitane (CL-20);

HOMO, Highest Occupied Molecular Orbital

IHE, Insensitive high explosives;

IP, Ionization Potential;

ISA, Iterated Stockholder Atoms;

KDE, Gaussian Kernel Density Estimate

MBD, many-body dispersion correction;

MC, monomer centred basis set;

MD, Molecular Dynamics;

Model0, non-empirical model fit to 1st order TNB dimer interaction energies;

MP, Møller-Plesset;

MPA, Mulliken Point Analysis;

NPT, Isothermal-isobaric;

NN, Neural Network;

optexptNO$_2$, multipole moments calculated in for the optimized structure with all nitro-groups in their experimentally observed torsion angles;

PCM, Polarizable continuum model;

PBE, Perdew-Burke-Ernzerhof general gradient approximation

PBE0, PBE combined with a portion of the exact HF exchange-correlation

PES, potential energy surface;

$_{pyr}$#, CSP generated, putative, mechanically stable crystal structure of pyridine;

RDX, 1,3,5-trinitroperhydro-1,3,5-triazine;

$RMSD_n$, the root mean square deviation, the optimum overlay of n molecules in two crystal structures excluding the hydrogen atoms.

R.T.P, room temperature and pressure;

RS Perturbation Theory, Raleigh-Schrodinger Perturbation Theory;

SAPT, Symmetry-Adapted Perturbation Theory

SAPT(DFT), Symmetry-Adapted Perturbation Theory based on a coupled Kohn-Sham Treatment of monomers

TNB, 1-3-5 trinitrobenzene;

$_{TNB}$#, CSP generated, putative, mechanically stable crystal structure of TNB;

TNT, trinitrotoluene;

TS, Tkatchenko-Scheffler method;

vdW, van der Waals

V_{max}, electrostatic potential surface maximum;

V_{mid}, trigger-bond (NO_2) polarity;

V_{min}, electrostatic potential surface minimum;

wRMSD, weighted root mean square deviation;

WSM, William-Stone-Misquitta method;

xminexpt, a lattice energy minimization, where the monomers are in their experimentally observed conformations;

xminopt, a lattice energy minimization, where the monomers are held in their isolated optimized molecular conformations;

Z', number of independent molecules in the asymmetric unit cell.

Mathematical symbols

R_{ik}, inter-atomic distance between interacting atoms i and k;

Ψ, wavefunction;

$\rho_M(r)$, the charge distribution of molecule M;

v, electrostatic interaction operator;

$\mathcal{H}'$, intermolecular interaction operator;

$\mathcal{H}_M^0$, Hamiltonian of the unperturbed ground state of molecule M;

E_M^0 is the unperturbed ground state energy of the unperturbed wave-function Ψ_M^0 of molecule M;

r, Cartesian vector $\mathbf{r} = \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix}$;

$\mathrm{T}_{\mathrm{tu}}^{\mathrm{ik}}(\mathrm{R}_{\mathrm{ik}}, \Omega_{\mathrm{ik}})$, orientation and distance transformation tensor;

Q_t^i, multipole moment on atomic site i of rank t;

C_{2n+6}, computed or empirically fit dispersion coefficients;

$\rho^{\iota\kappa}(\Omega_{ik})$, shape function of the pair of atom types ι and κ;

Ω_{ik}, relative orientation of atoms i and k;

E_{latt}, the lattice energy;

E_{int}, the intermolecular energy;

E_{intra}, the intramolecular energy;

$G_{free}(T)$, Gibbs free energy;

$A_{free}(T)$, Helmholtz free energy;

PV, Pressure-Volume term

TS, energy contribution due to the entropy of the system;

ZPE, zero-point energy;

$U_{therm}(T)$, internal energy of the system due to thermal effects

H_{latt}, lattice enthalpy;

$F_{vib}(T)$, free energy thermal correction;

E_{LR}, long-range energy;

E_{SR}, short-range energy;

$E_{elst}^{(1)}$, 1st order electrostatic energy;

$E_{pen}^{(1)}$, 1st order penetration energy;

$E_{exch-rep}^{(1)}$, 1st order exchange-repulsion energy, the subscript '$resp$' is included to indicate that coupled perturbation is used. $E_{exch}^{(1)}$ is also used in this thesis to indicate the exchange-repulsion energy;

$E_{IND}^{(2)}$, total 2nd order induction energy;

$E_{IND}^{(2)}(Reg)$, 2nd order regularized induction energy (polarization energy);

$E_{exch-ind}^{(2)}$, the electron exchange portion of the induction energy, the subscript '$resp$' is included to indicate that coupled perturbation is used;

$E_{POL}^{(2)}$, 2nd order polarization energy, a component of the induction energy;

$E_{CD}^{(2)}$, 2nd order charge-delocalization energy, also referred to as the charge-transfer energy;

$E^{(2)}_{DISP}$, total 2nd order dispersion energy;

$E^{(2)}_{disp}$, 2nd order dispersion energy, the subscript 'resp' is included to indicate that coupled perturbation is used;

$E^{(2)}_{exch-disp}$, the electron exchange portion of the dispersion energy;

δ^{HF}_{int}, higher-order induction correction;

$E^{(1)}_{int}[SAPT(DFT)]$, 1st order SAPT(DFT) dimer interaction energies

$E^{(2)}_{int}[SAPT(DFT)]$, 2nd order SAPT(DFT) dimer interaction energies

$V^{MN}_{int}(R_{ik})$, intermolecular interaction potential energy between bodies M and N;

V^{MN}_{LR}, long-range potential energy between molecules M and N;

V^{MN}_{SR}, short-range potential energy between molecules M and N;

$V^{(1)}_{elst}[DM]$, long-range multipolar electrostatic potential energy;

$V^{(2)}_{disp}[DM]$, long-range multipolar dispersion energy;

$V^{(2-\infty)}_{POL}[DM]$, long-range iterated multipolar polarization energy;

ΔQ^{ι}_{t}, the induced moment on atom i depends on its atomic polarizability ($\alpha^{\iota\iota'}_{tt'}$) and the net electric field (V^{ι}_{t});

β_{disp}, dispersion model damping parameter;

β_{pol}, polarization model damping parameter;

$S^{MN}(\boldsymbol{R}) = \int \rho^{M}(R)\rho^{N}(R)\,dR$, density overlap integral between molecules M and N;

$B_2(T)$), 2nd virial coefficients;

1 Introduction

1.1 Polymorphs & Crystal Structure Prediction (CSP)

1.1.1 Polymorphism

Crystallography will forever play a defining role in many areas of chemistry; through a solid comprehension of a species' crystal make-up one could forecast most physical and chemical behaviours of the material. From solubility, melting point, morphology, shock and mechanical stress responses, surface chemistry, reactivity and phase transformations, many characteristics of a species are often governed by its lattice structure. In industries like pharmaceuticals, organic photovoltaics and energetics,[1, 2] a great attraction is the ability to fully comprehend the behaviour of the organic molecular crystal under all probable conditions, reducing the likelihood of producing products that change forms while on the shelf or in use.[3-5] Over the past few decades, computational methods have been combined with experimental approaches in the determination of crystalline morphology and behaviour. With the ever increasing development of quantum chemistry and improvement of computing power, computational treatments are becoming progressively more powerful. Improving the accuracy and reliability of computational methods employed is of uttermost importance, especially in polymorphism studies due to the close relative energies of polymorphs, which is on the order of 2-5 kJ mol^{-1}.[6, 7] Accuracy and reliability is also important for specialty materials like energetics and pharmaceuticals, where experimental analysis can be difficult, due to the inherent dangers and costs.

A single chemical species can adopt a number of stable crystal structures depending on the crystallisation conditions.[6-8] These chemically equivalent structures are known as polymorphs and have become a hot topic of discussion over the past few decades. While there are a number of descriptions of polymorphism in literature,[6] this thesis defines polymorphism as *"Chemically-equivalent lattices with different crystal packing motifs of the molecule but identical liquid/vapour phases*". The condensed phases of organic molecules are an area of increasing interest, with the ability to simulate the differences in material properties between polymorphs being highly desirable for the specialty chemical industries.[9-11] With over 50% of materials suspected to exhibit polymorphism,[7] there are many examples of polymorphism in well-known substances, for instance the desired form of popular food additive L-glutamic acid is the α polymorph, however, the β polymorph can form if the conditions vary.[6] Infamous examples of polymorphism in pharmaceuticals include the patent litigation of the drug Zantac[6, 12, 13] and the temporary removal of the drug Ritonavir from production due to the discovery that it could transform into another polymorph while on the shelf.[4, 5, 14] It is easy to see the benefits of polymorph prediction for new materials, like not spending resources synthesising a molecule whose crystals do not exhibit the desired functional properties.

In the energetic materials industry there has always been a drive to improve safety. The sensitivity of an energetic material is not determined solely by the density but is related to a number of properties such as the composition, structure and the applied stimulus. A comprehensive knowledge of the sensitivity of each polymorph to different stimuli could reduce the risk posed by these materials. By reducing the sensitivity of energetic materials there is a reduced risk of accidental detonation from thermal or pressure shocks. Crystal Structure Prediction (CSP) is particularly important for energetic materials. CSP could in principle be used to develop new high performance materials with low sensitivity as structural properties like density influence characteristics such as sensitivity to mechanical and thermal shock, the evolution of hot spots and the detonation pressure and velocity.[15-17]

1.1.2 Crystal structure Prediction (CSP) studies & methodologies

1.1.2.1 Introduction

Crystal Structure Prediction (CSP) has emerged in the past decade or so, as a method of generating all probable, mechanically stable crystal structures of a molecule and ranking them in order of thermodynamic stability, using only the chemical diagram of the molecule that defines its atoms and covalent bonding.[18-20] In general, to calculate many crystal properties, accurate descriptions of the intermolecular and intramolecular interactions are required, and electronic structure methods (**1.3**) or atomistic potential energy functions also known as force-fields (**1.2**) are normally utilized. Unfortunately, the widely available transferable force-fields for organic molecules are usually not accurate enough for realistic crystal structure prediction.[21] Being able to generate custom force-fields for organic molecules is one of the main pursuits of CSP, and one of the most comprehensive tests of a newly generated force-field is its ability to predict the structures of the molecule’s polymorphs and their correct stability ranking.

One could say the zenith of crystal structure prediction studies would be reached if for any chemical species it was possible to:

- generate all probable, mechanically stable crystal structures,
- predict kinetically favoured structures or conditions under which one might be preferentially formed,
- calculate the intrinsic/extrinsic properties of the species,
- and subsequently simulate their physical behaviour under all conditions,

using solely its molecular formula and computational methods. As the world becomes progressively conscious of its finite resources, reducing the degree of experimentation in the development of novel materials is now vital and becoming a priority of many organisations in the field of science and technology.

1.1.2.2 The CSP treatment in this thesis

The CSP study in this thesis is based on total lattice energy E_{latt}, which is the difference between the internal energy of the system and the infinitely separated molecules in their lowest energy configuration. The total lattice energy can be separated into atom-atom intermolecular (E_{int}) interactions and intramolecular (geometry distortion, $\Delta E_{intra}^{molecule}$) energies of the molecules ($E_{latt} = E_{int} + \Delta E_{intra}^{molecule}$). If the molecular structures are held rigid, $\Delta E_{intra}^{molecule} = 0$ and the lattice energy composed solely of the intermolecular interaction energy. There are a number of methodologies for conducting CSP studies on unit cells (the smallest repeating unit in a crystal structure), as shown by the numerous blind tests,[22-27] such as using electronic structure methods, force-fields or a combination of both to explore the multidimensional, intermolecular interaction potential energy surface (PES) of a crystalline lattice. The number of dimensions (cell length and angles) depends on whether the molecule is held rigid or kept flexible, and the complexity of the search can increase dramatically for flexible molecules.[28, 29] The broad nature of CSP and the stringent requirement for accurate yet reliable methods is what makes it such a challenging field. Obtaining the true lattice energy is no easy feat and complexity, issues with implementation and computational cost means that a dependable CSP study on very large molecules is not yet possible.[30] Some approximations can be used to circumvent some of these challenges, like holding the molecule rigid and conducting a $Z' = 1$ search. It should be noted that in this thesis for systems where the molecule is held rigid, the terms potential, force-field and model refer to mathematical descriptions of the intermolecular potential energy surface and are interchangeable.

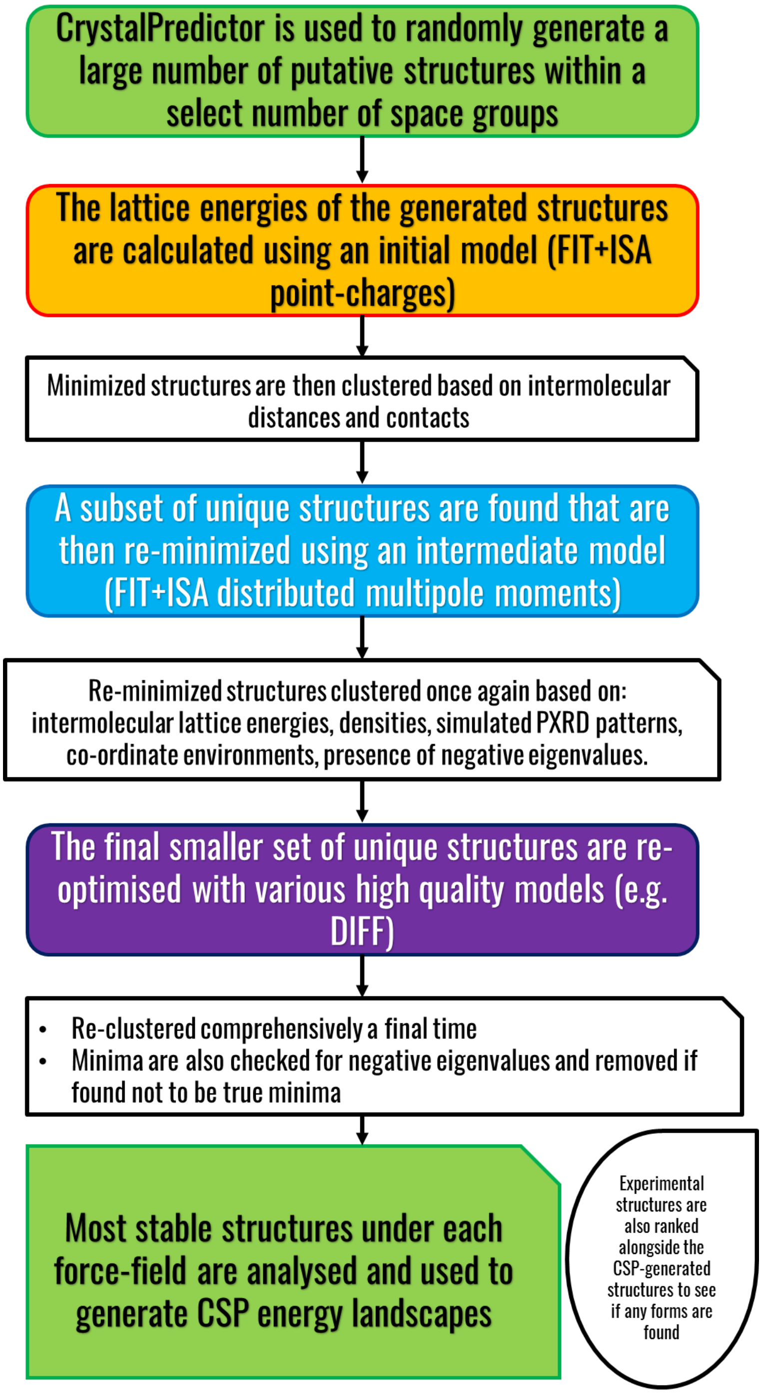


Figure 1: *An overview of the workflow for the pyridine CSP study in* **Chapter 3***. Python*[31] *and Bash scripting was used to develop this example workflow and automate the generation of suitable input files and directories for each structure, and then the mass-minimization, analysis, sorting, clustering, and ranking of the generated structures.*

1.1.2.3 Conformational flexibility

The effects of conformation on the intramolecular energy, the molecular charge distribution, the intermolecular interactions within the lattice, and ultimately the potential energy landscape are explored in detail in **Chapter 4**. An initial assessment of the conformational flexibility is a vital first step when performing a CSP search.[22] It gives insight into the dimensionality of the PES and also permits us to make approximations, which can greatly speed up the study without hindering accuracy. The dimensionality and cost of a search dramatically increases with the number of molecular conformations that might occur in the crystal structure.[29] Knowledge of which are the most important can greatly optimize the search (**Chapter 4**). Pyridine (**Chapter 3**) is a small molecule with no flexible functional groups, whose optimized structure is virtually planar and can be held rigid in its CSP study. Thus, the effects of conformational variation do not have to be investigated in as much detail as with the nitro-group energetics (**Chapter 4**).

It is important to understand the flexibility of functional groups within the molecule prior to the CSP search, and this is typically done by either calculating the relative intramolecular energies of a range of conformations of the isolated molecule using ab initio treatments, or by studying observed crystal structures within the Cambridge Structural Database (CSD).[32] The best treatments use both as this prevents biasing the study towards experimental data and allows access to possible unique structures.[33] The quantum chemical code NWCHEM was used to do so in **Chapter 4** to conclude that for the flexible energetic molecule TNB, it would be possible hold its molecular structure in its rigid planar conformation and still conduct a meaningful CSP search.

1.1.2.4 Crystal structure generation

For a rigid CSP study in the Price group,[19, 34] the program CrystalPredictor[35, 36] is used to generate a large number of putative $Z' = 1$ crystal structures (generally, on the order of a million) within a given energy range of the global minimum. This is typically done within the most probable space groups, but is dependent on the molecular crystal being studied (e.g. a sample of one enantiomer of a chiral molecule can only crystallise in a smaller number of chiral space groups). The structure generation method is tailored to the molecule. CrystalPredictor uses a quasi-random low-discrepancy Sobol Sequence[37], and Shoemake's algorithm[38] to initially search the input molecule's potential energy surface efficiently, as a function of cell lengths and angles. This method uses quaternion space to guarantee a uniform yet random sampling when generating initial structures.[35] For a flexible search, bond lengths and angles can also be explored by using the conformation searching program CrystalOptimizer.[29, 39] The density functional theory (DFT) (**1.3.1** & **2.2**) optimized isolated molecular structure is used as the starting structure to generate putative crystal structures. The energies of these crystal structures are initially calculated using a cheap but approximate force-field like the FIT potential (**1.2.2.1**) and Iterated Stockholder Atoms (ISA) (**2.5.1.7**) point charges, and then clustered based on their intermolecular distances, resulting in a subset of unique structures, in the case of pyridine this was around 10000 unique structures, that are then refined with more accurate methods.

1.1.2.5 Refinement

Typically, the subset of unique structures is minimized and re-ranked using a more elaborate force-field, i.e. the FIT potential with ISA distributed multipoles instead of just a point charge model. The resultant structures are subsequently re-analysed and further clustered based on intermolecular lattice energies, densities per asymmetric unit, simulated powder x-ray diffraction (PXRD) patterns, co-ordinated environments. Normally if negative eigenvalues of the energy were observed these structures would be re-minimised removing a symmetry element to find their true minima as higher Z' structures. However, due to limitations when using these new non-empirical models, the CSP searches in this thesis are all $Z' = 1$. These final set of structures are used as the core structures for searching the potential energy landscapes of the empirical and non-empirical intermolecular potentials using the more complete higher-order distributed multipole moments. Many other CSP workflows will use a hierarchical approach, doing lattice energy minimizations with increasingly accurate force-fields, and then finally using electronic structure methods, like DFTB3-D3 for the final stages of refinement of a small number of the most stable structures to optimize the lattice energy.[40-44] However, in this thesis, the crystal structures are minimized using solely intermolecular force-fields as a test of the state-of-the-art anisotropic atom-atom non-empirical potentials (**Chapters 3** & **6**). The most stable structures within a given energy range are then analysed in detail for their similarities to each other and gas-phase dimers.

1.2 Force-field development

1.2.1 Introduction

A sufficiently accurate analytical pair potential to account for all the physical properties of the inert gases was achieved in the 1980's.[45] These potentials also included the three-body dispersion term for the condensed phases. Water and other small polyatomic molecules now have very accurate potentials available, though these are often not used in Molecular Dynamics (MD) simulations because of the need to compromise between accuracy, speed of evaluation and the functional forms assumed in suitable MD codes.[46-56] In many atomistic force-fields, the interaction energy (E_{int}) is commonly given as an approximated sum of "pairwise" potentials, which only include two-body terms and ignore all many-body contributions to the lattice energy. The pairwise construction of the interaction energy is elaborated in **Chapter 2**, however, it is briefly discussed here as it is required for a proper discussion of force-fields and their use in crystal structure prediction (CSP).

The interaction energy between a pair of molecules M and N can be expressed approximately as sum over each atom-atom interaction between the two. The resultant intermolecular interaction potential energy ($V_{int}^{MN}(R_{ik})$) is given by the sum of all pairs of atoms i in molecule M with atoms k in molecule N, which have an inter-atomic distance of R_{ik}. Higher order contributions arising from 3-molecule, 4-molecule or larger interactions can be included as corrections, where $\Delta V_{MNO} + \Delta V_{MNOP}$ are the approximated triple-molecule and quadruple-molecule contributions for molecules M, N, O & P. These non-pairwise terms are generally absorbed as an average into the potential through empirically-fitted parameters, and often produce satisfactory results. Higher-order contributions have been found to contribute to approximately 30% of the interaction energy for water molecules in clusters.[52, 57, 58] Polarisation is the main contribution to the higher-order terms, and it is now possible to explicitly include the intermolecular polarisation energy in modelling organic crystal structures.[40, 59-61] A few studies suggest that the polarisation energy may be important to obtain the accuracies and reliability required for polymorph studies.[61-63]

1.2.2 Empirical force-fields

There are a number of empirical force-fields for organic molecular crystals,[19, 64-66] parameterised to describe various types of organic functional groups. In this thesis the FIT potential[67, 68] is the main empirical potential of choice, however, in **Chapter 3,** the Williams' potential[69, 70] is also employed. It is a force-field based on fitting experimental data with atomic types for the same atom in different hybridisation states (**1.2.2.2**). These empirical models utilize a fully isotropic short-range repulsion-dispersion description known as the exp-6 Buckingham potential,[71] which also includes an electrostatic model (normally GDMA (**2.5.1.3**)) to obtain the intermolecular interaction energy of two molecules M and N, from the summation of the atoms i in M and k in N of types ι and κ respectively.

$$V_{int}^{MN}(R_{ik}) = \sum_{i\in M,\ k\in N} A_{\iota\kappa}\exp(-B_{\iota\kappa}R_{ik}) - \frac{C_6^{\iota\kappa}}{R_{ik}^6} + \text{ electrostatic } (Q_{l,m}^{ik}\ l \le 4,\text{GDMA}, \Psi)$$

1

where $A_{\iota\kappa}$, $B_{\iota\kappa}$ and $C_6^{\iota\kappa}$ are empirically derived parameters that are element/environment specific, with units of eV, Å^{-1} and $eV\text{Å}^6$ respectively. R_{ik} is the intermolecular inter-atomic distance between the two interacting atoms in Å. This Buckingham potential does not damp the dispersion and so may have an unphysical maximum at very short range. In addition, the polarization is not explicitly calculated and so absorbed in the empirical parameterisation of the repulsion-dispersion terms.

1.2.2.1 The FIT potential

The empirical FIT force-field[67, 68] does not foreshorten the hydrogen interaction sites and is a revision of Williams' W84 force-field,[72] which had each element in combination with C and H only. It is not clear whether Williams corrected for the X-ray foreshortening during the deriving of the FIT potential. Unsurprisingly, these empirically based force-fields produce the best results for crystal structures closely related to the parameterization set. These models were also tested in CSP studies using both a point charge model and a distributed multipole model and were found to be sensitive to the quality of electrostatic model used and likely to perform poorly for atypical short contacts e.g. high-pressure studies.[73, 74] The absorption of errors, approximations and assumptions in the experiments, simplifies the functional form of the potential but also simplifies the description of the interactions. It should be noted that the intermolecular force-fields that have been most widely used in simulating organic condensed phases have been derived by empirical fitting to crystal structures and properties[75, 76] with increasing sophistication being used for the challenge of modelling crystal structures[11, 70, 77] to the extent that they rival popular periodic DFT calculations in accuracy.[11] However, the assumed functional form and neglect of thermal effects among other quantum effects can result in these models performing poorly for more complex systems or when sampling regions of the potential energy surface not sampled by the fitted data.

1.2.2.2 The Williams potential

The FIT potential uses many of the earliest potential parameters suggested by the Williams' group.[67, 68] These were first developed to be transferable between hydrocarbon compounds then extended to aza-hydrocarbons, oxy-hydrocarbons, etc., but not used for molecules capable of forming hydrogen bonds.[69, 70] The W99 force-field, which describes interactions for organic crystals containing H, C and O was adapted to include N as well producing the WILL01 developed specifically for molecules containing nitrogen atoms.[69, 70] In WILL01, nitrogen environments in a system are divided into four groups dependent on the presence on hydrogen: $N(1)$ – for systems with triple bonded nitrogen e.g.N_2, $N(2)$ – to describe environments with nitrogen bonded to no hydrogens (excluded the triply bonded nitrogen), $N(3)$ – for nitrogen bonded to a singular hydrogen, and $N(4)$ – to describe nitrogen bonded to two or more hydrogens.

The potential also contains parameters for various carbon and hydrogen atoms, for example, the $H(4)$ type is for hydrogen bonded to nitrogen. Another defining feature of this potential is that the potential was parameterised with foreshortened hydrogen positions; that is, the hydrogen interaction sites are shifted by 0.1Å from their neutron or ab initio optimized positions in towards the heavier atom it is bonded to in an H-X bond. The lattice minimization code used in this thesis, DMACRYS, can perform this foreshortening. The electrostatic model must also be foreshortened. Williams found that foreshortening improved the performance of the potential in comparison to the FIT model.[69, 70] This can be explained by the shifting of the electron density in the bonds to hydrogen, which has no core electrons. However, it should be noted that depending on the partitioning method used and the quality of the electrostatic model, hydrogen bond lengths could be underestimated using WILL01, this was probably due to the fact that distributed multipoles can result in stronger electrostatic forces than the point charge model used in William's parameterisation. This problem was profound for carboxylate ($O \dots HO$ and $N \dots HO$) hydrogen bonds but this was not an issue for the pyridine study in **Chapter 3**.

WILL01 was derived via fitting to the experimental data set of 76 aza-hydrocarbon crystal structures, many of which were $Z' = 1$, and 11 heats of sublimation. Even though the parameterisation data set included a range of structures determined at temperatures extending from below liquid nitrogen to room temperature, a large portion of these experimental structures were observed at conditions very close to ambient. So one can expect the force-field to not have accommodated for a number of temperature and pressure effects. Likewise, a significant number of crystal structures that were determined at low temperature are liquid or gas at R.T.P. and it is unlikely that this empirical force-field could extend to predict their behaviour in the solid-phase. The WILL01 force-field has been tested on each aza-hydrocarbon used in the training data-set, predicting the unit cell lengths to 3% or less. It was then tested on a number of peptide and nucleoside structures, including 8 common nucleosides in DNA and RNA, 15 dipeptides four tripeptides, two tetra-peptides and a $Z' = 2$ pentapeptide.[70] While none of these molecules were included in the parameterization sets for the potential, the calculated crystal structures were in very good agreement with experiment, predicting most unit cell lengths within 2%.[70] This force-field was primarily constructed for use in biochemical studies, thus it may also prove to be apt in a pyridine study (**Chapter 3**) as pyridine appears as a precursor to a number of biological structures.[78]

1.2.3 Non-empirical Force-fields

1.2.3.1 Calculating intermolecular contributions

The problems with selecting experimental data and the assumptions used in fitting a potential to the data can be avoided by fitting a force-field to ab initio calculations.[79] This approach of using ab initio derived analytical potential energy surfaces (PESs) for spectra of gas-phase clusters, has been extensively applied to water,[46, 80-83] other small molecule complexes, [84 85, 86] and various ab initio astrophysics applications.[87-90] The development of suitable ab initio methods to generate a potential energy surface for dimers, trimers and other clusters of molecules to be used for fitting pair potentials, and higher-order corrections is an active area of research.[91] Ab initio treatments can be divided into two categories. Supermolecular methods, which aim to get the most accurate total energy for a cluster and subtract the energy of the isolated molecule contributions, and perturbation methods, which evaluate the different contributions to the intermolecular energy. The supermolecule (total energy) approach of aiming for the best (approximate) solution to the Schrodinger equation is essentially the same, whether applied to a cluster, or using periodic codes, to a lattice of molecules (**1.3.1**). The perturbation approach using symmetry adapted perturbation theory (SAPT) is the most practical for systems larger than water and small polyatomic molecules (**2.3.2**). The use of SAPT(DFT) calculations to produce anisotropic atom-atom intermolecular potentials is central to this thesis (**Chapters 3, 5** & **6**).

1.2.3.2 Novel force-fields

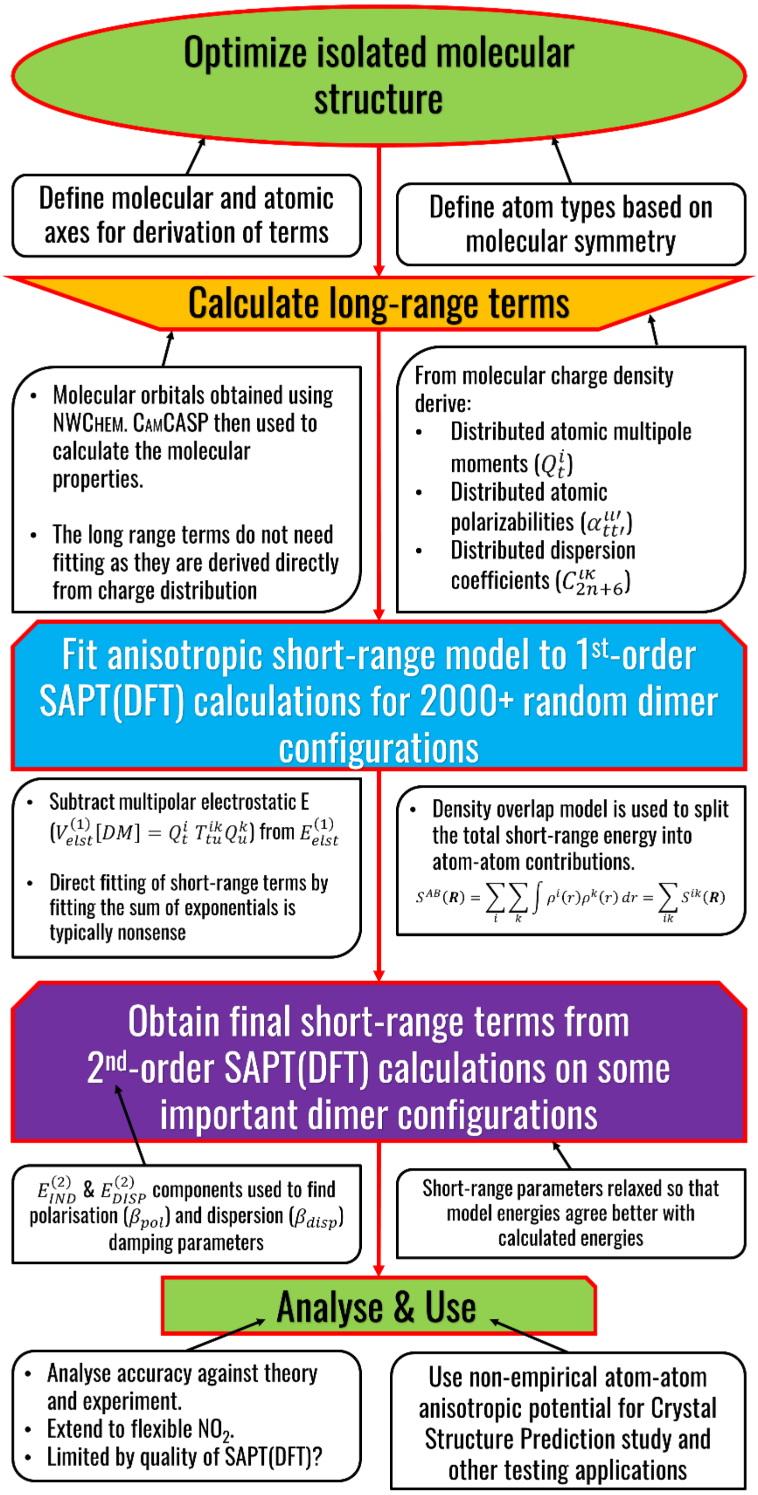


Figure 2: *The force-field development workflow that was used for trinitrobenzene and detailed in* ***Chapter 5****. It has been adapted from the workflow used on pyridine*[38] *and could be further adapted for other organic crystals. The workflow uses much of the theory discussed in* ***Chapter 2.***

Pioneering studies of the organic solid-state using analytical fits to SAPT(DFT) interaction energy calculations have mainly concentrated on energetic molecules,[92, 93] where predictive modelling requires reliable extrapolation into the repulsive region (regions of high temperature/pressure) that is not sampled in fitting empirical potentials. A non-empirical potential based on an early variant of the Distributed Intermolecular Force-Field (DIFF) approach for $C_6Br_2ClFH_2$,[60] was successful in predicting its crystal structure in the fourth Blind Test. It is therefore timely to assess the barriers to producing a sufficiently realistic model intermolecular potential from dimer calculations that can be used to predict the properties of a molecule in all phases, and assess the extent to which the exclusion of many-body terms is a limitation. Considerable advances have been made in the ability to calculate intermolecular pair potential energy surfaces accurately.[38, 79, 94-98] Computing sufficient number of points to define the PES leads to two challenges:

(1) the computational cost incurred, particularly for large molecules, and

(2) the need for sufficient accuracy.

Both needs can be satisfied by the SAPT(DFT)[77, 99-102] (**2.3.2.2**), which has now become the method of choice for many applications involving weakly-bound molecular interactions.[38, 60, 103-106] However, errors are introduced by the choice of analytical functional form and in the fitting process. Recently a method of automatically generating analytical intermolecular potential energy surfaces in an isotropic atom-atom functional form, to SAPT(DFT) numerical points reported a typical fit error of about 0.8 kJ mol^{-1} in the negative energy region.[79] Furthermore, the newly developed Slater-type Force-Field (Slater-FF) models exhibits average errors of 0.2 kJ mol^{-1} in the attractive region for wide range of interacting dimers.[96]

In this thesis, the alternative approach of using anisotropic atom-atom model potentials is explored, which are generated from distributed monomer properties and advanced fitting to the components of SAPT(DFT) dimer intermolecular interaction energies. This approach, using the suite of codes CamCASP has been described in detail and applied to the pyridine dimer.[38] This type of non-empirical distributed intermolecular force-field (DIFF) uses distributed multipolar models to describe the electrostatic, polarization and dispersion interactions, while the anisotropic atom-atom exponential short range terms are derived by fitting to SAPT(DFT) interaction energies.

While all these models represent considerable progress in accurate force-fields for dimers, it is important to determine how well they perform for the crystalline state. There are two issues that arise here, the first is that the lattice energy differences between 80% of polymorph pairs,[7] is less than 4 kJ mol^{-1}, with many observed polymorph energy differences being close to the reported errors in the above dimer potentials. The second issue is that so far these models have only been strenuously tested on gas-phase and calculated dimer properties, and there are many-body effects in condensed phases. A non-empirical model for the pyridine gas-phase, developed by Dr. Alston Misquitta,[38] is extended to the solid-state in **Chapter 3**, and the force-field functional form, a workflow (***Figure***

2) and treatments for developing the non-empirical distributed intermolecular force-fields (DIFFs) for the larger molecule trinitrobenzene are detailed in **Chapter 5**.

1.2.4 Computational chemistry software used in this thesis

A number of codes and interfaces are used in novel ways to generate the force-fields within this thesis. This has only been possible through careful debugging and a significant amount of Python[31] coding to process and reformat outputs, and also scale up processes that had previously been reserved for single-point calculations.

CAMCASP[107] is an intricate interface of a suite of programs created for computing interaction energies between gas-phase dimers, and the molecular properties (multipoles and frequency dependent atomic polarizabilities) of a material in single-site and distributed form to develop analytical intermolecular potentials. CAMCASP is used to derive and fit the non-empirical force-fields alongside the program ORIENT,[108] which is a program also developed to carry out various calculations on assembles of interacting molecules. NWCHEM is a quantum chemical software that is used alongside CAMCASP to computer the molecular properties of isolated molecules using quantum mechanical descriptions of molecular charge density. It is used in the force-field development process to obtain the electrostatic model and the subsequent long-range terms. It is also used in **Chapter 4** for the conformational analysis of energetic crystals. DMACRYS[59] is a program that uses atom-atom potentials, which include distributed multipoles, and other anisotropic terms, to calculate atomic contributions to the intermolecular lattice energies of organic crystal structures. Currently, the polarization module in DMACRYS can only do single-point polarization energy calculations and cannot calculate the forces due to polarization (**2.6.1.3**). This thesis pioneers developing and transferring the DIFF model using these codes to the solid-state phases of organic crystals. We seek to explore the issues involved in defining and using analytical force-fields that can transfer between phases.

1.3 Electronic structure methods

A variety of electronic structure methods are used in CSP studies. Electron density ($\rho(r)$) methods are typically less expensive than those that directly use the wave-function Ψ. Consequently, Periodic Density Functional Theory (DFT) methods, which had been traditionally reserved for small molecules, are becoming increasing prevalent in contemporary CSP. They have shown to be able to handle the subtleties found in intramolecular and intermolecular interactions with approaches like DFT-D[109] proving successful in the 6th Blind Test, where a record 12 groups used Periodic DFT methods.[22] The advantages of these methods are described next (**1.3.1**), and some detail on the theory of DFT is given in **Chapter 2**. However, the quality of DFT functionals that can be afforded, or even are available in periodic codes, is rather limited. Recent work[110-113] demonstrates that more expensive functionals such as PBE0 are needed as a refinement step. Hence, periodic calculations are done with a far worse molecular charge density description than that which can be used in developing intermolecular potentials or evaluating the conformational energy penalty (E_{intra}). Even when periodic electronic structure calculations become more accurate, electronic structure methods will have the disadvantage that they cannot be cheaply used for Molecular Dynamics (MD).

1.3.1 Density Functional Theory (DFT)

There are many examples of Periodic Kohn-Sham Density Functional Theory (Periodic DFT) being used in CSP.[114-117] With the optimization of codes and improvement in hardware, the computational cost is dropping at an astounding rate. Its growing popularity is due to the ability to calculate the intermolecular contributions like the induction and electrostatics without any need for parametrisation, experimental data or approximations of functional forms that can plague force-fields. That is not to say some DFT approaches do not use some corrections[118-120] to improve results.

The main drawbacks of DFT lie in it still being more expensive than current force-fields and its approximation of the exchange-correlation functional, E_{XC}. For organic molecular crystals correlation effects are important in capturing van der Waals dispersion interactions caused by associated fluxes between various regions of electron density, forming a long-range intermolecular interaction. In **Chapters 3** & **6** it is found that dispersive interactions can account for a large portion of the lattice energy. Standard DFT with its semi-local density functional approximations fails to properly describe these long range, non-local interactions. In response many functionals and corrections have been developed that use various approximations.

Dispersion corrected DFT (DFT-D)[22, 121] has become a popular solution as it includes a dispersion correction energy ΔE_{DISP} to E_{DFT} in order to account for E_{XC}. There are a few types of DFT-D non local functionals that include van der Waals dispersions (like vdW-DF1 & vdW-DF2) built-in and post-hoc corrections (like TS, MBD, and AC).[118-120] Some DFT-D corrections have been empirically fit to observed data,[119, 122] while more novel ones have been developed that are conformation dependent or derived from molecular polarizabilities.[123-125] There are also hybrid functionals available, which use part of the exact Hartree-Fock exchange to capture the missing exchange-correlation effects. Some hybrid functionals have been used to study energetic materials like B3LYP and B3LYP-BSSE (corrected for the basis set superposition error).[126, 127] For example, B3LYP and B3LYP-BSSE have been used to predict lattice energies of $-83.6\ kJ\ mol^{-1}\ and\ -44.4\ kJ\ mol^{-1}$ respectively, in comparison to the experimental value of $-130.2\ kJ\ mol^{-1}$ for the energetic crystal RDX.[128] The Wb97xd hybrid functional, which uses an empirically derived dispersion model (Grimme's Dispersion),[122, 129] has also been used and was found to perform well once corrected for BSSE. [128] Even with these corrections there is no systematic way of improving DFT, thus, treatments do not extend directly between systems. One that works amazingly for one system may fall flat on another. The contemporary goal is the development an accurate, transferable but computationally affordable depiction of all the intermolecular contributions in a system.

The scaling of DFT is ok, but the applied functionals are approximated and can result in a very poor description of the long-range correlations required to describe van der Waals dispersive interactions in crystal lattices like those found in energetic materials. There are methods available that contain an exact exchange-correlation, however, these are very expensive and cannot be used in periodic calculations but can be used as fragment based methods of evaluating the lattice energy.[130] Hartree-Fock itself has issues with electron-correlation[131, 132] but some post-Hartree-Fock methods are also worth note as they can circumvent the exchange-correlation issues found in DFT.

1.3.2 Post-Hartree-Fock methods

Møller-Plesset is a post-Hartree Fock, ab initio perturbation theory treatment that improves on some of the correlation issues found in common DFT. It is a relatively cheap perturbation method, and has been used in periodic calculations of energetic materials. However, it failed to deal with the long-range van der Waals dispersion interactions that were prominent in the RDX and HNIW.[133, 134] Revealing large errors in the computed cell parameters and the predicted densities. Correlated electronic structure methods like Coupled Cluster theory (CCSD(T)) work very well in quantitatively describing many-body systems and capturing intermolecular van der Waals dispersion to a very good accuracy.[135] CCSD(T) is a post-Hartree-Fock method that employs a many-body model, which scales sharply when applied to larger molecules or molecules with a sizeable number of heavier atoms. Even for dimer and trimer calculations quantifying the intermolecular interaction is an expensive challenge.

1.3.3 Summary

To summarise, there are many ways to utilize electronic structure methods in computational chemistry. One can carry out high quality molecular calculations using CCSD(T) as a fragment based method of computing the lattice energy or periodic intermolecular energy calculations can be carried out on the crystal structures of small molecular crystals. These periodic calculations can also require high-level, computationally exhaustive treatments like CCSD(T), large basis sets or a fine tweaking of dispersion corrections to obtain accurate results. Both CCSD(T) and MP methods struggle to decompose the total interaction energy into physically relevant contributions compared to SAPT (**2.3.2**). DFT-D and DFT-Tight-binding methods are becoming routine methods in CSP but are not yet methods cheaply accessible in Molecular Dynamics (MD) Simulations.[41, 136] On the other hand, SAPT(DFT) can be used to develop potentials with functional forms that can be readily used in standard MD codes.[38, 60, 103-106] Consequently, in this thesis we shall be using SAPT in the development of a truthful non-empirical intermolecular potential. In order to delve into the advantages of SAPT some background on its origins and theory must be given. These are detailed in **Chapter 2** but in the next section we shall examine some of its uses for predicting the crystal structures of some energetic materials.

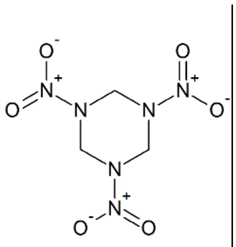
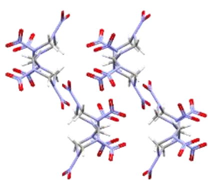
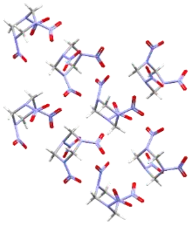
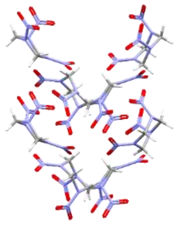
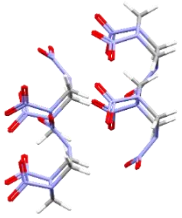
1.4 Energetic materials

1.4.1 Energetics and polymorphism

Energetic materials can be defined as compounds that can react chemically to produce heat and gas, and also emit light in an explosion. Classes of energetic materials include explosives, pyrotechnics, propellants and fuels, however, energetic materials are commonly referred to as explosives. Explosives have a range of different classifications.[17] The main two categories are high explosives and low explosives. Chemicals that cannot readily detonate i.e. pyrotechnics or propellants are known as “low explosives” while chemicals that can detonate are “high explosives”. The ease of detonation is ranked in descending order as primary, secondary and tertiary explosives. Tertiary explosives are also known as insensitive high explosives (IHEs).[17] In the design of energetic materials, there are two very important properties the industry would like to have complete control and understanding of:

- The detonation performance – the detonation velocity and pressures within the material, which dictate the power of the explosive compound.
- The sensitivity – the sensitivity of the compound to external stimulus, to initiate the reaction.

The energetics industry desire low sensitivity but high detonation performance materials and these properties experimentally appear to be closely related.[137, 138] There has been research into developing energetic co-crystals, combining two energetic molecules, which has produced materials that are still powerful but less sensitive to impact and thermal decomposition than their pure forms.[139] Both sensitivity and detonation properties depend on the physical properties of the solid, which usually vary with polymorph. The effects of the conformational differences between polymorphs on their impact sensitivity is explored in detail in **Chapter 4**. Polymorphism is certainly prevalent in energetic materials, some of the most popular energetic materials (RDX and TNT) show a range of polymorphs (***Table 1***).

 RDX	 Orthorhombic γ[140]	 Orthorhombic α[141]	 Orthorhombic β[142]	 Orthorhombic ε[143]
REFCODE	CTMTNA03	CTMTNA04	CTMTNA02	CTMTNA06
Notes	90K, 1.858 g cm^{-3}	150K, 1.82 g cm^{-3}	5.2GPa, 293K, 2.267 g cm^{-3}	5.7GPa, 293K, 2.266 g cm^{-3}
Lattice Energy (kJ mol^{-1})	-118.6	-106.2	-118.7	-118.0

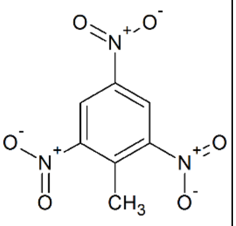
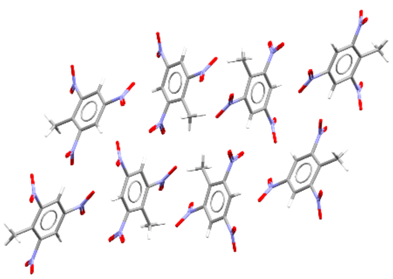
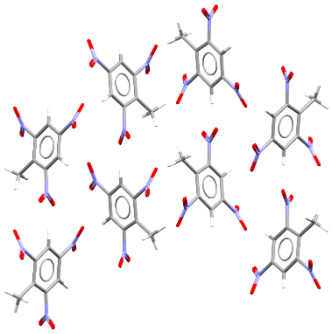
 TNT	 Monoclinic α[144]	 Orthorhombic β[144]
REFCODE	ZZZMUC08	ZZZMUC09
Notes	100K, 1.713 g cm^{-3}	123K, 1.704 g cm^{-3}
Lattice Energy (kJ mol^{-1})	-230.3	-230.1

Table 1: *The polymorphs of RDX and TNT with estimates of their intermolecular lattice energies. This was calculated using the empirical FIT potential and GDMA electrostatic model derived from the molecular charge density computed at PBE0/6-31G(d, p) quality on the experimental conformations with the hydrogen atoms corrected. These calculations ignore the effects of both pressure (PV) and temperature. The conformations used and their corresponding REFCODEs were obtained from the experimental Cambridge Structural Database entries.*[32] *The visual representations of the structures constructed using CCDC Mercury 3.6.*[145]

RDX is a widely used military explosive, that is often compounded with mineral jelly or polymers to form plastic explosives like C4 or PBX.[17, 146] RDX exhibits conformational polymorphism,[147] and there are two conformations (AAA and AAE) observed in experimental crystal structures.[140-143] These conformations refer to the positions of the NO_2 groups in the 6 membered ring, equatorial (E) and axial (A). The switching of the positions of these groups has an associated energy barrier as the ring has to go through a boat conformation. The β form is metastable at ambient temperatures and only the thermodynamically stable polymorph at pressures $P > 3.8\ GPa$.[142] The α form of RDX is the experimentally most stable polymorph at ambient conditions, however, the empirical FIT model, finds it to be the least stable in comparison to the three other high pressure forms. Its β form shows very large anisotropic displacement parameters associated with the first molecule's nitro groups and ring atoms,[142] this means there is significant libration about its position in the lattice. This suggests that it can readily transform into the more stable α form but also suggests a flexible study is more appropriate. Whereas, these lattice energy estimates are very rough as they also hold the molecular conformations rigid. The transition of $\beta \rightarrow \alpha$ could lead to deterioration of crystal, shear stresses and/or crystal defects affecting the creation of "hot-spots", hence, influencing performance and increasing sensitivity to detonation. It is clear a molecule like RDX could not be held rigid in a CSP study. It is also likely that within these AAE and AAA forms the rotation of the NO_2 groups around the N-N bond could also result in more conformational polymorphs.

TNT is another widely used, well known energetic material, with a metastable orthorhombic β form. More TNT is manufactured than any other military-grade explosive due to its low melting point and relatively low sensitivity.[148] During both the processes of casting TNT from the melt and growth in solution, the meta-stable orthorhombic form can be produced, which then transforms in the more stable monoclinic α form.[144] This change is problematic as it can lead to the development of defects (cracks and voids), and increased sensitivity to detonation. Another added complication is that the orthorhombic form can remain stable for over 12 months at ambient conditions without transforming, meaning that the transformations could happen in storage and be a safety risk. The empirical model also struggles to distinguish between the two polymorphs of TNT. As one can see in ***Table 1*** the two forms of TNT look virtually identical. This is because they are orientational polymorphs,[144] it is the relative packing that generates two polymorphs. TNT is a relatively rigid molecule meaning its conformation is unlikely to differ between polymorphs especially at ambient temperatures. It's the differences in packing that make this structure a large scale polytype. Polytypism is where the crystal structures differ in one dimension only (c axis for TNT), and is typically common to long chain molecules.[149-151] **Chapter 4** investigates the effects of conformation on lattice energies and molecular properties and determines the limits of the rigid model with regards to modelling the intermolecular interactions of nitro-groups in energetic materials.

1.4.2 Current methods of modelling energetic crystal structures

1.4.2.1 Density Functional Theory

	B3LYP	B3LYP (BSSE)	B3LYP-D (BSSE)	PBE-D (cluster)	PBE-D (cell)	Wb97xd	Wb97xd (BSSE)	Experiment
	Lattice Energy ($kJ\ mol^{-1}$)							
RDX	-83.6	-44.4	-151.4	-164.6	-154.4	-172.0	-127.9	-130.2
HNIW	-130.0	-29.1	-190.2	-192.7	-181.0	-227.3	-147.2	-168.7
FOX-7	-119.9	-90.2	-188.3	-209.2	-185.6	-205.9	-172.9	-108.7
TNT	-134.4	-18.2	-118.3	-129.5	-133.8	-127.7	-94.3	-104.6

Table 2: *The lattice energy ($kJ\ mol^{-1}$) for four energetic molecular crystals (RDX, HNIW, FOX-7 & TNT) obtained from experiments and calculations using various functionals. Data obtained from ref. 128*

Table 2 compares periodic DFT calculations with cluster based methods, the effects of including dispersion corrections, accounting for the basis set superposition error (BSSE[126, 127]) and the influence of different hybrid functionals on calculating the lattice energy of different energetic crystal structures. The study fails to clearly specify the basis sets used in each calculation, and the choice of basis set and basis set convergence (or method of estimating this by a BSSE correction) has a major influence on the computed DFT energy. As mentioned before (**1.3.1**) there is no way to systematically improve DFT so trialling different functionals is often the solution. Whereas one can systematically improve on a force-field's functional form and parameterisation to better model intermolecular interactions, certain functionals simply fail to capture intermolecular interactions when extended to other molecules.

B3LYP with a BSSE correction predicts the 2nd closest lattice energy to the observed energy for the energetic material FOX-7, but performs horrendously for the other energetic molecules. The Wb97xd functional[122, 129] does reasonably well once corrected for BSSE, and including a dispersion correction to the B3LYP functional only helps improve its TNT and RDX prediction. Considering the energy differences between related polymorphs,[7] these treatments could not be used for reliable CSP. The study in ***Table 2*** did not specify what polymorphs of the energetic molecular crystals were used and does not show whether variations in method affects the relative energies. Additionally, care should be taken in comparing predicted lattice energies (taken at $0K,\ 0GPa$ pressure) to the room temperature properties of crystal structures. This is because thermal effects will mean that there is a significant difference between lattice energies and enthalpies at practical temperatures. In **Chapter 3**, the effects of this type of approximation on the calculated properties of pyridine polymorphs are also examined.

1.4.2.2 Rigid force-fields using SAPT(DFT)

Symmetry Adapted Perturbation Theory (SAPT) based on the Kohn-Sham density functional theory (DFT) of monomers (SAPT(DFT)) has been a popular tool in the development of non-empirical force-fields for energetics.[104] This section discusses examples but the theory and implementation behind SAPT(DFT) and non-empirical potential development is investigated in **Chapters 2** & **5**.

1.4.2.2.1 RDX

Over 1000 SAPT(DFT) dimer interaction energy (up to the 2nd order SAPT(DFT) energy) calculations using an aug-cc-pVDZ basis[152] were used to derive an non-empirical isotropic atom-atom intermolecular force-field to model the potential energy surface of RDX.[104] The study aimed to develop force-fields to be later used for rigid MD simulations in order to examine the thermal and pressure response of RDX. This study uses an isotropic short-range description to enable the use of the force-field in current MD codes, however, it should be noted that codes for doing MD with anisotropic potentials are being developed.[153] The force-field used a damped C_6 isotropic dispersion term to model the long-range portion of the dispersion contribution alongside a point-charge electrostatic model. The main non-additive contribution, the polarization energy, a long-range component of the induction energy (**2.6.1**), was also absorbed into the C_6 dispersion coefficients. The C_6 dispersion coefficients were fit to the SAPT(DFT) polarization and dispersion energies of the interacting two-molecule configurations.[104] The other 2nd order induction and dispersion effects (the charge delocalization and the exchange-dispersion energies) were included by adjusting the short-range parameters to fit the total 2nd order SAPT(DFT) dimer energies. Higher order energy calculations were not carried out thus non-additive many-body contributions to the lattice energy calculations were not included, due to their computational expense outweighing their percentage contributions to the total energy. Coupled cluster (CCSD(T)) calculations on the RDX dimer, found the higher order induction contributions from the δ_{int}^{HF} energy (**2.6.3**) to be negligible.[104] In addition, the force-field does not include an intramolecular model for RDX and keeps its conformation fixed.

The RDX potential was used to carry out lattice energy minimizations, finding the dispersion energy to be comparable between all RDX polymorphs. Consequently, it was assumed that the relative energies would not significantly differ even if many-body contributions were included. The predicted crystal densities, lattice parameters and bulk moduli showed good agreement with observed experimental data. Though the potential had not been corrected for quantum effects like zero-point energy vibrations,[104] the treatment was accurate and efficient enough to predict the correct polymorph ranking and structure of RDX in a CSP study. The α form was found to be 4.7 $kJ\ mol^{-1}$ lower than the next most stable structure, and the crystallographic parameters corresponded well with experiment (***Table 3***). This study also compared its non-empirical model was with a bespoke empirical least squares (SRT) fitted model, which was created specifically for MD studies of RDX by fitting to the experimentally observed cell geometries and lattice energies of RDX.[154] A marked improvement in accuracy was observed when the non-empirical model was used. A key aspect of the non-empirical force-fieldfor RDX[104] was its simplified functional form. The simplification reduces the ability to identify the physical relationship between contributions. This simplified functional form did enable the potential to be used in an isothermal-isobaric Molecular Dynamics (NPT MD)[155] study to compare potential predicted lattice vectors with observed values.

	Energy $(kJ\ mol^{-1})$	Density $(g\ cm^{-3})$	a (Å)	b (Å)	c (Å)
Non-empirical	-128.95	1.778	13.259	11.634	10.754
Empirical (SRT)	-121.88	1.738	13.403	11.800	10.732
Experiment	-130.1	1.796	13.200	11.609	10.724

***Table 3:** The empirical and non-empirical computed properties for α-RDX extracted from[104, 154]. Experimental energy of sublimations obtained from [156] at 327-373 K. Experimental lattice vectors/density obtained from Ref.157*

The dispersion energy was found to be the most important attractive contribution to the dimer interaction energy, further reinforcing the idea that effective modelling of van der Waals dispersions is key. There are many cases where polarization/electrostatic interactions are the most important component of the lattice energy (heavily ionic or hydrogen-bonded structures) but even in these cases the dispersive forces are always important, neglecting dispersion would otherwise result in structures that were too weakly bound or not bound at all.

1.4.2.2.2 FOX-7

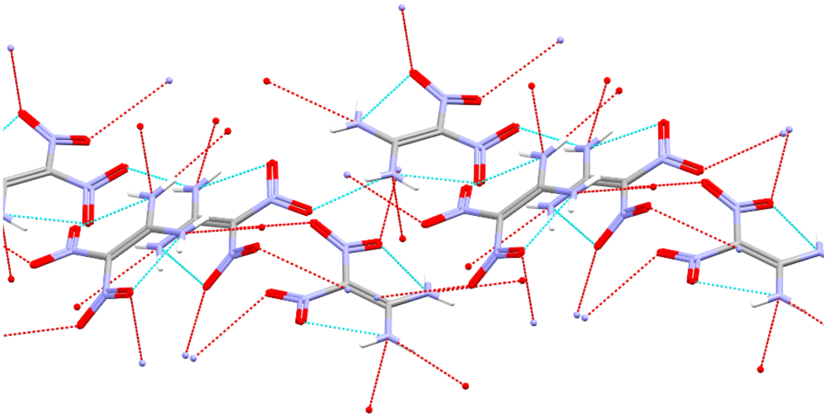


***Figure 3:** The network of short-contacts greater than the sum of van der Waals radii for the α polymorph of FOX-7.[158] These networks differ between polymorphs. Intramolecular interactions are given in blue while intermolecular interactions are in red. This visual representation was constructed using CCDC Mercury 3.6.[145]*

FOX-7 is a good example of an energetic material with very limited experimental data, accordingly, empirically fitting a potential for the crystal is very difficult, thus developing a non-empirical potential is the right direction. Conversely, due to the very limited experimental data results cannot be properly validated. FOX-7 is a promising energetic material as it has a high shock insensitivity and thermal stability, not decomposing till temperatures above 500K.[159, 160] The crystal structures of the FOX-7 polymorphs contain an extensive weak hydrogen-bonding network in the plane of the layers and weaker dispersive forces between layers, resulting in a unique balance of intermolecular forces.

Another good example of employing SAPT(DFT) to develop a non-empirical intermolecular force-field from first principles was presented for the organic energetic molecule FOX-7.[158, 159, 161] Over 1000 random $E_{int}^{(2)}[SAPT(DFT)]$ dimer interaction energies (of aug-cc-pVDZ quality[152]) were used to derive and fit a non-empirical pairwise potential. The force-field used an isotropic exp-6 Buckingham repulsion-dispersion definition that absorbed polarization effects coupled with a point-charge electrostatic model. The model for FOX-7 does attempt to include higher order induction contributions by including the δ_{int}^{HF} energy in the total energy used to fit the short-range parameters. This was because FOX-7 is a reasonably polar molecule and thus higher order induction contributions should be important. The contributions were included through the fitting of the short-range components, however, many-body dispersion effects were not included. The potential was used in an isothermal-isobaric (NPT) rigid MD simulation study to also compute the thermal and pressure dependent properties of the crystal.

FOX-7's nitro groups are known to rotate as temperature is increased, weakening intermolecular hydrogen bonds.[159] While SAPT(DFT) manages to correctly capture the long-range dispersion interactions and the close-range hydrogen bonding, the MD simulations in this study assumed a rigid molecule, meaning that this conformational change was not modelled. Hence, as the temperature increases the hydrogen bond network is unbroken, resulting in a stiffer material/bulk modulus. FOX-7 clearly needs a flexible force-field to better describe changes in geometry with external pressure or thermal change, and in turn improving the predicted thermal response. This is certainly the trajectory of current modelling, and the challenges faced in incorporating conformational flexibility of nitro groups are discussed in **Chapters 4, 6** & **7**.

1.4.2.3 Flexible force-fields using SAPT(DFT)

The development of a suitable work-flow for the easy generation and use of representative, tailor made force-fields for all molecules let alone energetic molecules is a goal within many research groups. A number of groups have been looking at strategies to automate and optimize the parameterisation and construction of tailor-made force-fields. A recent study of α-RDX[162] attempts to do so by developing a so-called self-consistent iterative force field

(SCFF) optimization technique. The resultant force-fields aim to model the total lattice energy by employing various electronic structure methods for parametrisation. The hybrid DFT functional B3LYP/6-31G* was used to optimize many monomer and dimer conformations of RDX, in order to generate a conformer training set to fit the intramolecular portion of the force-field. The point-charge electrostatic model was also fitted at the B3LYP/6-31G* level using the CHELPG scheme[163] (**2.5.1.2**). 1000 second-order and 2000 first-order SAPT(DFT) dimer interaction energies (this time at the PBE0/aug-Sadlej level of theory) were used, in conjunction with the overlap model methodology used by Misquitta,[38, 60] to fit an isotropic short-range potential with a damped isotropic C_6 dispersion model. In this study the polarization contribution is also absorbed into the dispersion coefficients, which were obtained through fitting to the induction and dispersion contributions of the 2nd order SAPT(DFT) dimer interaction energy and the short-range parameters were then scaled to the total energy curves. Assuming the polarization can be described using the same functional form as the dispersion C_6/R^6, is definitely an approximation one should avoid, as the polarization term is non-additive and should not be included with an approximately pairwise additive term like the dispersion (**Chapter 2**).[164] Furthermore, fitting the dispersion coefficients (a correlated property) directly to two molecule interaction energies will remove from the physicality of the model, as the dispersion coefficients can be derived directly from $\rho(r)$ and should be closely related to the electrostatic model (**Chapters 2** & **5**). The iterative aspect of the workflow occurred during the inclusion of the many-body dispersion correction using periodic DFT-D calculations. This is done by weighting the short-range, dispersion and intramolecular parameters and fitting the force-field computed energies to periodic DFT-D calculations using a minimization algorithm.[162] The force-field was then used to predict some crystalline properties of α-RDX, reproducing all the correct geometries for the remaining RDX polymorphs with good agreement with literature. Vibrational analysis from data generated by the force-field was then employed to categorise transition states. The force-field was then integrated into NPT MD simulations to predict the densities and lattice parameters of α-RDX under thermal and pressure conditions similar to those that explosives are subjected to. As seen with other force-fields,[161] the force-field struggled to describe some of the intramolecular configurations seen in α-RDX. The out-of-plane bending angle of the equatorial NO_2 group in the molecule was defined poorly in comparison to the axial groups. This is likely to be due to the potentials approximate representation of intramolecular forces and possibly its parameterisation process.

Nonetheless, the ability of this potential to maintain its functionality at these more “extreme” conditions suggests that the SCFF optimisation is a very valuable strategy for parameterisation. While the use of a point-charge electrostatic model, the functional form of the potential (absorbing the polarization), and the fitting process for the some of the core parameters (like the dispersion coefficients) was certainly questionable, the workflow was novel and the development process in this study was ambitious and one to learn from.

1.4.2.4 Summary of SAPT(DFT) derived force-fields for energetic molecules and crystals

The above non-empirical force-fields derived for energetic crystals were notable examples of atomistic potentials, obtained using ab-initio methods, which were capable of quantitatively determining crystal structures containing fairly sized molecules with a sizeable number of heavier atoms. The examples for FOX-7 and RDX use non-empirical potentials with simplified functional forms, which have an isotropic short-range description and no explicit polarization model, they instead absorb the effects of polarization into the dispersion coefficients. This description works well for small molecules, which are relatively rigid, like benzene.[103, 165] However, these models are unlikely to be extendable to larger more anisotropic systems like trinitrobenzene.[166] The dimensionality of a potential energy surface grows extensively with the degrees of freedom, and as such many potentials are limited to relatively small, rigid molecules with a few flexible groups. Accurately modelling conformational flexibility through including an intramolecular model is certainly a challenge.

1.5 Thesis overview

With all this in mind, this thesis aims to design a novel non-empirical atom-atom intermolecular force-field that outperforms contemporary empirical methods and develop a workflow that can be extended to larger more complex systems. The following chapters of this thesis aim to describe the theory of intermolecular forces and show how this leads to the development and use of ab initio force-fields (**Chapter 2**). The theory of intermolecular forces is applied to the pyridine crystalline lattice in a non-empirical potential, initially derived for the gas-phase[38] by Dr. Alston Misquitta, and then used for CSP (**Chapter 3**). The distributed intermolecular force-field (DIFF) for pyridine was the 2nd ever non-empirical anisotropic atom-atom potential developed for CSP. The only previous example being $C_6Br_2ClFH_2$, which neglected polarization effects.[60] **Chapter 3** is published work[167] that not only investigates the effects of polarization but pressure and free energy in a novel manner with unexpected results.[167] The thesis then progresses on to look at the influence of molecular geometry on the electrostatic properties of the crystal, and attempts to address the problem of flexible nitro-groups that one faces with energetic materials (**Chapter 4**). The published investigation[168] examines how conformation affects the molecular charge distributions, and in turn energetic properties and intermolecular interactions. It is determined whether or not one can hold trinitrobenzene (TNB) rigid and still obtain its genuine potential energy surface.[168] In the final chapters an ab initio non-empirical anisotropic atom-atom intermolecular force-field for TNB shall be derived (**Chapter 5**) and applied in a CSP study (**Chapter 6**) to determine whether one can extend this workflow to larger more challenging systems.

2 The development and implementation of the theory of intermolecular forces

2.1 Introduction

2.1.1 Fundamental notions within the theory of intermolecular forces

The notion that matter is composed of atoms and molecules is an idea that has existed for millennia, but the theory that the properties of matter are governed by forces and can even be predicted is relatively young in comparison.[164] The intermolecular forces determine many experimental properties of gases such as, second virial coefficient,[169] viscosity[170] and diffusion constants. In addition, modern spectroscopic techniques at low temperature provide detailed information about the potential energy surface (PES) around the global minimum, which can include information on barriers between equivalent minima and possible transformation pathways.[171-174] Furthermore, theoretical methods like intermolecular perturbation theory can give the components of the intermolecular interaction energy and allow for a detailed and accurate study of the intermolecular forces in a system.

It's safe to say these are concepts and methodologies, which a non-scientist would be forgiven for not being familiar with. However, some evidence for molecular interactions and intermolecular forces is very readily available. The cup of water on your desk, the ice in your fridge, the vapour that forms when you shower in the morning (if you shower in the morning) all show the different phases of water in a textbook illustration of intermolecular forces at work. Intermolecular forces play an important role in the physical properties of molecular solids and liquids, crystal and liquid structures (polymorphs, liquid crystals, amorphous phases etc.), being responsible for molecules condensing from the gas, temperature dependence, pressure, and properties such as viscosity and diffusion etc. Consequently, a profound understanding of intermolecular forces is crucial in a range of industries, such as the pharmaceutical industry for drug design and discovery methods.

H_2O is a definitive, open example of the nuances and importance of intermolecular forces. Ice, water and water vapour have differing structures and dynamics that result in different physical properties. The properties of every state of H_2O differs from that of hydrogen sulphide (H_2S) or the iso-electronic methane (CH_4) due to the differences in the intermolecular interactions between these systems. The strong hydrogen-bonds that form between the oxygen and hydrogen atoms of neighbouring water molecules dominates its physical properties resulting in a unique phase diagram, with over 15 phases.[48, 175] The directionality of the strong hydrogen bonds in water cause its crystal structure to exhibit a more open structure, hence it is less dense than its liquid phase, which has a substantially higher boiling point than its iso-electronic counterparts NH_3 or CH_4. The diversity of phases of water and their physical properties originates from a subtle balance of electrostatic, polarization, charge-delocalization and dispersion intermolecular interactions. For hydrocarbons like methane, van der Waals dispersive forces influences its behaviour more heavily, and for NH_3, which forms weaker hydrogen bonds than water due to nitrogen being less electronegative than oxygen. Consequently, its intermolecular interactions and phase diagram differs too.

In all states the distances between molecules are not equal, thus the effects of forces and interactions between neighbours will differ within co-ordination spheres. For each pair of rigid polyatomic molecules, the intermolecular energy and forces will depend on R, the centre of mass separation, and the orientation Ω. Generally, when describing interactions as a function of R we find there is a range where attractive forces, like hydrogen-bonding or van der Waals forces, dominate and a region at smaller R dominated by repulsive forces. The balance between attractive and repulsive forces will determine the equilibrium distance between two molecules, and in turn govern the overall thermodynamically stable configuration of an array of molecules like a crystal lattice structure. If one thinks of repulsive and attractive forces as compressing and stretching a foam ball, we know that if we squeeze it, the ball will naturally expand to its equilibrium shape and if the ball is stretched it will eventually contract (unless you rip it, then you'll need a new ball). This basic example illustrates the effects of deformation on a system and how a system will react to changes and move to its most stable state.

Varying external influences like temperature or pressure will sample different portions of the repulsive or attractive regions of the intermolecular PES, changing the interactions between molecules and thus the energy of the system, which are important in determining the properties. Due to the repulsive forces arising from the Pauli exclusion principle, two electronic charge distributions cannot reside on top of each other, therefore, the energy increases exponentially as $R \to 0$. This accounts for the low compressibility of condensed phases. Consequently, repulsion is active in the short-range and attractive forces are usually prevalent in the long-range, though as $R \to \infty$, the potential energy V of the system always tends to 0. This establishes a qualitative description of the potential energy $V(R)$ of a system, due to its intermolecular forces, as a function of R. ***Figure 4*** further illustrates this for the two TNB molecules in the S4 configuration, one of the most stable gas-phase dimer structures (**5.5.3.1.3**).

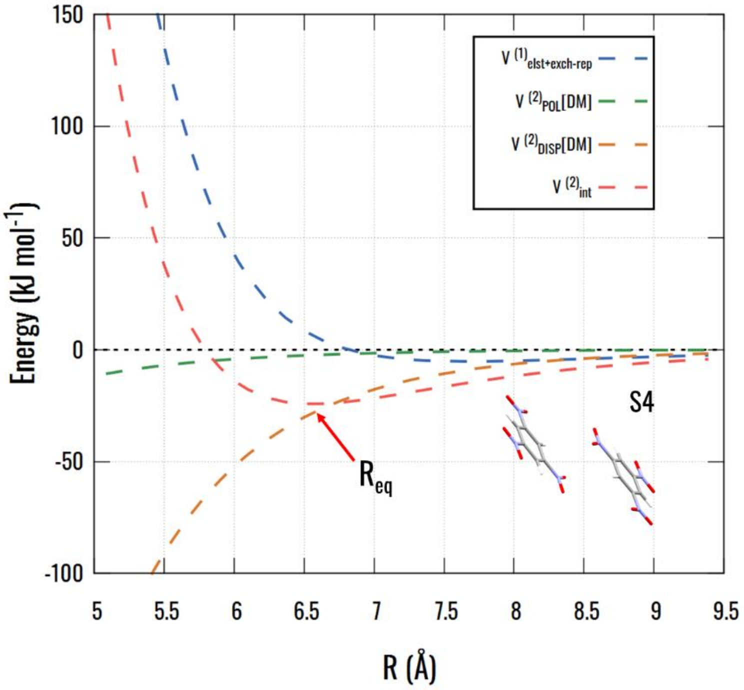


Figure 4: *The intermolecular potential energy ($V^{(2)}_{int}$) of a S4 gas-phase dimer configuration of trinitrobenzene (**5.5.3.1.3**) as a function of R, broken down into its energy components. The 1st-order exchange-repulsion and electrostatic potential energies are shown together as $V^{(1)}_{elst+exch-rep}$, while the polarization and dispersion energies are indicated as $V^{(2)}_{POL}$ and $V^{(2)}_{DISP}$ respectively. These components of the intermolecular potential energy are defined in (**2.3.3**). R_{eq} indicates the equilibrium distance between the two TNB monomers.*

In ***Figure 4*** we see that different components of the intermolecular interaction energy govern the stability of a system. There is a point of equilibrium (the equilibrium distance (R_{eq})) where the PES describing the intermolecular interactions is at a minimum. Having a precise form of $V(R)$ is crucial to getting the point of equilibrium correct and in turn obtaining the general features of the intermolecular PES. The importance of different energy components will change the overall potential energy surface, therefore, one can expect the PES of H_2O and CH_4 to differ significantly. The PES is very dependent on the orientations of the two interacting molecules, more so for water and TNB than for methane, thus, ***Figure 4*** is the description of $V(R)$ only for a specific orientation of two rigid TNB molecules. The different contributions vary in their orientation dependence, with the polarization and dispersion contributions always being attractive, whereas the electrostatic contribution can vary strongly in sign and magnitude. The intermolecular interaction in some orientations can be totally repulsive, when the electrostatic contribution is repulsive and larger than the attractive dispersive contribution. So if intermolecular forces can vary considerably between molecules, with different phases sampling different regions of the intermolecular potential energy surfaces, how we do accurately model it?

2.1.1.1 The pairwise additive approximation

The simplest model of the total interaction energy E of a system is one where E is composed only of one-body and two-body terms. In this thesis, the term “body” refers to an interacting unit. Internal geometric changes in the molecule lead to the one-body term, and this energy penalty may be compensated for by the intermolecular interaction energy stabilization within the dimer or crystal. Close-contacts between neighbouring molecules in crystals and liquids correspond to intermolecular interactions that can be stabilizing or destabilizing depending on orientation and separation. This intermolecular energy can be approximated as a pairwise summation between neighbouring molecules. Assuming pairwise additivity, the total energy E of a system is composed of the summation of all the intramolecular (or conformational) energies (E_M^{intra}) of the isolated molecules M in the system and all pairwise intermolecular interaction potential energies (V_{MN}) between molecules M and N. One must avoid double counting as $M \dots N = N \dots M$, thus we write the sum over pairs to avoid doing so.

$$E_{latt} = \sum_{M} E_M^{intra} + \sum_{M>N} V_{MN}$$

2

This pairwise model, however, is an initial approximation. Equation 2 has yet to include many-body (i.e. 3-body V_{MNO} or 4-body V_{MNOP}.) effects. The most important non-pairwise terms that are considered in this thesis are the many-body dispersion and polarization contributions.

$$E_{latt} = \sum_{M} E_M^{intra} + \sum_{M>N} V_{MN} + \sum_{M>N>O} \Delta V_{MNO} + \sum_{M>N>O>P} \Delta V_{MNOP} + \cdots$$

3

Typically, analytically including many-body terms drastically increases the dimensionality of the model and these contributions can diminish with increasing number of bodies.[176-179] Accordingly, they are usually included as approximate corrections to the energy or neglected completely. Conversely, there are many instances when these terms, especially the three-body terms, cannot be completely neglected[180-182] (**3.4.5.2**). For example, polarization effects depend on the symmetry of the environment and so effects can be cancelled out or compounded depending on crystal configurations. As we further develop the intermolecular model we will discuss this more but for now we will neglect the many-body terms and set aside the conformational energy of a system by holding it rigid, focusing on the pairwise intermolecular energy.

A rigid pairwise model can be used to approximate the intermolecular lattice energy of a crystal structure as a pairwise summation of the intermolecular interactions between monomers within the lattice.

$$E_{latt} = \sum_{M>N} V_{MN} \; [rigid\ molecule\ approximation]$$

4

2.1.1.2 Free energies

2.1.1.2.1 Background

The relative stability of a static crystalline lattice and the experimentally observed stability can differ significantly as the real crystal is not static and the molecules move even at $0\,K$, and this motion increases with temperature. The resultant phonon effects will certainly influence the crystal structure properties.[183, 184] Phonon and isotope effects can expand crystal structures, change relative stabilities and consequently result in polymorphic phase changes.[183, 185, 186] The energy difference between the hypothetical static crystal at $0K$ and a real crystal can be estimated by lattice dynamics within the rigid-molecule harmonic approximation.[11, 183, 185] The Gibbs free energy of a crystalline lattice $G_{free}(T)$, gives the thermodynamic stability of the crystal structure at a given pressure and temperature

$$G_{free}(T) = H_{latt} - TS$$

$$H_{latt} = U + PV$$

$$U = E_{latt} + ZPE + U_{therm}(T)$$

5

where T is the temperature of the system, P the applied external pressure, V the volume, S the entropy, E_{latt} the lattice energy, ZPE the zero-point energy, the energy due to the inherent lattice vibrations that occur at $0K$, $U_{therm}(T)$ the internal energy of the system due to thermal effects and H_{latt} the lattice enthalpy.[183] The summation of the energy due to thermal effects, and the lattice and zero-point energies gives us the total internal energy of the system U. The energy contribution due to the entropy of the system TS, can be added to the ZPE and $U_{therm}(T)$ to give the free energy thermal correction $F_{vib}(T)$. The pressure term (PV) is usually dropped in CSP studies as its contribution is only significant at high pressure but in **Chapter 3** one shall see how it can lead to new polymorphic forms.[167, 187] Keeping the pressure of the system constant gives the Helmholtz free energy of a system

$$A_{free}(T) = E_{latt} + F_{vib}(T)$$

$$F_{vib}(T) = -TS + ZPE + U_{therm}(T)$$

This treatment of obtaining the free energy correction $F_{vib}(T)$ requires the crystal structure to be a lattice energy minimum, with respect to the lattice parameters, including angles. It is good practice to evaluate $A_{free}(T)$ only at temperatures well below the crystal melting point as the harmonic approximation begins to lose its validity close to phase transitions. Methods for evaluating free energies of crystals from a MD simulation that samples the anharmonicity of the molecular motions require a very accurate intermolecular potential, and so very few of these kind of studies have been performed.[188]

2.1.1.2.2 Implementation in DMACRYS

DMACRYS[59] (originally DMAREL[189]) is a computational chemistry code, which models infinite lattices of rigid organic molecules, and can use anisotropic atom-atom intermolecular potentials. It was originally written to include the energies, forces, torques and some second derivatives arising from a distributed multipole electrostatic model.[190] The capability to accept anisotropic repulsion models was added later,[191, 192] and isotropic C_8, C_{10} dispersion terms in order to be able to use non-empirical dispersion coefficients instead of just an effective C_6 value.[40] The ability to use distributed dipolar polarizability tensors and model the polarizability explicitly was also implemented[40] with limited testing. DMACRYS can also calculate the effects of the PV pressure term,[40] the elastic constants[193] and $k = 0$ phonons within the rigid molecular harmonic approximation.[194] In **Chapter 3**, the effects of pressure are calculated by minimizing the cell geometry including the PV contribution to the energy with DMACRYS in order to obtain the lattice enthalpy.

The Helmholtz free energies of the polymorphs of pyridine are calculated in **Chapter 3** by sampling a number of k-points in reciprocal space. This is done by computing the phonons frequencies[195] and elastic stiffness tensors[194] for a number of linearly elongated supercells of crystal to sample the first Brillouin zone. The supercells are generated by selecting a default 0.12 Å^{-1} distance, in reciprocal space, between each k-point in each direction[11, 183] which samples about 26 k-points. If the isolated molecule is small one does not require very large supercells. This was the case with pyridine (**Chapter 3**). The Debye frequency contribution to the acoustic phonons and a Gaussian Kernel Density Estimate (KDE) of the optical density of states (the default bandwidth is 3 cm^{-1}) are calculated from the elastic tensors and phonons respectively, and then used to calculate the free energy thermal correction F_{vib}, as the sum of the vibrational zero-point energy and the thermal energy at various temperatures. A quasi-harmonic approximation could be used to calculate the thermal expansion due to the free energy. The pioneering work using this approximation to estimate thermal expansion suggested that the calculations are very computationally expensive yet may have a negligible effect on the relative free energies in comparison to the computational expense.[185, 196, 197]

For a properties calculation in DMACRYS to include the phonons and an accurate elastic tensor, the second derivatives have to refer to the moment of inertia axes. All second derivatives are calculated numerically, while the first derivatives are calculated analytically in testing that the starting point is at a minimum. The 1st and 2nd derivatives are needed for structure minimization, but an estimate of the second derivative matrix with respect to the molecule fixed-axis is built up during the minimization process.

The polarization term has to be numerically iterated to consistency (**2.6.1.1**). Thus there are no analytical first derivatives, let alone second derivatives of the polarization energy, and numerically calculating the second derivatives of the polarization term is very expensive. Hence, crystal structures cannot be routinely optimized when the intermolecular potential includes an explicit, distributed polarizability model.

Splines were introduced in DMACRYS to reduce the numerical noise in energies and first derivatives due to the cut-off in the direct lattice summations. They result in continuous forces and allow better optimization with smaller lattice sums. Splines are used throughout this thesis to allow a much smaller direct lattice summation cut-off. The details of the changes to the methods of summing electrostatic and polarization contributions using Ewald summations and splines can be found in the DMACRYS manual.[59] However, some properties calculations have been found to give poor results when splines are included, presumed due to the splines not giving continuous second derivatives. Thus splines are not used for phonon calculations. Subsequently, with so much likely numerical noise, phonon (and hence free energy) calculations in this thesis do not include the polarization term.

2.1.2 Intermolecular interactions: A delicate balance of contributions

The theory of intermolecular forces allows us to decompose the total intermolecular interaction energy of a system into various contributions, that one can then use to derive a non-empirical intermolecular force-field. The most important contributions to the intermolecular energy arise from a few physical contributions. The most dominant interactions for most organic molecules containing heteroatoms are the electrostatic, the universal exchange-repulsion and dispersion interactions. The leading (and often most important) term is the electrostatic interaction, the Coulombic interaction between the unperturbed molecular charge densities ($\rho(r)$) of molecules M and N. The exchange-repulsion term arises from the Pauli exclusion principle, while at higher orders there are additional contributions from charge-delocalization and dispersion (and mixed) terms. The intermolecular interactions can be either attractive or repulsive depending on their origin, orientation and distance and can be partitioned and classified as long-range ($V(\boldsymbol{R}) \sim R^{-n}$) and short-range ($V(\boldsymbol{R}) \sim e^{-\alpha R}$) contributions, depending on whether there is overlap of the molecular charge densities. A number of factors and terms like hydrogen-bonding, $\pi \dots \pi$ interactions, resonance (**2.1.2.1**) or van der Waals forces (generically termed) are used to qualitatively describe the intermolecular forces for specific types of organic molecules, giving intermolecular potential energy surfaces as diverse and rich as the range of organic molecules themselves. The intermolecular forces in organic solids are sufficiently weak (in comparison to metallic, ionic or covalent bonding) that one can treat it as a perturbation on the isolated molecular charge distribution (**2.3**). The gravitational and magnetic effects can be excluded, as can resonance forces if only closed shell molecules are considered. We will not attempt to list and explain every single one in detail but we will touch on the most prominent long-range and short-range terms and then investigate the most important terms when developing intermolecular force-fields for weakly bound organic materials.

2.1.2.1 Long-range terms

Long-range contributions refer to terms that survive at large separations, when there is negligible overlap of charge distributions, with the largest long-range terms for organic materials being the electrostatic, induction and dispersion contributions. These terms also have short-range contributions, however, the long-range and short-range expressions are modelled quite differently. The electrostatic energy (E_{elst}) is the simplest to understand. It is a 1st order energy that arises from classical interactions between the static charge density of molecules M and N, when $\rho_M(r)$ interacts with $\rho_N(r)$. It is orientation (Ω) dependent and strictly additive, however, it can be both attractive and repulsive depending on the orientation and separation. The induction energy (E_{IND}) is a 2nd and higher order energy that arises from the distortion of the molecular charge density ($\rho_M(r)$) due to the electric field created by neighbouring molecules. The response to the electric field means the interaction is always attractive, however, as the field of neighbouring molecules can cancel it out or reinforce it, it is strongly non-additive. Therefore, the induction contribution within a gas-phase dimer interaction is very different from the induction in the condensed phase and can be very difficult to calculate as it requires numerical iteration (**2.6.1.3**). The dispersion energy (E_{DISP}) is also a 2nd and higher order energy that cannot be explained classically as it arises from correlated fluctuations in electron density ($\rho_M(r)$ and $\rho_N(r)$) as the electron moves. It is non-additive but in systems bound by the electrostatic/induction energies, the dispersion can be approximated as being additive. The way the dispersion coefficients (C_n) are computed ensures that the intramolecular non-additivity is included, thus all that is left is the intermolecular non-additivity from many-body terms. The two-body dispersion energy is always present and attractive. Consequently, the average effect is a lowering of energy and strengthening of the correlation effect as molecules move closer together.

Resonance & magnetic interactions can also result in long-range effects. Resonance is a non-additive contribution that occurs when either one of more of the interacting molecules are in a degenerate state or when molecules M and N are identical and one is in an excited state. Therefore, resonance does not occur in the closed shell systems in this study. Magnetic interactions are additive and involve electron or nuclei. With regards to electrons, magnetic interactions typically occur when both molecules have unpaired spins. This typically results in a small but non-negligible energy contribution. Magnetic effects for closed-shell systems[198] can also occur when nuclei have non-zero spin ($I \neq 0$), which is frequent but the resultant energy contribution is many orders of magnitude smaller than the electron case so is negligible. For example, a recent study found that even for coronene, the thermodynamic effects of magnetic interactions were very small, too small to be included in intermolecular pair potentials.[198]

Long-range terms are also present at shorter ranges. For instance, the multipolar expansion electrostatic contribution remains finite until nuclei make contact. The difference between the total electrostatic energy and the multipolar electrostatic energy is the penetration energy, the electrostatic short-range term when Ψ overlap. Long-range contributions are typically described as a power series in $1/R$, which begin to diverge as R tends to zero. Thus, the expansion is only valid for large R. Furthermore, R assumes a distance between two points, thus, even where the expansion converges, because each molecule is treated as a single point rather than extended in space it may still be in error. When the expansion is then distributed between atomic sites this error remains and is known as the penetration error. We shall avoid this by absorbing the penetration term into the short-range portion of our non-empirical potential.

2.1.2.2 Short-range terms

The long-range components can be derived without the need for full anti-symmetrization of the dimer wave-function (**2.3.1**). The influence of anti-symmetrization becomes very important when the molecular wave-functions (Ψ_M) overlap is significant, which leads to exchange energies.[164] In practice the main region of interest is the start of overlap, around van der Waals contact distances. At small R the perturbation (**2.3**) of the Hamiltonian becomes complex and non-trivial to compute analytically because the exchange of electrons between the two molecules means you can no longer identify an electron within a molecule. We remedy this through fitting the short-range energy contributions (**2.9** & **Chapter 5**). Two effects come into play at short-range, the orbital overlap needs to be accounted for and the $1/R$ expansions need to be damped (**2.8**) to make them well-behaved. At short-range the exchange-repulsion interaction is the most dominant contribution. The exchange-repulsion ($E^{(1)}_{exch-rep}$) is a 1st order energy, which is exactly additive at 1st order.[164] There is an attractive portion of the exchange-repulsion, which is the exchange due to electron delocalization between molecules, which increases the uncertainty in the position of electron. Due to the uncertainty principle

$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

7

the uncertainty in the momentum decreases, and subsequently, the energy decreases.[164] There is also a repulsive effect which is more dominant. This is due to the molecular wave-functions ψ wanting to maintain Pauli antisymmetry. This costs energy, and the energy cost increases as $R \rightarrow 0$. We group these two phenomena together as the exchange-repulsion energy $E^{(1)}_{exch-rep}$, and the combined term is always repulsive for closed shell systems. There are other exchange terms that come into play when the molecular wave-functions overlap, like the exchange-induction and exchange-dispersion but the other notable short-range effect is charge-delocalization (E_{CD}). Charge-delocalization is a non-additive attractive term, which describes the electron delocalization between the high energy occupied orbitals in donors and the lower energy unoccupied orbitals in acceptors (**2.6.2**). For an accurate representation of the genuine intermolecular potential energy surface, getting the balance between the short-range and long-range terms is crucial. Moreover, not all terms are as important depending on the molecule, for instance, the polarization contribution to the intermolecular interaction energy for water clusters will be more significant than in methane clusters. However, accurately modelling these terms in a molecule is vital.[22, 47, 60, 199, 200]

2.2 Supermolecular methods: Density Functional Theory (DFT)

To summarise, contributions to the intermolecular interaction energy can be analytically calculated at long-range using the molecular properties, however, this is non-trivial at short-range. The intermolecular interactions at short-range can be calculated *ab initio* using either supermolecular methods or intermolecular perturbation theory.

The supermolecule method is a concept that says the intermolecular energy between bodies M and N is simply the internal energies E_{intra} of the two subtracted from the total energy.

$$E^{MN}_{int} = E^{MN}_{tot} - E^{M}_{intra} - E^{N}_{intra} \tag{8}$$

There are some notable disadvantages to this method. There is no variational principle as the result is a difference of energies, and the intermolecular interaction energy has no components, thus no physical interpretation. Moreover, for a dimer interaction calculation, each monomer's basis functions are used to improve the description of the other, which means the basis set localized on one monomer can behave as diffuse functions for the other monomer leading to a lowering of the computed energy. This is known as the basis superposition error (BSSE).[126, 127] The BSSE can be addressed either using a larger basis set which can be very expensive or using the counterpoise (CP) correction. The CP treatment tries to counteract this lowering in energy by performing all calculations in identical basis sets, and using additional 'ghost' centres to place basis functions at points corresponding to the nuclei of the two interacting monomers in a dimer calculation.[201] The supermolecular method with Density

Functional Theory (DFT) can be fast, however, BSSE remains an issue, reliable calculations require dispersion corrections[121] (**1.4.2.1**) and the intermolecular interaction energy still has no physically meaningful components.

2.2.1 The DFT energy

Kohn and Sham[202] computed the DFT energy $E[\rho]$ of a system as a functional of its electron density $\rho(r)$. Kohn proved that the DFT energy is exact if the exchange-correlation energy E_{XC} (which describes non-classical interactions between electrons, like dispersive interactions) is exact.

$$E[\rho] = E_{ext}[\rho] + T_s[\rho] + J_{ee}[\rho] + E_{XC}[\rho] \tag{9}$$

$E_{ext}[\rho]$ is the interaction energy between nuclei and electrons, $T_s[\rho]$ is the kinetic energy of electrons (the kinetic energy of the nuclei is zero as its position is fixed) and $J_{ee}[\rho]$ is the Coulomb energy due to the interaction between electrons. All functionals in Equation 9 are exact except for $E_{XC}[\rho]$, in principle it can be exactly calculated from the density, but this exact functional is unknown, thus it is approximated.

2.2.2 The exchange-correlation

While DFT is comparatively inexpensive, the approximate nature of its exchange functionals result in an intrinsic inaccuracy in calculations. Most DFT functionals are local (LDA) or semi-local (GGAs) in form, and can struggle to exactly define the long-range correlations required to encompass all important dispersion interactions. An erroneous treatment of the van der Waals forces can cause an underestimation in the attractive forces between the molecules in these crystals and thus an overestimation in the overall cell lengths. This has resulted in a number of functionals that use different approximations of $E_{XC}[\rho]$ to be applied on various systems. Local Density Approximation (LDA) and Generalised Gradient Approximation (GGA) are two groups of common functionals. In GGAs, $E_{XC}[\rho]$ is a functional of ρ and the density gradient.[203] Hybrid functionals that include different proportions of the Hartree-Fock exchange term are also used. In this thesis, and generally for computing molecular charge densities in CSP, the hybrid functional PBE0 is used.[204, 205]

2.2.3 Common Corrections

The most common problems with local and semi-local functionals are the self-interaction error[206] and its inability to describe the dispersion contribution at long-range. As there is no systematic way to improve the quality of DFT,[207] it is common to use system dependent corrections. The self-interaction error can be partially amended using an asymptotic correction (AC), which uses marginal additional computational resource. Tozer and Casida[208, 209] proposed to merge a $-1/R$ asymptotic tail to an exchange-correlation potential that does not display the appropriate asymptotic behaviour to correct it. The inexpensive asymptotic correction demonstrates considerable reductions in the average errors in excitation energies for molecules[210] by aiding in fixing the single electron self-interaction error, while the remaining error can be reduced by range-seperation.[208, 209]

The errors in the long-range dispersion can be partially addressed using a dispersion correction appropriate to the system studied. The simplest of these corrections is the Grimme 2006 correction.[122] There are also more complex models[41, 121, 211] and some post-hoc dispersion corrections like the Tkatchenko-Scheffler (TS) method[212] that includes pairwise interactions between atoms into the DFT functional. There is also the Many-Body Dispersion (MBD) correction; a marked improvement on the TS correction.[213]

Density Functional Theory (DFT) is not used to directly calculate the lattice energy in this thesis but it is used in the force-field development process. This thesis is not an electronic structure methods study so we shall not delve too deeply into the theory behind DFT and its contemporaries. While it is not as cheap as force-fields, its growing popularity[22] means that currently one cannot mention Crystal Structure Prediction without the mention of periodic DFT-D (**Chapter 1**). We use DFT chiefly to optimize isolated molecular structures (**Chapter 3** & **5**), calculate relative conformational energies (**Chapter 4**) and obtain the molecular orbitals to be used to construct our electrostatic model (**2.5**).

2.2.4 A note on basis-sets

The quality of the basis set is very important as the interaction energies are heavily dependent on the outer regions of the wave-function. There is a zeta 'ζ' hierarchy for the Gaussian basis sets typically used in electronic structure calculations. For instance, a triple-ζ (TZ) basis set is when three contracted sets of basis functions are used to describe each atomic occupied shell.[214] Often the virtual (unoccupied) orbitals are described using progressively smaller basis functions. The most commonly used high-quality basis sets are valence-only indicated by a 'V', and 'correlation-consistent polarization' basis sets indicated by a 'cc-p'. This means they include increasingly larger shells ($d, f, g\ etc.$) of polarized correlating functions. Augmentations to basis sets have also been developed to include additional, diffuse functions to better represent the polarizability of molecules and describe weak interactions.

Different basis set types, like monomer-centred (MC) and dimer-centred (DC) basis sets, can be used in calculating intermolecular interactions. In a MC basis set, each molecule in an interaction is described in the same basis used for the isolated molecule. Consequently, MC basis sets exclude excited states that describe charge-delocalization (**2.6.2**).[215] DC type basis sets are needed to accurately describe charge-delocalization states. In dimer-centred basis types, the Hessians and molecular orbitals (MOs) of each monomer in the dimer are made in the entire dimer basis. Functions between covalent bonds are sometimes included in DC basis sets (known as DC+) as they are very important for obtaining basis converged dispersion energies.[215] In monomer-centred basis types, the Hessians and MOs of each monomer are made in the basis of the monomer but often mid-bond and "far-bond" functions are included between two monomers (MC+), these can be smaller but near equivalent to DC basis types,[215] thus cheaper.

2.3 Perturbation theory

2.3.1 Raleigh-Schrodinger (RS) Perturbation Theory

The distortion of charge densities due to the intermolecular interactions between molecules is relatively small for organic molecules, thus the intermolecular interactions can be treated as perturbations of the unperturbed isolated molecular charge density.[164] Rayleigh-Schrodinger (RS) perturbation theory[216] uses this concept to derive the contributions to the interaction energy from the perturbation for the intermolecular interaction operator, $\mathcal{H}'$. RS perturbation theory approximates a solution to a non-trivial problem using the exact solution of another.[217] The Hamiltonian for the two interacting molecules can be given as

$$\mathcal{H}_{MN} = \mathcal{H}_M + \mathcal{H}_N + \lambda\nu$$

10

where the Hamiltonians for molecules M and N are $\mathcal{H}_M$ and $\mathcal{H}_N$, and ν is the interaction operator that defines the interaction between the electrons and nuclei of the two molecules.

$$\nu = \frac{1}{4\pi\epsilon_0}\sum_{i\in M}\sum_{k\in N}\frac{q_i q_k}{R_{ik}}$$

11

ϵ_0 is the permittivity in vacuum, the charges of the particles i and k in M and N are q_i and q_k, with a separation of R_{ik}. If the eigenfunctions and eigenvalues of the unperturbed Hamiltonians are known then, for two independent non-interacting molecules M and N in their unperturbed ground states, their time independent Schrödinger's equations is

$$\mathcal{H}_M^0\Psi_M^0 = E_M^0\Psi_M^0$$

12

$\mathcal{H}_M^0$ is the Hamiltonian of the unperturbed ground state, while E_M^0 is the unperturbed ground state energy of the unperturbed wave-function Ψ_M^0. For non-interacting two molecules, where $\lambda = 0$, the wave-function of the system is $\Psi_{MN}^0 = \Psi_M^0\Psi_N^0$, thus the resultant time independent Schrödinger's equation is

$$\mathcal{H}_{MN}^0\Psi_{MN}^0 = E_{MN}^0\Psi_{MN}^0$$

$$E_{MN}^0 = E_M^0 + E_N^0$$

13

The interactions between the electrons and nuclei of the two molecules result in the perturbation of the ground state ($\lambda \neq 0$). Thus, a perturbation expansion for the wave-function can be written as

$$\Psi = \Psi_{MN}^0 + \lambda\Psi_{MN}' + \lambda''\Psi_{MN}'' + \cdots,$$

14

and the perturbed Hamiltonian (up to first order) for the molecules is

$$\mathcal{H}_{MN} = \mathcal{H}_{MN}^0 + \mathcal{H}' = \mathcal{H}_M^0 + \mathcal{H}_N^0 + \mathcal{H}'$$

15

where $\mathcal{H}'$ is the 1st order perturbation due to interactions between the two molecules. From RS perturbation theory[216] the total intermolecular interaction energy, going from 0th order to infinity, between two bodies M and N can be written as

$$E = E_{MN}^0 + E_{MN}' + E_{MN}'' + \cdots$$

16

At first order the interaction energy can be given in terms of unperturbed densities $\rho(r)$,

$$E_{MN}' = \langle\Psi_M^0\Psi_N^0|\nu|\Psi_M^0\Psi_N^0\rangle$$

$$= \iint \rho_M(r_i)\,\nu(r_i, r_k)\rho_N(r_k)dr_i dr_k$$

17

This is the 1st order electrostatic interaction energy $E_{elst}^{(1)}$.The higher order perturbations give the other contributions to the intermolecular interaction energy, for example, the 2nd order energy E''_{MN} describes the induction and dispersion contributions, however, the influence of higher order terms above 2nd order greatly diminishes.[164, 216] The second-order energy E'' can be given by

$$E'' = -\sum_{mn\neq 00} \frac{\langle \Psi_M^0 \Psi_N^0 | \nu | \Psi_M^m \Psi_N^n \rangle^2}{E_m^M + E_n^N - E_0^M - E_0^N}$$

18

where Ψ^0 is the ground state wave-function, and Ψ^m and Ψ^n are the excited state wave-functions of quantum numbers m and n. Both m and n cannot equal 0. Excluding the term where both interacting molecules M and N are in the ground state, gives three terms to the 2nd order energy, which are the induction and dispersion energies

$$E'' = E_{ind}^M + E_{ind}^N + E_{disp}$$

19

The induction energy for M (E_{ind}^M) describes the change in the energy of the M in response to the electric field of N, and thus is when the change in the charge distribution of M is expanded in terms of its excited states and N is in the ground state, likewise E_{ind}^N is when molecule N is polarized and M is in its ground state. The dispersion energy E_{disp} is when both densities are distorted.

$$E_{ind}^M = -\sum_{m\neq 0} \frac{\langle \Psi_M^0 \Psi_N^0 | \nu | \Psi_M^m \Psi_N^0 \rangle^2}{E_m^M - E_0^M}$$

$$E_{disp} = -\sum_{m\neq 0}\sum_{n\neq 0} \frac{\langle \Psi_M^0 \Psi_N^0 | \nu | \Psi_M^m \Psi_N^n \rangle^2}{E_m^M + E_n^N - E_0^M - E_0^N}$$

20

The dispersion energy (**2.7**) and the polarization portion of the induction energy (**2.6.1**) can be derived from the polarizabilities defined in the second order energy, using the frequency-dependent density susceptibility function and expressed with the multipole expansion.[164]

The underlying issue with this treatment is that the unperturbed wave-function Ψ_{MN}^0 and the perturbed wave-function Ψ'_{MN} are not anti-symmetric under the permutation of electrons. They do not obey the Pauli exclusion principle as $R \rightarrow 0$. Including an anti-symmetrizer operator $\widehat{\mathrm{A}}$ for Ψ_{MN}^0, results in product states that are antisymmetric and satisfy the Pauli principle, however, an intermolecular exchange of electrons between M and N then results in wave-functions that are no longer eigenfunctions of the ground state Hamiltonian $\mathcal{H}^0$.[218] This can be remedied with symmetry-adapted-perturbation theories (SAPT) that can use $\hat{A}\Psi_{MN}^{0}$ in its perturbation equations, and can reliably give intermolecular energy contributions up the third order.[218]

2.3.2 Symmetry-adapted perturbation theory (SAPT)

Symmetry Adapted Perturbation Theory (SAPT) is a framework for directly computing the interaction energy using intermolecular perturbation theory. Here "symmetry" refers to the anti-symmetrisation between the interacting molecules. The most commonly used SAPT is based on symmetrized Raleigh-Schrodinger (SRS) perturbation theory.[219]

2.3.2.1 SAPT

When Symmetry Adapted Perturbation Theory is based on Hartree-Fock (HF) orbitals as the starting point for the perturbation theory we get what is commonly known as "SAPT". In this theory both intramolecular (**2.3**) and intermolecular correlations need to be included using perturbation theory. The Hamiltonian of two interacting molecules is given as

$$\mathcal{H}_{MN} = \mathcal{F}_M + \mathcal{F}_N + \xi W_M + \eta W_N + \lambda \nu$$

21

where $\mathcal{F}$ are the Fock operators, W the Møller-Plesset (MP) operators, and ν the intermolecular interaction operator.[164, 218] The Fock and MP decomposition of the Hamiltonian makes SAPT a triple perturbation theory. SAPT assigns a 0th order state to the wave-functions of isolated molecules and models the interactions between each molecule as perturbations in the system; therefore, the interaction energy between two molecules can be computed directly, avoiding the basis set superposition error (BSSE) inherent to electronic structure methods that exploit a cumulative supermolecular approach such as regular DFT. It does not need to compute the total energy of monomers or dimers (Equation 8). An advantage of this is that one only has to carry out correlation corrections to the order required by each individual term in the interaction energy.

2.3.2.2 SAPT(DFT)

SAPT(DFT) is Symmetry Adapted Perturbation Theory (SAPT) based on the Kohn-Sham density functional theory (DFT) of monomers.[100, 102] Kohn-Sham orbitals and orbital energies are used in the Hamiltonian to reduce the scaling of SAPT as intramolecular correlation effects on the interaction energy can be better described by DFT.[220] A few of the computationally demanding intra-monomer electron correlation terms found in SAPT do not have to be included when it is coupled with DFT to form SAPT(DFT). This is because the 0th-order monomer states, and hence the molecular orbitals, are computed at the DFT level, accordingly, the intramolecular correlation is included intrinsically.[104] The 1st order components of the SAPT(DFT) intermolecular energy, $E^{(1)}_{elst}$ and $E^{(1)}_{exch}$, can be computed using the following Hamiltonian

$$\mathcal{H}_{MN} = \mathrm{K}_M + \mathrm{K}_N + \lambda\nu \tag{22}$$

where K are the Kohn-Sham operators, and ν is the intermolecular perturbation. The above Hamiltonian as written is only approximate, however, SAPT(DFT) is formally exact (at second order). The SAPT(DFT) Hamiltonian can use the same terms as those used for RS perturbation theory in Equation 10. Subsequently, each term can then be expressed as properties that can be computed by DFT (i.e. density, density matrix, response functions etc.) and these are used to obtain exact SAPT(DFT) interaction energies. The second-order contributions to the intermolecular energy, $E^{(2)}_{IND}$ and $E^{(2)}_{DISP}$, can be computed using Kohn-Sham linear response theory as they are response energies (**2.6.1** & **2.7**).

Typically, SAPT(DFT) is faster than pure SAPT,[161] as the treatment of the electron correlation within the SAPT method using the triple perturbation theory can be expensive when developing potentials. Therefore, employing SAPT(DFT) offers a way to include intramolecular correlation effects cheaply and simply.[99, 221, 222] However, it is vital to use the appropriate asymptotically corrected function to account for the 1-electron self-interaction error that leads to overly diffuse densities, and a HOMO-LUMO gap that is too small, which results in errors in the derived intermolecular contributions. In this thesis, SAPT(DFT) is used to concisely describe the intermolecular interactions in the dimers of weakly-bound organic molecules, and the components of the SAPT(DFT) intermolecular interaction energy (***Table 4***) are used in force-field parameterisation.

In the force-field construction/parameterisation, SAPT(DFT) is used to generate numerical values for the short-range interaction terms, which can then be used in the fitting of an intermolecular force-field. Misquitta and Stone have shown how SAPT(DFT) can be combined with methods for distributed molecular properties to derive most of the terms in the intermolecular model, yielding accurate many-body models with a small number of SAPT(DFT) evaluations.[223, 224] Accordingly, SAPT(DFT) has been central to the derivation of our long-range terms and the generation of the short-range portion of the non-empirical force-fields we have created in this study. With SAPT(DFT) we have greatly increased the accuracy and sense of realism of our models.

	Contribution	
$E_{elst}^{(1)}$	Electrostatic	1st order electrostatic interaction energy classically derived from the unperturbed Kohn-Sham molecular charge distributed
$E_{exch}^{(1)}$	Exchange-Repulsion	1st order short-range interaction energy arising from closed-shell repulsion and electron exchange between overlapping ψ
$E_{ind,resp}^{(2)}$	Induction	2nd order energy arising from changes in molecular charge distribution due to the field from the unperturbed Kohn-Sham charge distribution of its neighbours
$E_{exch-ind,resp}^{(2)}$	Exchange Induction	The electron exchange portion of the induction energy
$E_{disp,resp}^{(2)}$	Dispersion	A quantum mechanical energy arising from correlated fluctuations in the Kohn-Sham charge distributions of neighbours
$E_{exch-disp}^{(2)}$	Exchange Dispersion	The electron exchange portion of the dispersion energy
δ_{int}^{HF}	Higher-order Induction Correction	An approximation of the 3rd to infinite order induction terms

Table 4: *Analysis and summary of the components of the intermolecular interaction energy calculated in SAPT(DFT).*[164] *The subscript 'resp' indicates that coupled perturbation is used, and is omitted elsewhere in this thesis.*

2.3.3 The SAPT(DFT) intermolecular interaction energy and its components

Typically, the most important contributions to the total interaction energy are first and second-order terms, with the higher order terms being very computationally expensive to compute accurately with almost negligible energy contributions.[167] For reasons described in detail in **Chapter 5**, the 3rd to infinite-order terms can be carefully omitted for a molecule like trinitrobenzene (TNB). Consequently, the SAPT(DFT) intermolecular interaction energy, is the interaction energy components summed to 2nd order. On the right hand side of Equation 23, "(n)" indicates the order of the contribution, however, on the left hand side "(2)" indicates the components of the interaction energy have been summed up to 2nd order. Thus, in this thesis, $E_{int}^{(1)}$ & $E_{int}^{(2)}$ (or any contribution on the LHS) indicates that the energy has been summed up to 1st or 2nd order.

$$E_{int}^{(2)} = E_{elst}^{(1)} + E_{exch-rep}^{(1)} + E_{IND}^{(2)} + E_{DISP}^{(2)}$$

23

where $E_{elst}^{(1)}$ and $E_{exch}^{(1)}$ are the 1st order electrostatic and exchange-repulsion energies respectively, while $E_{IND}^{(2)}$ is the total 2nd order induction energy, which includes the exchange-induction energy, and $E_{DISP}^{(2)}$ the total 2nd order dispersion energy, which includes the exchange-dispersion energy.

$$E_{IND}^{(2)} = E_{ind}^{(2)} + E_{exch-ind}^{(2)} \text{ and } E_{DISP}^{(2)} = E_{disp}^{(2)} + E_{exch-disp}^{(2)}$$

24

The SAPT(DFT) energy components benefit from having well-defined multipole expansions (**2.4**), and are usually employed in the distributed sense.[99, 164, 225, 226] This allows one to determine the long-range forms of these terms analytically from these expansions. That leaves only the damping and the short-range penetration and exchange terms to be derived through fitting. The long-range multipolar electrostatic potential energy $V_{elst}^{(1)}[DM]$ can be given as $\sum_{i \in M, k \in N} Q_t^i T_{tu}^{ik} Q_u^k$ where Q_t^i is a multipole moment on atomic site i of rank t, using the angular momentum notation $t = 00, 10, 11c, \ldots$ to describe the multipole moment and its related axis (**2.4.1**).[164] T_{tu}^{ik} is the interaction tensor[164] determined by our axes definitions (***Figure 6***). The same spherical tensor notation (tu) used for multipole components is used for the polarizabilities and transformation tensors. The long-range multipolar dispersion energy $V_{disp}^{(2)}[DM]$ is determined exactly by a sum of isotropic, orientation independent atom-atom dispersion coefficients $-\sum_{n=0}^{2} \frac{C_{2n+6}^{\iota\kappa}}{R_{ik}^{2n+6}}$ and a damping parameter β_{disp} fit from $E_{DISP}^{(2)}$ calculations (**5.5.4**). The final long-range term, the iterated multipolar polarization energy, is the long-range portion of the induction energy, which is dependent on the atomic multipole moments and atomic polarizabilities. This is the only many-body term in the potential, although it is written in a two-body form, many-body polarization effects are included when the polarization model is iterated.[47, 164]

$$V_{POL}^{(2-\infty)}[DM] = \sum_{i \in M, k \in N} f_n\left(\beta_{pol}, R_{ik}\right)\left(\Delta Q_t^{\iota} T_{tu}^{\iota\kappa} Q_u^{\kappa} + \Delta Q_t^{\kappa} T_{tu}^{\kappa\iota} Q_u^{\iota}\right)/2$$

$$\Delta Q_t^{\iota} = -\sum_{i \in M, k \in N} \alpha_{tt'}^{\iota\iota'} f_n\left(\beta_{pol}, R_{ik}\right) T_{t'u}^{\iota'\kappa}\left(Q_u^{\kappa} + \Delta Q_u^{\kappa}\right)$$

25

This term is damped using a damping function that is independent of orientation, and a damping parameter β_{pol}, which is determined using regularized induction energies $E_{IND}^{(2)}(Reg)$. The induced moment ΔQ_t^{ι} on atom-type ι

depends on the distributed polarizability ($\alpha_{tt'}^{\iota\iota'}$) for sites ι and ι'. $\alpha_{tt'}^{\iota\iota'}$ describes the response of Q_t^{ι} to the t'-component of the field at ι', and in order to obtain the induced moment on ι we must solve Equation 25 iteratively (**2.6.1.1**). As written Equation 25 uses the non-local polarizability $\alpha_{tt'}^{\iota\iota'}$, however, in this thesis the polarizability is localized and $\alpha_{tt'}^{\iota\iota'}$ is replaced with $\alpha_{tt'}^{\iota}\delta_{\iota\iota'}$, where $\alpha_{tt'}^{\iota}$ is the localized polarizability tensor and $\delta_{\iota\iota'}$ is the Kronecker delta (**2.6.1.1**).[165, 227]With this in mind, using the axis definitions and atomic types in ***Figure 6***, and a combination of programs: CamCASP,[107] Psi4,[228] DMACRYS[59] and ORIENT[108] (**1.2.4**), the total interaction energy can be represented as non-empirical potential, whose long-range components are represented using the distributed multipole expansion[164] and the short-range components represented as an exponential with the following functional form

$$U_{M,N}(R_{MN},\Omega_{MN}) = \sum_{i\in M,k\in N} U_{INT}^{ik}(R_{ik},\Omega_{ik}) = G\exp\left[-\alpha^{\iota\kappa}\left(R_{ik}-\rho^{\iota\kappa}(\Omega_{ik})\right)\right] + Q_t^{i}T_{tu}^{ik}Q_u^{k} - \sum_{n=0}^{2} f_{2n+6}\left(\beta_{disp},R_{ik}\right)\frac{C_{2n+6}^{\iota\kappa}}{R_{ik}^{2n+6}}$$

$$+f_n\left(\beta_{pol},R_{ik}\right)(\Delta Q_t^{\iota}T_{tu}^{\iota\kappa}Q_u^{\kappa} + \Delta Q_t^{\kappa}T_{tu}^{\kappa\iota}Q_u^{\iota})/2$$

26

2.4 Multipole moments

2.4.1 Defining molecular multipole moments

The molecular charge density can be represented as multipole moments (charge, dipole, quadrupole (Q_{lm}) etc.), which are defined relative to a central position in the molecule. The first non-zero moment is independent of centre, so for example, the dipole moment of carbon monoxide and the quadrupole moment of carbon dioxide are independent of centre. The lowest ranked molecular multipole moment, the total molecular charge (q_M), can be written as the integral of the molecular charge density $\rho_M(r)$ with respect to position r.

$$q_M = \int \rho_M(\boldsymbol{r})\, d\boldsymbol{r}$$

27

The dipole moment, is a rank 1 tensor (a vector quantity) in the x, y, z directions and the total dipole moment is composed of three components

$$\boldsymbol{\mu} = \left(\mu_x, \mu_y, \mu_z\right)$$

28

For either of the three components α, to obtain the value of the dipole moment in the ground-state wave-function, $|\Psi^0\rangle$, one can take the expectation value of the dipole moment operator $\hat{\mu}_\alpha$

$$\mu_\alpha = \langle\Psi^0|\hat{\mu}_\alpha|\Psi^0\rangle$$

29

This can also be expressed as an integral of the molecular charge density $\rho_M(r)$ over electronic co-ordinates to give the dipole moment for a fixed nuclear configuration, where z component of the dipole moment is expressed as

$$\mu_z = \int \rho_M(\boldsymbol{r}) z \, d\boldsymbol{r}$$

30

The subsequent multipole moment is the quadrupole moment. A quadrupolar charge distribution contains four charges of equal magnitude (two positive and two negative). It is a rank 2 tensor (a 3×3 matrix) with 9 components ($\Theta_{xx}, \Theta_{yy}, \Theta_{zz}, \Theta_{xy}$ ...) and is traceless ($\Theta_{xx} + \Theta_{yy} + \Theta_{zz} = 0$). In Cartesian tensor notation for molecule M in position R with charge q_M the operator of the quadrupole moment can be defined as

$$\widehat{\Theta}_{\alpha\beta} = \sum_M q_M(\frac{3}{2}R_\alpha R_\beta - \frac{1}{2}R^2\delta_{\alpha\beta})$$

31

α and β are the x, y or z components and $\delta_{\alpha\beta}$ is the Kronecker delta, where $\delta_{\alpha\beta} = 0$ if $\alpha \neq \beta$ and 1 if $\alpha = \beta$. The tensor notation is very useful as it makes manipulating multipole moments much easier. We see that the quadrupole moment is not only traceless but also symmetric as $\Theta_{\alpha\beta} = \Theta_{\beta\alpha}$. Hence there are only 5 independent non-zero components of the quadrupolar moment of an asymmetrical molecule. For a multipole moment of $Rank = l$ there are $2l + 1$ independent moments, which is also the number of Rank l spherical harmonics. They are directly proportional.[229] Defining multipole moments in terms of spherical harmonics makes advanced applications of them easier.[164] In spherical tensor notation, a multipole moment operator can be defined as

$$\widehat{Q}_{l\kappa} = \sum_M q_M R_{l\kappa}(\mathbf{r})$$

32

l denotes tensor rank, κ the component (a member of the series $0, 1c, 1s, 2c, 2s$ ...) and $R_{l\kappa}(\mathbf{r})$ are the spherical harmonics, and the ground state expectation value is

$$Q_{l\kappa}^{M} = \int \rho_M(\mathbf{r})\, R_{l\kappa}(\mathbf{r})\, d^3\, \mathbf{r} \qquad 33$$

2.4.2 The validity of the multipole expansion

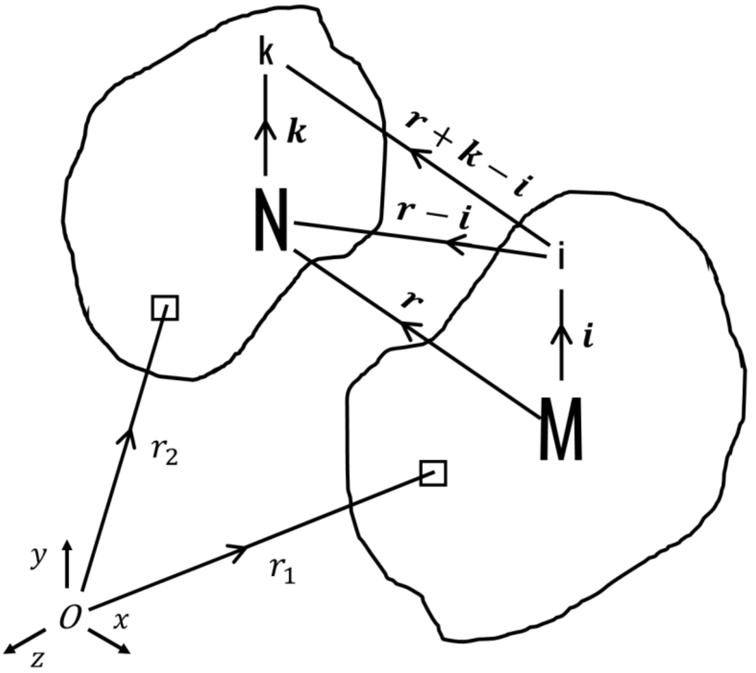


Figure 5: *An illustration of the interacting molecular charge densities of molecules M and N in a space fixed axis system with origin O. Molecules M and N contain atoms i and k respectively. The positions of atoms i and k is given by the position vectors r_1 and r_2. The distributed contributions to the molecular charge density of M and N by atoms i and k at these positions can be given as $\rho_i(r_1)$ and $\rho_k(r_2)$ respectively. It is also important to specify whether the directionality of the interaction is **MN** or **NM** as this changes our definition of **R**.*

Consider molecule M with a centre of mass in a position **M** from the origin O, and the atomic-centres i in the molecule at positions **i** relative to **M** (***Figure 5***). The potential at the position **N** can be calculated by the integration over all contributions to the total molecular charge density ($\rho_M(\boldsymbol{r})$) of the molecule M at position **M** over all-space.

$$V^M(\mathbf{N}) = \int \rho_M(\boldsymbol{r}) \frac{1}{4\pi\epsilon_0 |\boldsymbol{r} - \boldsymbol{N}|} d^3\, \boldsymbol{r}$$

The electrostatic energy from the total ground-state charge densities of the two interacting molecules M and N is hence given as

$$E_{elst} = \int \frac{\rho_M(r_1)\rho_N(r_2)}{4\pi\epsilon_0|r_1 - r_2|} d^{\mathbf{3}}\, r_1\, d^3\, r_2$$

35

Equation 35 uses the molecular charge densities of M and N, but one can partition the electrostatic potential into its atom-atom contributions by partitioning the total charge density into regions centred on atoms i and k. We use distributed multipole partitioning methods like GDMA[230] or ISA[231-233] to do so (**2.5.1**). The distributed multipole expansion is an improvement on central multipole expansions as it converges better at short range (small R).[164] The standard multipole expansion of the electrostatic potential about the centre (M) typically converges at distances that are much further from the origin than the furthest point of the charge distribution itself, depending on the order of the first non-zero multipole moment.[234] This is not so much a problem in gas-phase studies as neighbouring molecules are separated by large distances, however, it is a major issue in the condensed phases of non-spherical molecules, where neighbours are in van der Waals contact, and the multipole expansion is not valid, let alone convergent, giving an unrealistic representation of the potential outside of the charge distribution. Hence, it is necessary to split the molecular charge distribution $\rho_M(r_1)$ into atomic contributions $\rho_i(r_1)$. The electrostatic interaction can hence be written as

$$E_{elst} = \sum_{i\in M} \sum_{k\in N} \int \frac{\rho_i(r_1)\rho_k(r_2)}{4\pi\epsilon_0|r_1 - r_2|} d^{\mathbf{3}}\, r_1\, d^3\, r_2$$

36

where the sum of the atomic electron densities is the total molecular electron density $\sum_i \rho_i\,(r) = \rho_M(r)$. Once partitioned, the multipolar moments can be written in terms of the atomic positions. For example, the potential of M can be written as a sum of $Rank = 0$ atomic-moments (charges) of i in position **i** in molecule M, which can then be expanded as a Taylor series about R to include the higher multipole moments. Expanding the Taylor series allows us to define the potential as distributed multipole moments and their respective interaction tensors ($Rank = 0$ (Q_{00}) = charge, Rank 1 ($Q_{1\kappa}$) = dipole, Rank 2 ($Q_{2\kappa}$) = quadrupole etc.). For instance, the charge ($l = 0$), dipole ($l = 1$) and quadrupole ($l = 2$) moment contributions to the electrostatic potential at **N** due to molecule M can be written as

$$V^M(\mathbf{N}) = Tq^M - T_\alpha \hat{\mu}_\alpha^M + \frac{1}{3} T_{\alpha\beta} \widehat{\Theta}_{\alpha\beta}^M \dots$$

37

where $\hat{\mu}_\alpha^M$, $\widehat{\Theta}_{\alpha\beta}^M$ are the dipole and quadrupole moment operators, and the multipole moment interaction tensors, $T = 1/4\pi\epsilon_0 R, T_\alpha = -R_\alpha/4\pi\epsilon_0 R^3$ and $T_{\alpha\beta} = (3R_\alpha R_\beta - R^2\delta_{\alpha\beta})/4\pi\epsilon_0 R^5$.[164] Knowing the potential V^M at point **N** due to molecule M, one can evaluate the electrostatic interaction operator (ν) between M and say a point charge N is at position **N**. In Cartesian form this can be written as

$$\nu = q^N V^M(\mathbf{N})$$

38

Expanding the electrostatic potential of M for the higher order multipole moments gives

$$\nu = q^N [Tq^M - T_\alpha \hat{\mu}_\alpha^M + \frac{1}{3} T_{\alpha\beta} \widehat{\Theta}_{\alpha\beta}^M - \cdots]$$

39

The electrostatic interaction energy E_{elst} of two bodies in non-degenerate states is the expectation value of ν and is obtained by replacing each multipole operator ($\hat{\mu}_\alpha^M$, $\widehat{\Theta}_{\alpha\beta}^M$) by its expectation value, $\mu_\alpha^M, \Theta_{\alpha\beta}^M$. One can write these expressions for an assembly of molecules (a crystal) by summing over the distinct atom pairs, and the explicit formula for the interaction between the multipole moments of N with those of M would be a more elaborate equation than that shown above. The multipole moments of a molecule can be described in either a Cartesian or Spherical tensor formulation. Using the spherical tensor formulation and indexing the *molecular* multipole moments by angular momentum labels the interaction operator can be written as

$$\nu = \hat{Q}_t^M T_{t\,u}^{MN} \hat{Q}_u^N$$

40

$T_{t\,u}^{MN}$ is the interaction tensor function between the moments of M and N and depends solely on the relative positions of their molecular axis systems and their separation R, varying with the ranks of the two multipoles l_1 and l_2 as $R^{-(l_1+l_1+1)}$[164]. It is the spherical tensor formation ($Q_{l\kappa}$) that we use in DMACRYS. For manipulation and many other purposes using the spherical tensor form for the electrostatic interaction is more useful. Considering the induction energy of molecule M from Equation 20

$$E_{ind}^{M} = -\sum_{m\neq 0} \frac{\langle \Psi_M^0 \Psi_N^0 | \nu | \Psi_M^m \Psi_N^0 \rangle^2}{E_m^M - E_0^M}$$

$$= -\sum_{m\neq 0} \frac{\langle \Psi_M^0 | \nu[\rho_N] | \Psi_M^m \rangle^2}{E_m^M - E_0^M}$$

41

The induction energy is the 2nd order energy response of M to the electrostatic field of its neighbour N. The neighbour does not necessarily have to be another molecule but could be replaced with the electrostatic potential of the environment. Using the multipole expansion in long-range (only up to the dipole-dipole moment), intermolecular interaction operator ν is

$$\nu = Tq^M q^N + T_\alpha (q^M \hat{\mu}_\alpha^N - \hat{\mu}_\alpha^M q^N) - T_{\alpha\beta} \hat{\mu}_\alpha^M \hat{\mu}_\beta^N + \cdots$$

42

Substituting ν into Equation 20 yields the polarizability, also known as the charge-density susceptibility, $\alpha_{\alpha\alpha'}$

$$E_{ind}^{N} = -\frac{1}{2}\left(q^M T_\alpha - \hat{\mu}_\beta^M T_{\alpha\beta} + \cdots\right) \alpha_{\alpha\alpha'}^{N} \left(q^M T_{\alpha'} - \hat{\mu}_{\beta'}^M T_{\alpha'\beta'} + \cdots\right)$$

$$\alpha_{\alpha\alpha'}^{N} = \sum_{n\neq 0} \frac{\langle 0 | \hat{\mu}_\alpha^N | n \rangle \langle n | \hat{\mu}_{\alpha'}^N | 0 \rangle}{E_n^N - E_0^N}$$

43

and integrating over the coordinates of M will yield the expectation values of the multiple moment operators. $\alpha_{\alpha\alpha'}$ is distributed among atomic sites in the molecule fixed axis frame to give the distributed polarizability model. The multipolar expansion of the intermolecular interaction operator, $\mathcal{H}'$, is only valid outside the charge distribution of the molecule. For most molecules, the charge density is far from spherical (***Figure 6***) and formally extends to infinity with the exponential decay of the charge distribution with radius (***Figure 8***).

2.4.3 Co-ordinate systems

There is an an orientation dependence to the intermolecular atom-atom interactions. In **Chapters 3** & **5** we see that atomic shape (anisotropy) is crucial to accurately describe interactions, differences in crystal structures/polymorphs and their associated properties. Therefore, it is important to define the co-ordinate systems used as to define our orientations. It is vital to define these precisely and consistently to enable replication of results and the transfer of the non-empirical force-fields. In this study, we will use atom fixed, molecule fixed and space fixed axes. The atom fixed axis is on each atomic site and is also referred to as the local axis in this thesis. It is attached to the distributed atomic sites within our molecule. There is a molecule fixed axis with an origin at the centre of mass of the molecule. It is molecule fixed as it moves with the molecule and within the force-field development process and the lattice energy minimization process the same molecule fixed axis will be used. The direction of both the local and molecule-fixed axis is referenced by atom positions and one should aim to set axes to simplify symmetry terms within the potential. The axes determine the $\mathrm{T}_{\mathrm{tu}}^{\mathrm{ik}}(\mathrm{R}_{\mathrm{ik}}, \Omega_{\mathrm{ik}})$ tensors in the theory of intermolecular forces and help define anisotropic terms, which are used to represent the non-sphericity of atoms. The space fixed axis is defined by x, y, z Cartesian axes, and its direction defined in reference to macroscopic features such as the electric field.

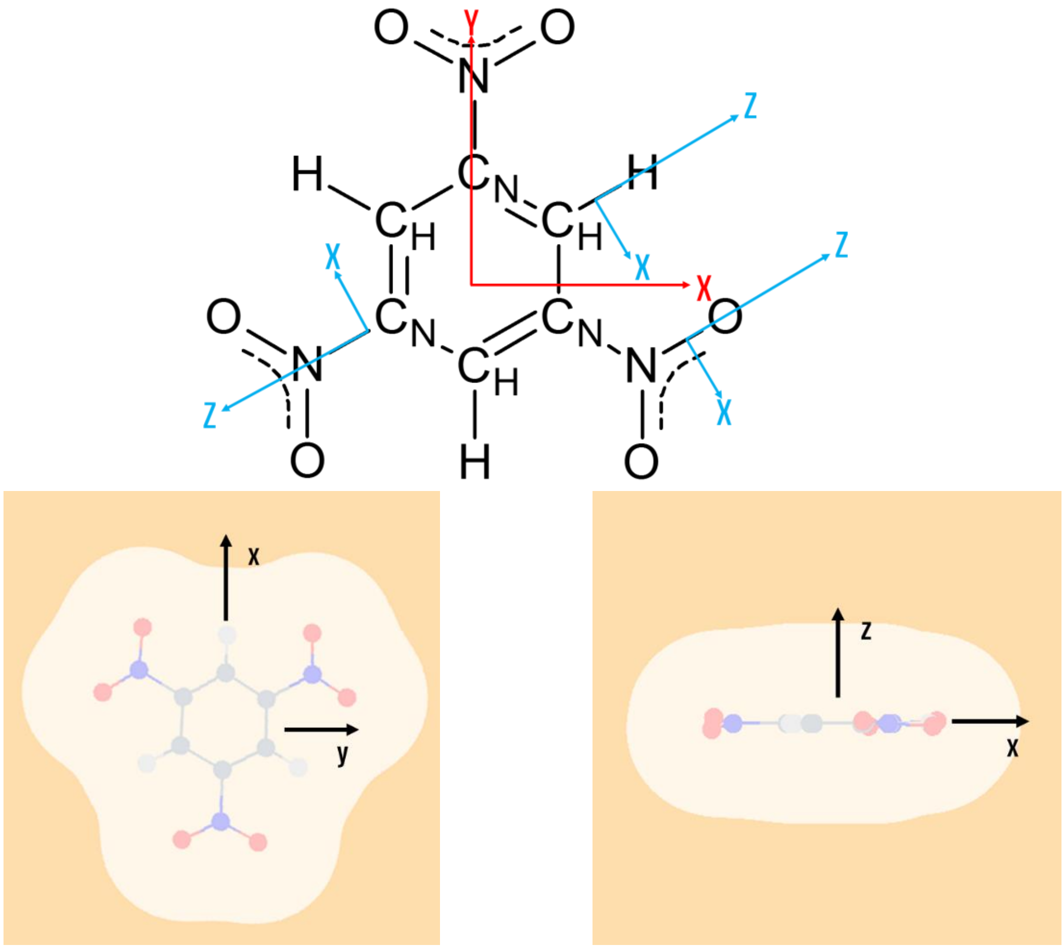


Figure 6: *(Top) The molecule-fixed axis (red) and local axis (atom fixed, blue) definitions used for the energetic organic crystal trinitrobenzene (TNB) to derive the atomic multipoles and polarizabilities. The axes are defined by the atoms involved for use in DMACRYS. Some programs use other axis systems by default, such as the inertia axes,*[235] *which change with conformation. The atomic properties must be transformed to the correct axis system (see* ***Chapter 3*** *for pyridine example). (Bottom) The charge density and shape of TNB, in the molecule-fixed axis.*

Figure 6 is an example co-ordinate system that is used in **Chapter 5** & **6.** Setting a reproducible co-ordinate system allows us to now progress onto the development of our non-empirical intermolecular force-fields.

2.5 The electrostatic model

The electrostatic potential around a molecule can be calculated from the charge density distribution of a molecule. It can be evaluated by considering the potential of the molecule at a point $\mathbf{M}$ in a global co-ordinate system, outside an iso-density surface of surface of 10^{-3} $electron/bohr^3$ around the static isolated molecule. The electrostatic potential is evaluated on a surface of this size for the sake of comparison to previous studies.[137, 138, 236-239] The charge distribution within the molecule $\rho_M(\mathbf{r})$ can be partitioned into atom-centred electron densities $\rho_i(\mathbf{r})$ whose atomic multipole moments can be used in an electrostatic model to describe the electrostatic interactions between neighbouring molecules in a crystal.[164]

2.5.1 Methods of partitioning

Distributed Multipole Analysis (DMA) is a description of the molecular charge density for calculating the electrostatic potential of a molecule at points *outside* (long-range) the charge density itself. Using a distributed multipole expansion assumes that the potential is being evaluated outside the charge density of the molecule and requires a suitable partitioning of the computed molecular charge density into atomic densities. The method of partitioning is vital to properly describing the behaviour of a system, and there are many methods of doing so.

2.5.1.1 Mulliken Population Analysis (Point Charges)

Mulliken Point Analysis (MPA)[240] was historically used in approximating the atomic partial charges, however, in comparison to other methods, the use of Mulliken charges for estimating the charge at a given atom is notoriously dependent on basis set and conformational changes.[241-243] MPA partitions the molecular density to obtain solely atom-centred point charges[240, 244] to model the entire charge distribution of a molecule, which does not capture spatial contribution of the charge distribution at all. MPA's estimation of higher order molecular moments like dipoles or quadrupoles has been found to be erroronous,[244] and higher order multipole moments are crucial to describing the bonding behaviour seen in weakly bound organic materials.

2.5.1.2 Charges from Electrostatic Potentials using a Grid-based method (CHELPG)

The Charges from Electrostatic Potentials using a Grid-based method (CHELPG) is an empirical approach developed by Breneman and Wiberg,[163] in which the partial atomic charges are obtained via fitting to a grid of points from an ab initio derived electrostatic potential of the isolated molecule. CHELPG aims to give the best reproduction of the electrostatic potential around the molecule,[236, 245] however, these charges are not directly linked to the charge distribution around the atom and are less physically interpretable. Modelling the atomic charge density simply as spherical atomic point charges completely neglects penetration effects. Moreover, the resultant atomic charges will be most accurate and heavily dependent on the grid of points to which they have been fitted, and lack transferability between molecules and even conformers.[236, 245] The molecular charge distribution is known to greatly change with conformation, thus charges fitted to one conformer transfer very poorly to others. This is less of an issue with the atomic charges calculated from the Iterated Stockholder Atoms (ISA) approach used in this thesis (**Chapter 4**, ***Figure 25***).[233] The best approach for a model is to obtain multipole moments from the distributed multipole analysis of a high quality wave-function rather than to fit to molecular electrostatic potential surfaces. Additionally, CHELPG does not work well in reproducing the atomic charges of internal atoms in large systems that are not close to the points outside the molecule where the electrostatic potential surface is calculated.

2.5.1.3 Gaussian Distributed Multipole Analysis (GDMA): GDMA1 & GDMA2

The first Gaussian type orbital (GTO) treatment of distributed multipole analysis was by Lavery et al,[246] however, Gaussian Distributed Multipole Analysis (GDMA) was one of the first electrostatic models to include higher order multipole moments, and GDMA1 (then known as DMA) was used to investigate dipole and up to hexadecapole terms in hydrogen fluoride.[234] These additional terms allowed one to investigate how the atomic charge distribution changes with bond formation. GDMA electrostatic models are classically coupled with empirical potentials, which absorb the effects of penetration. In GDMA, the charge density is expressed as a density matrix containing the product of two GTOs, $\phi_i(r)$ and $\phi_k(r)$, with an expansion coefficient C_{ik}. The product of the Gaussian orbitals is used to construct the atomic wave-function.

$$\rho_M(\boldsymbol{r}) = \sum C_{ik}\phi_i(\boldsymbol{r})\phi_k(\boldsymbol{r}) \tag{44}$$

Each atom is the logical centre of its own multipole contributions but what about the centres for the bond contributions? When ϕ_i and ϕ_k are not on the same atoms, a complete atomic multipole expansion could be created by moving the bond contributions from the distributed expansion to neighbouring atoms, which may not necessarily be the atoms at the end of the bond. With GDMA one can shift some bond multipoles to neighbouring atoms (based on the molecular chemistry) or have mid-bond multipoles.[234] Atomic terms (when both orbitals ϕ_i and ϕ_k are on the same atom) remain centred at the atoms and all contributions to bonding are collected into a single multipole expansion about the centre of the bond. By using just s and p-type Gaussian functions higher order multipole moments (octupole, hexadecapole etc.) will be created by the shifting, and their magnitudes depend on the how far from the atom centre the contributions are moved. In principle, this results in a more accurate representation of the electrostatic potential outside the charge distribution that relates to the atom.

GDMA1 gives a more complete generalization of Mulliken Population Analysis (MPA). For instance, lone pairs can be represented by a dipole or quadrupole moment on an atom, which is more physical/intuitive, compared to contributing to the atomic point charge, which results in the misrepresentation of higher multipoles. While it is a relatively modest representation of charge distribution, it can be evaluated quickly, effectively and provide a great foundation to build new methods of multipole analysis. However, there was the caveat that higher terms were very sensitive to small errors in the wave-function and fluctuated heavily with basis set.[247] Calculated moments changed significantly with basis set quality even though the computed electrostatic potential did not.[247] In the newer approach of GDMA, GDMA2, the more compact functions are handled in the same way as in GDMA1. However, GDMA2 stabilizes fluctuations in higher multipole moments due errors in the wave-function and changing basis sets by integrating the more diffuse contributions to the molecular electron density using a grid-based quadrature.[247]

2.5.1.4 Hirshfeld (Stockholder approach)

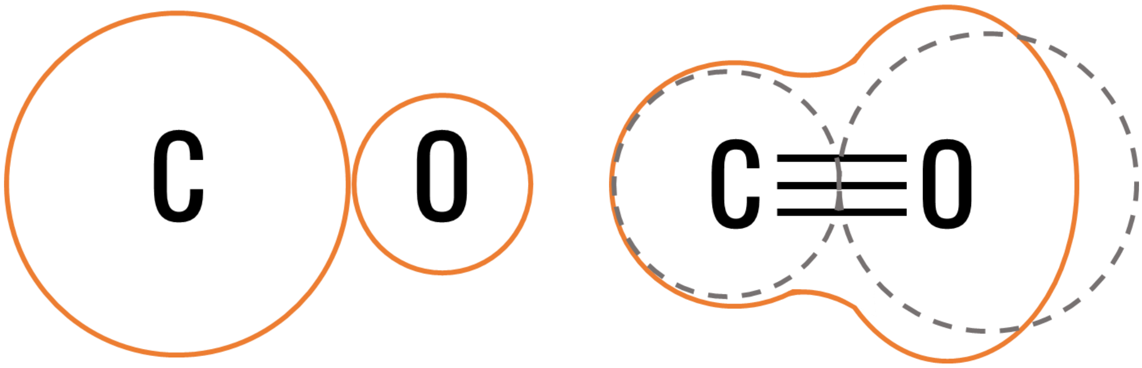
The Hirshfeld scheme[248] is a straightforward one that yields atomic charge densities that are closely related to those in the isolated molecule.[249] The Hirshfeld scheme uses a weighting function w_i on the molecular charge density $\rho_M(r)$ to partition it into an atomic density $\rho_i(r)$. It does not iterate these weighting functions. Instead it uses reference charge densities in its partitioning.

$$\rho_i(r) = w_i(r)\rho_M(r) = \frac{\rho_i^0(r)}{\rho_{mol}^0(r)}\rho_{mol}(r) \tag{45}$$

$\rho_i^0(r)$ is the ground state atomic density, and the reference charge density for atom i, which is usually the free atomic density. $\rho_{mol}^0(r)$ is the promolecular density, the sum of the $\rho_i^0(r)$, and $\rho_{mol}(r)$ is the true molecular density. This results in two key disadvantages:

- The Hirshfeld approach requires pre-calculated shape functions, which are obtained from *free* atom calculations i.e. free carbon or free oxygen and are consequently spherical.
- Thus, it cannot describe well the changes in atomic electron densities that occurs with chemical bonding.

For example, the electron densities of free carbon and oxygen atoms are very different from the O and C in carbon monoxide (***Figure 7***).



(a) *(b)*

Figure 7: *(a) the charge distribution around a free carbon atom and, the more electronegative, free oxygen atom. (b) the charge distribution around a carbon monoxide molecule. One can see that the electronegative O pulls charge density towards it. The spherical averaging in the Iterated Stockholder Atoms (ISA) approach*[231] *addresses this issue by trying to find the optimum atomic density shapes to best represent the molecular charge density*.

The more electronegative oxygen free atom has a less diffuse electron density ρ, however in carbon monoxide, the more electron-withdrawing oxygen pulls more electron density around itself so the ρ is more diffuse (***Figure 7***). The Hirshfeld approach fails to capture this, and raises other questions. The computed charges tend to be very close to 0, and seemingly unphysical. It can be debated that this is possible as the exact charges of atoms are unknown in reality (charges don't really exist), however other models (HI, ISA, GISA, GDMA2, BS-ISA) show larger atomic charges.[231, 233] This issue stems from the fact that the pre-calculated Hirshfeld weighting factors are derived from the isolated atom. Furthermore, the atomic charges are only available for neutral molecules as choosing a promolecular density $\rho^0_{mol}(r)$ for charged systems is non-trivial.[250, 251] The method of weighting the atomic charge densities was made iterative to combat this in the HI stockholder scheme.[243, 252]

2.5.1.5 Iterated Hirshfeld (HI) Stockholder approach

The Iterated Hirshfeld (HI) approach, dubbed Hirshfeld-I (HI), modifies the Hirshfeld approach using by iterating the weight functions using a simple iterative method,[252] where w_i is a weighted average of the neutral atom i and ionic densities.[252]

$$w_i(r) = \frac{\rho^0_i(r)}{\rho^0_{mol}(r)} \tag{46}$$

In the HI approach the atomic electronic population A_i (number of electrons per atom), dubbed the Hirshfeld population is calculated from the molecular density $\rho_i(r)$.

$$A_i = \int \rho_i\,(\boldsymbol{r})d\boldsymbol{r}$$

47

The isolated atomic density is then constructed by calculating the weighted average of atomic density from the reference promolecular density. These new isolated atomic densities are summed to create a new promolecular density that is then used to re-compute new atomic weighting functions. This procedure is iterated until the atomic electronic population is identical to the isolated atomic density for each atom. This works well generating large atomic charges but requires previous knowledge of neutral atomic densities and ionic densities. A principal benefit of this methodology, as with ISA, is that this does not depend explicitly on the basis set.[253] Furthermore, the number of ionic densities required depends on the number of ionic states required to describe the majority of the atom's ionic density.[243] The spherical averaging of the stockholder atomic density around each nucleus (Equation 48) and the fact the ISA does not use reference densities at all makes it an improvement on the HI approach.

2.5.1.6 Iterated Stockholder Atoms (ISA) Approach

The Iterated Stockholder Atoms (ISA) approach is another impressively simple yet effective method for obtaining an atomic electron density $\rho_i(r)$ that is capable of reproducing the molecular electron density.[231, 233] The ISA method partitions the molecular density according to spherically symmetric weighting functions centred at the atomic nuclei within a given molecule.

$$\rho_i(r) = \rho_M(r)\frac{w_i(r)}{\sum_k w_k(r)}$$

48

where $\rho_i(r)$ is the atomic electron density of atom i at position r, $\rho_M(r)$ is the molecular electron density, and $w_i(r)$ is the spherically symmetrical weighting function around the nucleus of atom i, also referred to as the shape function.[233] The sum $\sum_k w_k(r)$ includes shape functions of all the other atoms in the molecule. The w_k shape function is obtained by calculating the electron density of the free atom k in the gas-phase, and subsequent spherical averaging about the nucleus to mimic the spherical atom shapes. This spherical averaging requires either an extra calculation on each atom or a pre-calculated table of electron densities. This methodology is similar to the Hirshfeld approach,[252] however, the shape functions in the ISA method are defined by iterating the stockholder procedure from an initial guess, any initial guess in fact. In the subsequent iteration the shape functions are taken to be the spherical average of the stockholder atoms around each nucleus, thus

$$w_i(r) = \langle \rho_i(r) \rangle_{sph}$$

49

The only requirement for the ISA approach to work is that the shape function, $w_i(r)$, is positive and can be integrated over all space. Accordingly, the atomic electron densities are positive everywhere as long as the initial shape functions are positive. The angular brackets "$\langle \quad \rangle_{sph}$", represent the spherical averaging around the centre of atom i. These new weighting functions are inserted into Equation 46 to give a new set of stockholder atoms, to be spherically averaged again using Equation 48. This process is repeated until both equations can be solved simultaneously, indicating convergence. The conventional ISA approach is found to converge within 30-50 iterations for most small molecules.[231]

The principal benefits of the ISA method, and other iterated stockholder approaches, are that computed atomic electron densities, $\rho_i(r)$, are typically reasonable:[231, 233]

- They generate approximately spherical atoms, so that the higher atomic multipole moments are less important.
- A better distribution scheme, with electron densities that decrease smoothly away from the atomic centre.
    - The ISA-computed atomic electron densities display radial behaviour, near-exponential decay (***Figure 8***).
- They produce more natural, physical atomic charges than Mulliken Population Analysis and DMA methods,[233] which can be obtained by subtracting the integral of the atomic electron density over all space from the nuclear charge N_i

$$q_i = N_i - \int \rho_i(r)\, dr \tag{50}$$

    - Electronegative atoms have negative charges, making the ISA charges more realistic than MPA or DMA charges
- The ISA partitioned atomic densities can also be used to obtain distributed 2nd order quantities like the frequency dependent atomic polarizabilities, in a methodology referred to as the ISA-Pol (**2.6.1.1**).[200]

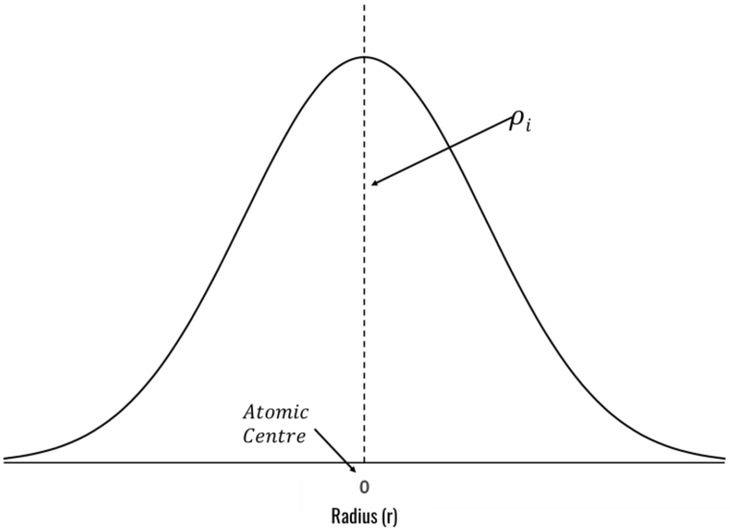


Figure 8: *The atomic electron density modelled by ISA. By using a separate and large basis set for better density tails combined with iterative spherical averaging helps capture the exponential decay of the charge distribution better.*

2.5.1.7 Basis-Space ISA (BS-ISA)

For small molecules like water, hydrogen sulphide or nitrogen, the conventional ISA converges within 30-50 iterations, however the number of iterations before convergence grows quickly with molecule size.[233, 243] The rate of convergence can be improved by using basis set expansions for all the quantities in Equation 46. Though charge conservation is not an issue if the partitioning of the atomic densities is done in real space, convergence is substantially faster in basis space, taking just 10 iterations in some cases.[233] This is due to the spherical (s-shaped) averaging of the stockholder atoms. This does add some basis dependence to the method, but this converges quickly with basis-set, and only using a small unbalanced basis set could lead to poor convergence.

$$\tilde{\rho}(r) = \sum_k d_k \mathrm{X}_k(r) \tag{51}$$

where $\tilde{\rho}(r)$ is the molecular density expanded in basis space, and d_k is the expansion coefficient of the auxiliary basis set $\mathrm{X}_k(r)$. The expansion of the atomic electron density can thus be given by

$$\rho_i(r) = \sum_k c_k^i \varepsilon_k^i(r) \tag{52}$$

Since the atomic density is expanded in a basis of spherical Gaussian Type Orbitals (GTOs), the atom shape function w_i is inherently spherical as it is the s-function part of $\rho_i(r)$. c_k^i is the expansion coefficient determined by minimizing the appropriate ISA functional and $\varepsilon_k^i(r)$ is the basis function (typical GTO). The shape function is hence

$$w_i(r) = \sum_{k \in s-func} c_k^i \varepsilon_{k,s}^i(r) \tag{53}$$

The s subscript in $\varepsilon_{k,s}^i$ denotes the s-character of the expansion. Spherical averaging becomes substantially easier when using basis expansions as the atomic functions. This methodology is comparable to the Gaussian Iterated Stockholder Atoms (GISA) approach,[254] which employs a series of s-function expansions to define the shape function but there are many differences between GISA and Misquitta's BS-ISA approach. These differences are vital to obtaining well-converged atomic density tails in the BS-ISA approach. The BS-ISA uses an additional auxiliary basis set for the fitting step of the atomic density tails, because the s-functions used for $\rho_i(r)$ are not flexible enough to describe the density tails well, and these need to be described well for the ISA solution to converge reliably.[233] The BS-ISA method also minimizes a different functional, employs larger basis sets, includes an exponential tail requirement and the ability to add off-atomic sites.[200, 233]

Using large diffuse basis sets in the density-fitting step results in well-defined density tails and accurate high order multipole moments, outperforming GDMA1 and GDMA2 methods in calculating the higher ranked multipole moments for formaldehyde.[233] Moreover, conventional ISA charges have been found to be very sensitive to conformation. Although in that study the multipole expansion was truncated at $Rank\ l = 0$, using only the atomic charges, ISA was deemed not robust enough to accommodate for changes in conformation of penta-alanine.[243] The BS-ISA charges on the other hand were. The form of BS-ISA used in this thesis uses *three* separate basis sets, a main basis, an auxiliary basis and an atom-auxiliary basis.[233] The BS-ISA method of obtaining distributed multipoles has been extended to generate the distributed polarizabilities and dispersion coefficients for both pyridine (**Chapter 3**) and trinitrobenzene (**Chapter 5**).

2.6 The induction model from atomic polarizabilities

A molecule in a crystal experiences an external potential $V(R)$ due to the potentials of its neighbours. This external potential has an associated electric field $F = -\partial V/\partial R$. If molecule M is in an external electric field F, the electrons experience opposite forces to the nuclei and the electron distribution of the molecule distorts. The results in an induced dipole moment $\Delta\mu$. The induced moment can be expressed in terms of the applied field and the polarizability tensor α from Equation 43 as

$$\Delta\mu = \alpha F$$

2.6.1 The polarization energy

Semi-empirically using Regularized SAPT(DFT),[255, 256] the 2nd order induction energy $E_{IND}^{(2)}$ (Equation 24) can also be partitioned into two components, contributions associated with delocalization of charge between molecules (**2.6.2**), the charge-delocalization energy $E_{CD}^{(2)}$, and those not associated with charge delocalization, the polarization energy $E_{POL}^{(2)}$. The polarization contribution here is the same interaction as the polarization in inorganic materials, however, in inorganic crystals the polarization energy is dominated by ion induced dipole-ion interactions which are rarely observed in organic molecular crystals. The polarization energy in organic crystals like pyridine is dominated by induced dipole moments arising from much weaker fields around the molecules. Note that the term polarization is sometimes used to describe changes of charge distribution within the isolated molecule caused by electronegativity differences, but not in this thesis.

The strength of the polarization contribution is dependent on the molecular charge density's ability to distort under the external field. The change in the distribution of charge density within a molecule at a point **r**, due to an external field caused by other molecules at **r'**, is known as the charge-density susceptibility $\alpha(\mathbf{r},\mathbf{r'})$ (Equation 43). Moreover, the response to the external field can be non-uniform across the molecule creating a field gradient. Distributed multipolar polarizabilities tensors are needed to describe this effect, as the field polarizes the atoms in the molecule, and its charge distribution is distorted inducing a multipole moment.[61, 164] This effect is non-additive,[164] and consequently, the polarization term is a truly many-body term. It is the only many-body term in the non-empirical force-fields created in this thesis.

2.6.1.1 The atomic polarizabilities model

The total and distributed frequency dependent polarizabilities are calculated using a constrained density fitting algorithm (like the multipole moments). The ISA-Pol method[200] within CamCASP, which follows on from the ISA-DMA method of deriving the multipole moments, can be utilized to derive the non-pairwise additive distributed polarizabilities models. This method is flexible and can be used to obtain an accurate polarizability model using constraints imposed by the user. The flexibility comes in the form of the ability to break down the models into different ranks, such as, an $L1$ (anisotropic dipole-dipole) polarization model, an $L2$ (anisotropic dipole-quadrupole & quadrupole-quadrupole) polarization model or a fully isotropic description of $L1$. Fully anisotropic or isotropic models up to $Rank\ l = 4$ can be derived and truncated using the ISA-Pol method.

The distributed polarizabilities from CamCASP also include non-local contributions. ORIENT can be used to transform these contributions away via its localization module, if the recipient molecular mechanics code can only take localized polarizabilities (which is the case with DMACRYS). Furthermore, point-to-point polarizabilities, responses to frequency-dependent point-charge perturbation can be calculated within CamCASP. We do not use these within this thesis as their added complexity makes them less accessible to available codes like DMACRYS.

2.6.1.2 Rotation of the atomic polarizabilities

CamCASP computes the atomic polarizabilities in the local axis (***Figure 6***), however, DMACRYS is only capable of using atomic polarizabilities in the molecule-fixed axis. The ISA-Pol localized atomic polarizabilities must be appropriately transformed and transferred for use in the lattice energy minimization calculations. The method of obtaining the localized atom polarizabilities has been detailed in a previous study.[200] The anisotropic polarizabilities are first computed and localized using the LW method[257] and weighting scheme number 3 (**2.9.5.2**) with an arbitrary weighting coefficient of 10^{-3}. The final weighted static polarizabilities must then be rotated into the molecule-fixed axis using ORIENT and re-formatted to be used as a DMACRYS input. The rotated DMACRYS polarizability input files can be found in **Appendix 3.A** and the **Appendix CD**.

2.6.1.3 Minimizing with Polarization

Computing the polarization energy (Equation 25) and the forces due to polarization contributions is an iterative process and generally very expensive especially for many-body systems. Currently, DMACRYS is capable of calculating the analytical first derivatives for all energy terms except polarization, which needs to be converged by iteration to a tolerance (δ_{pol}) of around $10^{-8}\ kJ\ mol^{-1}$. DMACRYS also has some analytical second derivatives, but not all the cross terms for the all energy terms, like the polarization. Normally, minimizations within DMACRYS do a quasi-Newton Raphson optimization, building up an approximate Hessian that is used to detect whether a valid minimisation (zero strain matrix, so no forces or torques) corresponds to a true lattice energy minimum, and then estimates the elastic tensor. When the polarization is explicitly included in the force-field, the numerical first derivatives of the polarization energy are included in the forces. Since the numerical derivatives of an iterated quantity are expensive to calculate, a structure should be first minimized without polarization and then, ideally, minimized again using this structure as a starting point. The induced moments must also be fixed along the same search direction in order for the energy to converge well. Consequently, in all but two (**Chapter 3**) lattice energy minimization calculations in this thesis we do not calculate the forces due to polarization.

2.6.1.4 The Polarizable Continuum Model

Is it reasonable to approximate the effects of polarization using the polarizable continuum model? As the polarization term is derived from the charge distribution it is most appropriate to have this term absorbed into the electrostatic model. The long-range effects of polarization can be absorbed into the electrostatic model by recalculating the multipole moments in a dielectric polarizable continuum solvation model (PCM), where the electric field due to neighbouring molecules is approximated as a polarizable continuum. An enhanced version of the COSMO Solvation model aka the "Conductor-like Screening Model" in NWCHEM[258, 259] was used to calculate the Restricted Kohn-Sham DFT (DFT) energies of the isolated TNB molecule under a specific dielectric by determining the solvent reaction field self-consistently with the solute charge distribution (**6.2.2**).

2.6.2 Charge-delocalization

The charge-delocalization term is a stabilizing energy that describes the delocalization of charge density between two molecules in close proximity. While the charge-delocalization effect occurs at the ground-state, for two interacting monomers it can be described as partly localized excitations on the neighbouring molecule, hence it is a 2nd order term. The charge-delocalization energy has been previously described as the difference between the 2nd order SAPT(DFT) induction energy ($E_{IND}^{(2)}$) calculated in a dimer centred (DC) basis and a monomer centred (MC) basis.

$$E_{CD}^{(2)} = E_{IND}^{(2)}[DC] - E_{IND}^{(2)}[MC]$$

55

This presumes that charge-delocalization effects are not allowed using MC basis types as the MC basis is smaller, incomplete and centred on each individual monomer with no sharing of charge density, whereas the larger DC basis functions are diffused between the two monomers (**2.2.4**).[264] This treatment, however, has its pitfalls as the charge-delocalization term here is clearly basis dependent. As the MC basis becomes larger and more complete $E_{IND}^{(2)}[MC]$ converges towards $E_{IND}^{(2)}[DC]$; reducing the apparent $E_{CD}^{(2)}$. Thus, the symmetry adapted perturbation theory methods SAPT and SAPT(DFT) face major roadblocks due to the basis-set dependence of the charge delocalisation term at small R. They do not describe a true 2nd order polarization energy but instead the total 2nd order induction energy $E_{IND}^{(2)}$, which is the SAPT(DFT) polarization and charge-delocalization energy combined.[255] Nonetheless, it has been suggested that SAPT(DFT) can be semi-empirically regularized in order to obtain the components of $E_{IND}^{(2)}$.[255] Under this treatment the 2nd order polarization energy is the regularized second-order induction energy, which can then be used to damp the distributed polarization model to best mimic the regularized SAPT(DFT) induction energy $E_{IND}^{(2)}(Reg)$.[38, 167] Subtracting $E_{IND}^{(2)}(Reg)$ from the 2nd order induction energy would also give the charge-delocalization contributions to the induction energy $E_{CD}^{(2)}$

$$E_{POL}^{(2)} = E_{IND}^{(2)}(Reg)$$

$$E_{CD}^{(2)} = E_{IND}^{(2)} - E_{IND}^{(2)}(Reg)$$

56

2.6.3 Estimating the higher-order induction terms

In SAPT(DFT) the complete total two-body interaction energy (including all higher up to infinite order perturbation terms) is

$$E_{int}^{(1-\infty)} = E_{elst}^{(1)} + E_{exch}^{(1)} + E_{IND}^{(2)} + E_{DISP}^{(2)} + \delta_{int}^{HF}$$

57

The polarization of the system due to an electric field continues to change the multipole moments of each molecule and the induced moments further polarize neighbours, leading to higher-order energies contributions. Thus, higher-order contributions are important and make up the remainder of the total interaction energy. Typically, they are too large to ignore, especially for highly polar/polarizable molecules. In **Chapter 3** the total two-body interaction energy is employed, however, in **Chapter 5**, the perturbation expansion is truncated to 2nd order.

In the previous non-empirical potential for pyridine,[38, 167] the 3rd order and higher induction energies were absorbed into a term known as δ_{int}^{HF}; the Hartree-Fock energy of the system that contains all induction effects[38] but no correlation corrections[265, 266] and is defined as

$$\delta_{int}^{HF} = E^{HF} - E_{elst}^{HF} - E_{exch-rep}^{HF} - E_{index}^{HF}$$

58

Equation 58 is the Hartree-Fock interaction energy with the correlation corrections (E_{index}^{HF}), first order electrostatic and exchange-repulsion Hartree-Fock energies subtracted.[164] The remaining contributions after these terms have been deducted from E^{HF} are those involving the change in charge distribution of each molecule due to the electric field from neighbouring bodies and higher order charge-delocalization terms, i.e. the remainder of the induction energy. Subsequently, under the SAPT(DFT) partitioning of intermolecular interaction energies, the δ_{int}^{HF} correction energy contains all higher order two-body interactions, like the effect of higher order induction terms $E_{IND}^{(3-\infty)}$.[164] Further detail on its implementation and its effect on the potential energy surface of gas-phase dimers and crystal structures is detailed in **Chapters 3** and **5**. The explicit calculation of the polarisation contribution to the lattice energy requires iteration so all symmetry related molecules have the same charge distribution. This is equivalent to including the higher order terms, i.e. the induced moments on one molecule induce further changes in the induced moments of the nearby molecules, but this is approximate as it uses distributed multipoles and polarizabilities within a crystalline, many-body environment.

2.7 The dispersion model

The dispersion energy is an approximately pairwise-additive, attractive energy arising from *correlated* fluctuations in the electron distribution of each molecule that create instantaneous multipole moments within a molecule. These multipole moments induce other multipole moments within neighbouring molecules, resulting in correlated fluctuation of molecular charge densities. These correlations favour lower energy configurations and lower the overall energy of the system.[164] We say it is approximately pairwise-additive as many-body effects, like the 3-body dispersion can be important.[164, 167] It is a non-classical phenomenon and the dispersion energy is typically an enormous contribution to the intermolecular energy and can be dominant in organic crystals depending on whether the molecule is capable of strong electrostatic interactions. After all, dispersion is the van der Waals interaction that allows spherical neutral molecules to form condensed states.[267] Accordingly, an accurate and reliable model of this component is crucial. The long-range perturbation theory approach (**2.3.1**) used to calculate the intermolecular energy in a force-field, reduces the dispersion coefficients to products of the individual-molecule properties, which allows the dispersion energy coefficients to be calculated just once for a rigid molecule.[164] Quantum electrodynamics has been utilized by Casimir and Polder[268] to better describe the dispersion energy, which was further developed and simplified[269] allowing the dispersion energy to be derived from the polarizability components of two interacting bodies at imaginary frequency.[164] The energy is accurately calculated as a function of frequency-dependent dipolar polarizability from ab initio calculations, but can also be derived from experimental data. Substituting $\mathcal{H}'$ from Equation 42 into Equation 43 gives us the multipole-expanded form of the dispersion energy. If we consider only up to the dipole-dipole term in $\mathcal{H}'$

$$E_{disp}^{(2)} = -T_{\alpha\beta}T_{\gamma\delta} \sum_{m\neq 0, n\neq 0} \frac{\langle \Psi_M^0 | \hat{\mu}_\alpha^M | \Psi_M^m \rangle \langle \Psi_M^m | \hat{\mu}_\gamma^M | \Psi_M^0 \rangle \langle \Psi_N^0 | \hat{\mu}_\beta^N | \Psi_N^n \rangle \langle \Psi_N^n | \hat{\mu}_\delta^N | \Psi_N^0 \rangle}{E_m^M - E_0^M + E_n^N - E_0^N}$$

Using the full spherical-tensor notation of the intermolecular interaction operator $\mathcal{H}'$ and summing over repeated spherical-tensor suffixes t, t', u and u' gives the multipolar 2nd order dispersion energy as

$$E^{(2)}_{disp}[MP] = -\sum_{m\neq 0, n\neq 0} \frac{\langle 00|\hat{Q}^M_t T^{MN}_{tu} \hat{Q}^N_u|mn\rangle\langle mn|\hat{Q}^M_{t'} T^{MN}_{t'u'} \hat{Q}^N_{u'}|00\rangle}{E^M_m - E^M_0 + E^N_n - E^N_0}$$

60

The matrix elements can be factorised components of M and N, however, the denominator cannot be factorised, making it awkward to handle. The following identity can be used to obtain an exact formulation of the dipole-dipole term in the dispersion energy

$$\frac{1}{A+B} = \frac{2}{\pi}\int_0^\infty \frac{AB}{(A^2+v^2)(B^2+v^2)} dv$$

61

This is valid for positive A and B, and can be applied to the energies in the denominator of Equation 59, where $\hbar A = E^M_m - E^M_0 = \hbar\omega^M_m$. Expanding Equation 60, and still summing over repeated spherical-tensor suffixes t, t', u and u', allows one to identify the components of the polarizability at imaginary frequency (iv) to obtain a general form of the dispersion energy, that will yield $R^{-(2n+6)}$ contributions to the dispersion energy. Giving

$$E^{(2)}_{disp} = -\frac{2\hbar}{\pi} T^{MN}_{tu} T^{MN}_{t'u'} \int_0^\infty {\sum_m}' \frac{\langle\Psi^0_M|\hat{\mu}^M_\alpha|\Psi^m_M\rangle\langle\Psi^m_M|\hat{\mu}^M_\gamma|\Psi^0_M\rangle\omega^M_m}{\hbar((\omega^M_m)^2+v^2)} \times {\sum_m}' \frac{\omega^N_n\langle\Psi^0_N|\hat{\mu}^N_\alpha|\Psi^n_N\rangle\langle\Psi^n_N|\hat{\mu}^N_\gamma|\Psi^0_N\rangle}{\hbar((\omega^N_n)^2+v^2)} \mathrm{d}v$$

62

These contributions are partitioned and implemented in the potential as atom-type specific dispersion coefficients C^{2n+6}, which are only valid asymptotically.

$$E^{(2)}_{disp} = -\frac{\hbar}{2\pi} T^{MN}_{tu} T^{MN}_{t'u'} \int_0^\infty \alpha^M_{tt'}(\mathrm{i}v)\alpha^M_{uu'}(\mathrm{i}v)\mathrm{d}v$$

63

The orientation dependence of the dispersion is defined in the tensors T^{MN}_{tu} and $T^{MN}_{t'u'}$. This expression is exact in the limit of large R. Furthermore, R^{2n+7} contributions are zero as they cancel out due to symmetry for centrosymmetric molecules and atoms or are very small, therefore, R^{2n+7} contributions are not included in our distributed dispersion model.

2.7.1 Empirical dispersion models

In empirical force-fields like FIT[68] or WILL01,[69, 70] the dispersion is fitted together with the repulsion in an isotropic atom-atom Buckingham exp-6 potential, which is summed for all atoms i and k of type ι and κ in molecules M and N.

$$V_{rep-disp}^{MN}(R_{ik}) = A^{\iota\kappa} \sum_{i \in M,\ k \in N} \exp[-\alpha^{\iota\kappa} R_{ik}] - \frac{C_6^{\iota\kappa}}{R_{ik}^6}$$

64

The $C_6^{\iota\kappa}$ dispersion coefficient, $A^{\iota\kappa}$ and $\alpha^{\iota\kappa}$ are empirically fitted to experimental data and are atom specific, and in the case of the Williams potential, specific to atomic-connectivity. In the WILL01 potential are different parameters for different nitrogen bonds for example. In empirical potentials combining rules are also used to calculate the parameters for heterogeneous pairs,[270]

$$A^{\iota\kappa} = \sqrt{A^{\iota\iota} + A^{\kappa\kappa}}$$

$$\alpha^{\iota\kappa} = \sqrt{\alpha^{\iota\iota} + \alpha^{\kappa\kappa}}$$

$$C_6^{\iota\kappa} = \sqrt{C_6^{\iota\iota} + C_6^{\kappa\kappa}}$$

65

However, the non-empirical force-fields in this study do not use combining rules. Parameters are derived (in the case of the dispersion) or fitted for both heterogeneous and homogeneous pairs. This is done using the dispersion module within CamCASP when deriving the isotropic pair-wise dispersion model.

2.7.2 Non-empirical dispersion models

The Iterated Stockholder Atoms approach[200] can be utilized to derive the pairwise dispersion model as it leads directly from the ISA-Pol method by integration over the distributed polarizabilities. The dispersion contribution in our force-field is expressed accurately in terms of a power series in $R^{-(2n+6)}$, damped with single parameter Tang-Toennies damping functions (Equation 67).

$$V_{disp}^{MN}(R_{ik}) = \sum_{ik} -f_6(\beta_{MN} R_{ik}) \frac{C_6^{ik}}{R_{ik}^6} - f_8(\beta_{MN} R_{ik}) \frac{C_8^{ik}}{R_{ik}^8} - f_{10}(\beta_{MN} R_{ik}) \frac{C_{10}^{ik}}{R_{ik}^{10}}$$

66

where M and N are two interacting molecules and C_{2n+6} are the computed dispersion coefficients. The higher order C_8 and C_{10} terms are included to better capture van der Waals interactions at the short-range as R^{-8} & R^{-10} become important at small R. Atoms are assumed to be spherical, which is an approximation, but necessary as anisotropic dispersion models cannot be used in DMACRYS. Furthermore, the dispersion is kept isotropic for the sake of simplicity as the issues and complexities associated with an anisotropic model currently outweighs its benefits.[223]

2.8 Damping

As previously stated, the long-range distributed multipole expansions ignore the effects of molecular overlap and must be damped at short-range in order to prevent them from blowing up and behaving unphysically. Suitable damping functions must recover the long-range formula at large separations and supress singularity at small separations. In this thesis the Tang-Toennies damping function[271] is used. The Tang-Toennies damping function uses a single parameter (β_{MN}) and tends to 0 as $R \rightarrow 0$, and 1 as $R \rightarrow \infty$. The Tang-Toennies damping function is given as

$$f_n(\beta_{MN}R_{ik}) = 1 - \exp(-\beta_{MN}R_{ik})\left(\sum_{m=0}^{n}\frac{(\beta_{MN}R_{ik})^m}{m!}\right)$$

67

where β_{MN} is the damping parameter for a molecular pair M and N. Using a single damping parameter is not ideal. Not only are our terms dependent on the atom pair but also orientation, therefore, atom specific or even orientation specific damping functions should be used. However, the single parameter Tang-Toennies is the only functional form accepted by DMACRYS and also allows the force-field to be more transferable between ORIENT and CAMCASP.

The electrostatic expansion is not damped as this is not necessary, when the penetration energy is taken into account,[272] however, the dispersion must be damped as it mathematically diverges relatively quickly. The polarization energy behaves in a similar fashion and is also damped to suppress singularity and avoid errors when the wave-functions of two point polarizabilities overlap, known as the polarization catastrophe.[165, 227] The final dispersion (β_{disp}), and polarization damping parameters (β_{pol}), are obtained by sampling important dimers on a molecule's potential energy surface and calculating their second-order SAPT(DFT) dispersion ($E^{(2)}_{DISP}$) and polarization ($E_{POL}^{(2)}$) energies, as detailed for TNB in **Chapter 5**. A first guess of β_{disp} can be initially derived from the ionization potential of the two monomers M and N.[273]

$$\sqrt{2I_M} + \sqrt{2I_N} = \beta_{disp}$$

68

This method is used to obtain the dispersion damping parameter for the initial model derived from the first-order SAPT(DFT) interaction energies, which is used for the higher order calculations.

2.9 Fitting the short-range contribution

2.9.1 Partitioning the short-range energy

Interaction effects that occur at short distances, like the dominant exchange-repulsion term, charge-delocalization, electron tunnelling, electron exchange and electrostatic penetration, cannot be directly derived from the molecular charge density using our current methodology. Instead corrections are made to the potential through fitting. The remaining effects not described in the long-range part of our potential are included in the short-range component, where the most dominant effect is the energy due to exchange-repulsion. The exchange-repulsion describes the repulsive energy due to the anti-symmetrisation of the electron exchanges between the wave-functions (Ψ) of two overlapping molecules (**2.3.1**). This interaction cannot be described by a simple product of their isolated-molecule wave-functions, thus we do not derive this portion of the potential but instead fit the total short-range energy (E_{SR}) in two stages using CamCASP,[38] and split the energy into the dominant 1st order short range energy ($E_{SR}^{(1)}$) and the 2nd order short-range energy ($E_{SR}^{(2)}$).

$$E_{SR} = E_{SR}^{(1)} + E_{SR}^{(2)}$$

$$E_{SR}^{(2)} = (E_{IND}^{(2)} - V_{pol}^{(2-\infty)}[DM] - E_{CD}^{(2)}) + (E_{DISP}^{(2)} - V_{disp}^{(2)}[DM])$$

69

The first step in developing a complete description of the short-range is to develop a short-range repulsion model from 1st order dimer interaction energies. The 1st order SAPT(DFT) interaction energy ($E_{int}^{(1)}[SAPT(DFT)]$) is composed of the 1st order electrostatic ($E_{elst}^{(1)}$) and exchange-repulsion ($E_{exch}^{(1)}$) energies and can be given as

$$E_{int}^{(1)}[SAPT(DFT)] = E_{elst}^{(1)} + E_{exch}^{(1)}$$

70

The total electrostatic energy ($E_{elst}^{(1)}$) can be broken down into two contributions, a long-range multipolar electrostatic energy ($V_{elst}^{(1)}[DM]$), which we calculate by the distributed multipole expansion. $V_{elst}^{(1)}[DM]$ is known to differ from the non-expanded electrostatic energy ($E_{elst}^{(1)}$) by an energy defined as the electrostatic penetration energy ($E_{pen}^{(1)}$).[38, 233]

$$E_{elst}^{(1)} = V_{elst}^{(1)}[DM] + E_{pen}^{(1)}$$

71

The 1st order short-range energy is composed of the 1st order exchange energy $\left(E_{exch}^{(1)}\right)$ and the 1st order penetration energy $\left(E_{pen}^{(1)}\right)$, a component of $E_{elst}^{(1)}$,

$$E_{SR}^{(1)} = E_{pen}^{(1)} + E_{exch}^{(1)}$$

72

$E_{pen}^{(1)}$ is the electrostatic energy due to the penetration of ψ at shorter distances when the molecular charge distributions of two molecules overlap. The penetration energy can be absorbed into our short-range exponential term as with other small terms, like the charge-delocalization and exchange-dispersion energies by fitting. Subsequently, the first-order short-range energy can be given as

$$\therefore E_{SR}^{(1)} = E_{elst}^{(1)} - V_{elst}^{(1)}[DM] + E_{exch}^{(1)}$$

73

We do not analytically derive the short-range potential energy but fit the calculated short-range energy E_{SR} from Equation 69 by fitting to parameters within our short-range exponential V_{SR}^{MN} and approximating induction and dispersion damping parameters using 2nd order interaction energy calculations.

2.9.2 The short-range exponential

The energy at short-range can be represented by an exponential function in non-empirical force-fields. For the atom-atom interactions of atoms i of type ι with atoms k of type κ in molecules M and N, the short-range potential energy V_{SR}^{MN} can be defined as:

$$V_{SR}^{MN} = G \sum_{i \in M,\ k \in N} \exp[-\alpha^{\iota\kappa}\left(R_{ik} - \rho^{\iota\kappa}(\Omega_{ik})\right)]$$

74

Where G is a constant to define the energy scale, $\rho^{\iota\kappa}(\Omega_{ik})$ is the shape function of the pair of atom types ι and κ, an expansion in the angular co-ordinate that can be approximated as $\rho^{\iota\kappa}(\Omega_{ik}) \cong \rho^{\iota}(\phi_i, \psi_i) + \rho^{\kappa}(\phi_k, \psi_k)$. Ω_{ik} is the relative orientation of two sites and $\alpha^{\iota\kappa}$ the hardness parameter. For example using the local axis described in ***Figure 6***, for the anisotropic terms 00, 10, 20 on sites i and k would be described as

$$\rho^{\iota\kappa}(\Omega_{ik}) = \rho_0^{\iota\kappa} + \rho_1^{\iota}(\vec{z}_i \cdot \vec{R}_{ik}) + \rho_1^{\kappa}(\vec{z}_k \cdot \vec{R}_{ik}) + \rho_2^{\iota}\left(3[\vec{z}_i \cdot \vec{R}_{ik}]^2 - 1\right)/2 \ + \rho_2^{\kappa}\left(3[\vec{z}_k \cdot \vec{R}_{ik}]^2 - 1\right)/2$$

75

$\vec{z}_i$ is a unit vector in the z direction (along the bond on our local axis system) on atom i, while $\vec{R}_{ik}$ is a unit vector in the inter-site direction between sites i and k. This assumes that the anisotropy is cylindrically symmetrical about this inter-site vector. The exponential terms $\alpha^{\iota\kappa}$ and $\rho^{\iota\kappa}(\Omega^{\iota\kappa})$ can be obtained by fitting to penetration energy ($E_{pen}^{(1)}$) and the exchange energy ($E_{exch}^{(1)}$).

2.9.3 The overlap model: fitting the first-order short-range energies to terms in the exponential

As it is non-trivial to directly fit the sum of exponentials, let alone sufficient anisotropy coefficients, the density overlap model is used to split the energy into atom-atom contributions. The molecular density of a molecule M can be partitioned into a sum of atomic densities for sites i in M

$$\rho_M(r) = \sum_i \rho_i(r)$$

76

The same is also true for sites k in molecule N. The density overlap integral for molecular densities of M and N at a relative position and orientation $\boldsymbol{R}$ can be described as

$$S^{MN}(\boldsymbol{r}) = \int \rho_M(r_1)\rho_N(r_2)\, dr$$

77

If we partition the density overlap integral between the atomic sites we can define it as

$$S^{MN}(\boldsymbol{r}) = \sum_{i\in M}\sum_{k\in N}\int \rho_i(r)\rho_k(r)\, dr = \sum_{i\in M, k\in N} S^{ik}(\boldsymbol{r})$$

78

$S^{ik}(\boldsymbol{r})$ is the total density overlap $S^{MN}(\boldsymbol{r})$ partitioned into a sum of atom-atom interaction terms between i and k using a density-fitting procedure and thus depends on the type of distribution scheme used. As with the multipole moments, the ISA partitioning was used in the overlap model. Hence, one can determine the distributed 1st order short range energy $E_{SR}^{(1)}(ik)$ using the sum of the product of the atom-atom density overlap integrals and their density overlap parameter (K^{ik})

$$E_{SR}^{(1)} = \sum_{i\in M, k\in N} K^{ik} \int \rho_i(r)\rho_k(r)\, dr = \sum_{i\in M, k\in N} K^{ik} S^{ik}(\boldsymbol{R}) = \sum_{i\in M, k\in N} E_{SR}^{(1)}(ik)$$

79

2.9.4 Simplifying terms using symmetry

2.9.4.1 Atom types

The number of terms in the potential can be significantly reduced using the symmetry of the molecule, and if the molecules within the dimer are identical and symmetrical, the total number of atom types can be further reduced. The overlap matrix can be symmetrized; constructed to use all the overlaps of one type between atom types ι and κ of each site i and k in each molecule

$$S_p^{\iota\kappa} = \sum_{i=\iota} \sum_{k=\kappa} S_p^{ik}(NoSymm)$$

80

$S_p^{\iota\kappa}$ is the total density overlap matrix for atom-types ι and κ, while $S_p^{ik}(NoSymm)$ is the overlap integral matrix for atomic sites i of atom-type ι in molecule M with sites k of atom-type κ in molecule N. Furthermore, if the dimer is comprised of identical molecules (M ... M), an additional symmetrization of the overlap matrix $S_P^{\iota\kappa}$ can be performed to symmetrize between the atoms ι molecule M and κ in molecule N.

$$S_P^{\iota\kappa}(\text{symm}) = \frac{1}{2}(S_P^{\iota\kappa} + S_P^{\kappa\iota})$$

81

This provides a numerical dataset of overlap values for each pair of atomic types, which depends on the relative separation of the atoms and the orientation of the two atomic local axis relative to the intermolecular interatomic vector.

2.9.4.2 Anisotropic terms

The anisotropy terms can also be reduced using the symmetry of the molecule, we will use the local axis (blue) defined in ***Figure 6*** and the molecule trinitrobenzene (TNB) as an example. As the molecule is planar the y terms are equal and opposite above and below the plane of the ring, and y-dependent terms like $11s$ and $21s$ should be zero by symmetry. In addition due to TNB's D_{3h} symmetry the terms containing anisotropy in the x axis are also equal and opposite for all atoms except oxygen, which has different chemical environments on either side. The important anisotropy terms for atomic site type ι in the atom-atom pair $\iota - \kappa$ in TNB is shown in ***Table 5***.

K_1	K_2	L_1	L_2	J	Function	Term	Atoms
0	0	1	0	1	$\vec{z} \cdot \vec{R}$	10	All
1	0	1	0	1	$\vec{x} \cdot \vec{R}$	$11c$	$Oxygen$
-1	0	1	0	1	$\vec{y} \cdot \vec{R}$	$11s$	$None$
0	0	2	0	2	$(3(\vec{z} \cdot \vec{R})^2 - 1)/2$	20	All
1	0	2	0	2	$\sqrt{3}(\vec{z} \cdot \vec{R})(\vec{x} \cdot \vec{R})$	$21c$	$Oxygen$
-1	0	2	0	2	$\sqrt{3}(\vec{z} \cdot \vec{R})(\vec{y} \cdot \vec{R})$	$21s$	$None$
2	0	2	0	2	$\sqrt{3}\left((\vec{x} \cdot \vec{R})^2 - (\vec{y} \cdot \vec{R})^2\right)/2$	$22c$	All
-2	0	2	0	2	$\sqrt{3}(\vec{x} \cdot \vec{R})(\vec{y} \cdot \vec{R})$	$22s$	$None$

Table 5: *Allowed anisotropy terms in DMACRYS*[59]*, and the atoms in TNB that require them by symmetry. Here the terms are for the site ι in an atom-atom pair $\iota - \kappa$ with the directionality $\iota \rightarrow \kappa$. The anisotropic term $\rho^{\iota\kappa}(\Omega_{ik})$ is defined by an S function that can be defined by the terms K_1, K_2, L_1, L_2, J.For example, for an atom-atom pair $\iota - \kappa$, the 11c term on type ι can be defined as $J, K_1, L_1 = 1, K_2, L_2 = 0$ while the 11c term on type κ can be defined as $J, K_2, L_2 = 1, K_1, L_1 = 0$.*

2.9.5 Weighting and Fitting method

Once we have defined all the distinct atom-atom interaction types, (there are 15 in the TNB dimer), the short-range energy can be split into atom-atom contributions using the 15 $S_p^{\iota\kappa}(Symm)$ datasets. Each of the exponential terms can be fit to their atom-atom contributions. This method is robust as it allows one to truncate or elaborate the functional form of our short-range potential. Depending on the axis-system used by various molecular modelling codes the allowed anisotropy terms in **Table 5** and the functional form of our short-range potential could change.

2.9.5.1 Estimating parameters for each atom pair from overlaps

Dimer energies can either be fit to an isotropic or an anisotropic (including directional terms like $10, 11c, 11s$ and above) short-range exponential. The calculated 1st order short-range energies $E_{SR}^{(1)}$ are first fitted to the density overlap, firstly using a single overlap proportionality parameter K_0. For atomic types ι and κ in molecules M and N with overlap proportionality parameters of $K^{\iota\kappa}$ ($K^{\iota\kappa} = K_0$, initially) and density overlap matrices of $S^{\iota\kappa}$

$$E_{SR}^{(1)} = e \equiv E_{exch}^{(1)} + E_{pen}^{(1)} \Rightarrow K_0 S \Rightarrow \sum_{\iota\in M,\kappa\in N} K^{\iota\kappa} S^{\iota\kappa}(\boldsymbol{R})$$

82

i.e. K_0 is first an initial estimate used as a starting point for fitting $K^{\iota\kappa}$. Distributed density overlap integral calculations are carried out on each dimer configuration in order to obtain the overlap integrals S^{ik} for each atom in the molecule. The distributed density-overlap parameters (K^{ik}) are then obtained using the ISA-based partitioning of the density. The only difference between the ISA-based partitioning algorithm used to obtain the multipole moments and the algorithm used for the overlap model is that the charge density is partitioned in real-space as opposed to basis space for the overlap model, as it is more accurate. CamCASP initially calculates the total overlap model for each $E_{SR}^{(1)}$ with just one overlap parameter, i.e. K^{ik} is the same as K_0 for all atom pairs. K_0 changes to minimize the weighted variance χ^2

$$\chi^2 = \sum_p w(e_p)\left(\frac{K_0 S_p}{e_p} - 1\right)^2$$

83

$w(e_p)$ is a weighting of a given data point p that depends on the energy e_p and the appropriate weighting scheme used (**2.9.5.2**). Using only a single K for all atom pairs allows for a very rough initial fit, however, the resulting short-range terms have no physical meaning and are not representative of the landscape sampled by the 1st order SAPT(DFT) energy calculations. Hence, K is allowed to vary for each atom type. Firstly, preliminary potential parameters are generated using the initial density-overlap parameter K_0, and then the Equation 84 is minimized to obtain the distributed density-overlap parameters K^{ik}. A penalty function is used to slightly constrain K^{ik} from varying too significantly from K_0 (Equation 84). Some of these parameters can be ill-defined and negative, however, CamCASP re-calculates the distributed overlap model, using the K_0 as an anchor for the K^{ik} values. While it is not easy to tighten the constraints on K^{ik} for each pair, one can tighten the constraints of all K^{ik} parameters.

$$\frac{\chi^2}{\sum_p w(e_p)} + \sum_{ik} \lambda \left(K^{ik} - K_0\right)^2$$

84

The overlap model constraint parameter λ is included to constrain the fit and the weighted variance is now

$$\chi^2 = \sum_p w(e_p) \left(\frac{\sum_{i \in M, k \in N} K^{ik} S_p^{ik}}{e_p} - 1 \right)^2$$

85

The root mean square derivation of the weighted variance (χ^2) is then used to ascertain the quality of the fit. For each atom-atom interaction point (p) in a dimer configurations, the total density overlap values S_p^{ik} and fitted K^{ik} parameters are outputted.

2.9.5.2 Available weighting schemes

In ORIENT & CamCASP, there are multiple weighting schemes available to use for the overlap model and relaxation/fitting procedure. The choice of weighting scheme is crucial in the fitting of the potential as it ensures the most important structures are sampled correctly. Generally, the five main weighting schemes that can be used for weighting the short-range and total SAPT(DFT) interaction energies of the dimer configurations are:

1. $w(e_p) = 1$ – all energies are given an equal weighting.
2. $w(e_p) = \left(1 + (e_p/e_0)\right)^{-1}$ – equivalent to assuming the uncertainty in the calculated energy e_p is proportional to *magnitude* of the energy.
3. $w(e_p) = \left(1 + (e_p/e_0)^2\right)^{-1}$ – similar to the above scheme, the uncertainty is proportional to the magnitude of energy but with a heavier bias towards smaller energies.
4. $w(e_p) = (e_0/e_p) + 1$
5. $w(e_p) = \exp\left[-\alpha\left(\ln(e_p/e_0)\right)^2\right]$ – Gaussian-log weighting scheme.

where e_p is the calculated energy of data point p and e_0 is the chosen reference energy to weight against. In the force-field development process of non-empirical potentials for pyridine and trinitrobenzene, the Gaussian-log weighting[274] scheme is used in the fitting of the short-range parameters to the calculated $E_{SR}^{(1)}[SAPT(DFT)]$ energies, which are repulsive ($e_p > 0$). This weighting function is good for data that spans multiple orders of magnitude, and $\alpha = 1/\ln 10$ by default. However, the Gaussian-log function can only be used for the short-range energies which are positive as it is undefined when $e_p = 0$, i.e. when the total energies go through 0. In the 2nd stage of the fitting process where the models are relaxed to the second-order SAPT(DFT) intermolecular interaction energies (**5.6.1**), weighting scheme 3 is used, which allows one to include the most repulsive of energy without over-biasing the potential. The shape of the intermolecular potential in the region that is most important for simulations (***Figure 4***) means we must use different weighting functions for different parts of the fitting process.

2.10 Developing distributed intermolecular force-fields

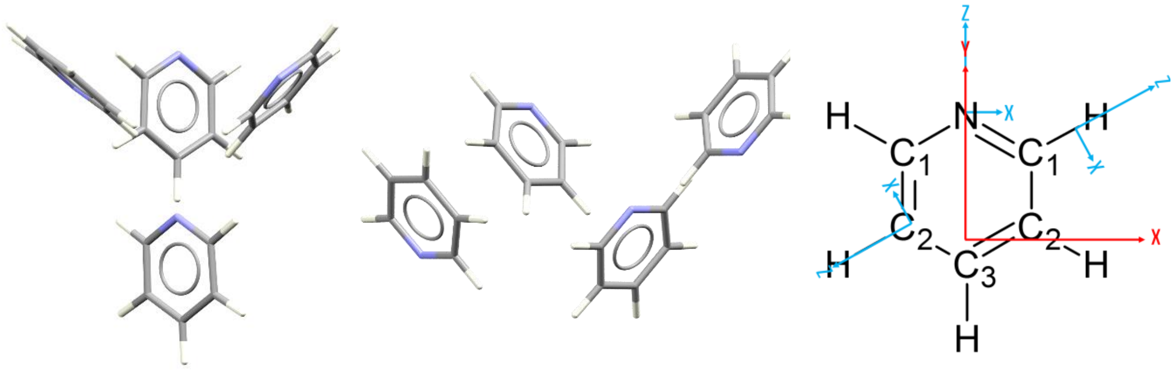
By carefully combining the theory of intermolecular forces with conventional and novel quantum chemical codes, one can build each component of the non-empirical atom-atom anisotropic intermolecular force-field from a solid foundation. In this chapter, we emphasise that the construction of these potentials is unique to each molecule and very involved. The potentials are state-of-the-art and the distributed intermolecular force-fields (DIFFs) in this study are the first two non-empirical anisotropic force-fields with an explicit polarization model ever used in CSP. The DIFF is extended from the gas-phase to the solid-state of the pharmaceutical building block pyridine in **Chapter 3,** and in **Chapter 5** the DIFF workflow and its development is detailed for the energetic building block trinitrobenzene (TNB). In **Chapter 6** the theory of intermolecular forces and the approximations discussed are truly tested in a CSP study of TNB.

3 From the gas-phase to the crystalline state: Pyridine[167]

Now that background has been given into the theory of intermolecular forces and force-field development, we shall examine its application. This chapter investigates the development of a non-empirical anisotropic atom-atom intermolecular potential for modelling pyridine's solid state, which is referred to as a Distributed Intermolecular Force-Field (DIFF). This DIFF is equivalent to the Model 3 previously developed by Misquitta.[38] This chapter further develops work initially established by Misquitta and Stone to study the pyridine gas-phase.[38] The theoretically based many-body pair potential is applied to model the intermolecular interactions in the pyridine crystalline-state. The potential uses distributed atomic multipoles, polarizabilities and dispersion coefficients derived from the molecular charge distribution, and an anisotropic atom-atom repulsion model fitted from SAPT(DFT) dimer interaction energy calculations to model the genuine potential energy surface of pyridine. It is established that these new models are capable of accurately modelling pyridine crystal structures and even capable of identifying an unreported high pressure phase of pyridine, unlike existing empirical force-fields.[167] The workflow developed here is completely novel and thus many adaptions to existing codes have been made to apply this work to the solid-state and obtain the results published.[167] This work on pyridine led to refinements in the development and applications of a non-empirical DIFF for trinitrobenzene (TNB)(**Chapter 5**).

3.1 Pyridine

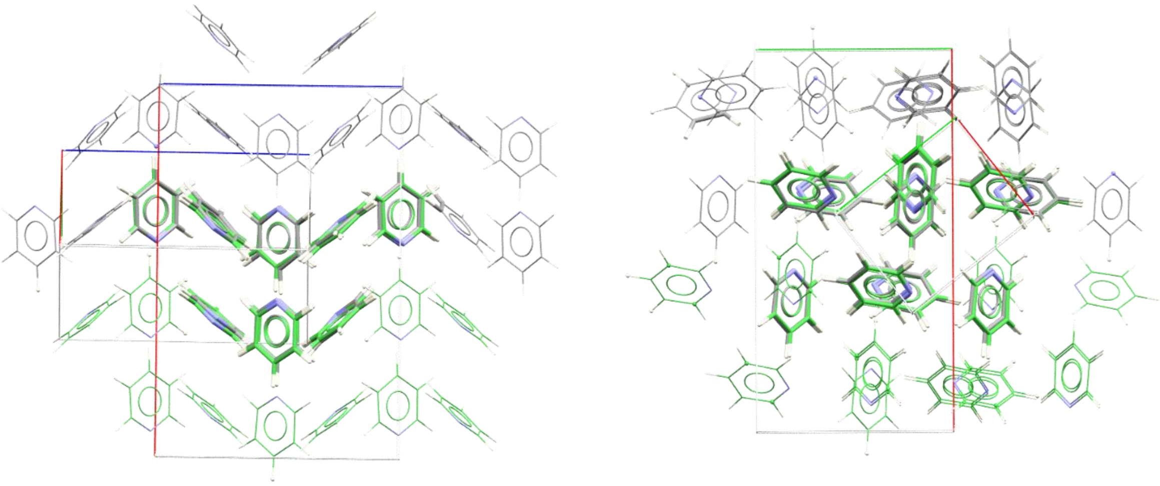
With a melting point of $231.6K$,[275, 276] pyridine is a liquid at ambient conditions. Pyridine (CSD refcode: PYRDNA), isoelectronic to benzene, is a relatively small molecule of 11 atoms, 42 electrons and no flexible functional groups. Consequently, its molecular structure changes very little between its experimentally observed crystalline structures,[186, 275-278] and thus, when modelling, one can assume pyridine to be a rigid molecule that adopts its isolated static molecular structure (***Figure 9***) in the crystalline state. Accordingly, the isolated molecule extracted from the PYRDN05[186] experimental structure is optimized using the Gaussian09 program,[235] the PBE0 functional[203-205] and the d-aug-cc-pVTZ Dunning basis.[214] The structure is kept rigid in all subsequent simulations and the axis frames described in ***Figure 9*** (c) are used throughout this study. Different axis definitions were used in the initial study of pyridine by Misquitta[38], but these were changed in order to allow the transfer of anisotropic terms between CamCASP[107]/ORIENT[108] and DMACRYS.[59]



(a) (b) (c)

Figure 9: *The experimentally observed crystal structures of pyridine, (a) Form I ($d5$-I PYRDNA04*[186]*) and (b) Form II (green $d5$-II PYRDN05*[186]*). (c) The molecule-fixed axis (red) and local axis (blue) definitions used for pyridine.*

An accurate potential is required to model the pyridine solid state as it has an observed dipole moment of ~2.2 Debye,[279] and is capable of forming weak $CH \dots N$ hydrogen bonds.[280] Therefore, the attractive dispersion, electrostatic and polarization forces are competitive with the short-range repulsive forces in determining the structure of the different phases, as has already been shown in the design of an intermolecular potential for simulating liquid pyridine.[281] The earliest observed crystal structure of pyridine is unusually complex,[275, 276] with 4 independent molecules in the asymmetric unit ($Z' = 4$) and 16 molecules in the unit cell (form I, $Pna2_1$) (***Figure 9***). After an early crystal structure prediction (CSP) study showed that simpler structures were thermodynamically competitive, a low-temperature form II ($P2_12_12_1$ $Z' = 1$) was crystallized from pentane for deuterated-pyridine ($d5$) though not for protonated-pyridine ($h5$).[277] The two polymorphs are rather similar, with a large common coordination cluster (***Figure 10***) but differ in the stacking so that a $CH \dots \pi$ interaction in form I becomes a $CH \dots N$ interaction in the low-temperature, high-pressure form II.[186] A later CSP study, using periodic DFT-D calculations, suggested that the lattice energy of form II is slightly more stable than form I by less than $0.1\ kJ\ mol^{-1}$. Moreover, the difference increases with pressure, and alternative structures are within $0.1\ kJ\ mol^{-1}$ of the observed structures.[282] The crystallization behaviour of pyridine is isotope-dependent, as well as depending on pressure and temperature.[186, 280] Additionally, towards the conclusion of this work, it was discovered that a third form ($d5$-III) had been crystallized at around $2GPa$ pressure,[278] and the challenge of identifying its structure was set, providing a test of how well the intermolecular force-fields model the repulsive wall.



(a) (b)

Figure 10: *Comparison of the two observed polymorphs of pyridine; form I in grey ($d5$-I PYRDNA04[186]) and form II in green ($d5$-II PYRDN05[186]). They show an optimum overlay of 11/15 molecules with an $RMSD_{11} = 0.149$Å. (a) shows the overlay of the lattice cells of the two polymorphs including molecules that do not overlay. It is viewed along the* ***b*** *axis with the* ***a*** *and* ***c*** *axes shown in red and blue respectively. (b) shows the same overlay of forms I & II as (a) but along the* ***c*** *axis with the* ***a*** *and* ***b*** *axes shown in red and green respectively.*

3.2 Methodology

3.2.1 Transferring the analytical models

The most accurate version of the models developed by Misquitta[38] is used in this study. Misquitta derived this model from theoretical calculations on the monomer and dimer. The main focus of the following methodology, is the aspects of the model that have been modified for use in DMACRYS.[59] Furthermore, the necessary sanity checks required to ensure the potentials are equivalent and reproducible are described. The pyridine DIFF functional form can be given as

$$U_{M,N}(R_{MN},\Omega_{MN}) =$$

$$\sum_{i\in M,k\in N} U^{ik}_{DIFF}(R_{ik},\Omega_{ik}) = G\exp\left[-\alpha^{\iota\kappa}\left(R_{ik}-\rho^{\iota\kappa}(\Omega_{ik})\right)\right] + Q^i_t T^{ik}_{tu} Q^k_u + \sum_{n=0}^{2} f_{2n+6}\left(\beta_{disp},R_{ik}\right)\frac{C^{\iota\kappa}_{2n+6}}{R^{2n+6}_{ik}}$$

$$+f_n\left(\beta_{pol},R_{ik}\right)(\Delta Q^i_t T^{ik}_{tu} Q^k_u + \Delta Q^k_t T^{ki}_{tu} Q^i_u)/2$$

where the relative position (R_{ik}) and orientations (Ω_{ik}) of atom i of type ι in molecule M and atom k of type κ in molecule N is defined by the relative position and orientation (R_{MN}, Ω_{MN}) of the molecules M and N. The components of the above equation are discussed in detail in **Chapter 1** & **2**. This DIFF uses the axes required by the conventions within the DMACRYS, which enable crystal structure modelling program NEIGHCRYS[283] to transform the multipole moments to maintain a right-handed axis system when molecules are generated by inversion or other symmetry operators that would otherwise invert the axis definition. Due to the change in axis systems, the short-range coefficients had to be refitted and the distributed polarizabilities were rotated into the molecule-fixed axis (**Chapter 2**). Nonetheless, this DIFF is essentially identical to Model 3[38] bar the axis definitions. In order to clarify the symmetry relationships within pyridine, the parameters of the DIFF are given in full in **Appendix 3.A** in the ORIENT format.

3.2.1.1 Long-range terms

Equation 86 includes the standard multipolar expansion of the long-range electrostatic term, $V_{elst}^{(1)}[DM]$, given as $Q_t^i T_{tu}^{ik} Q_u^k$, and polarization term, $V_{POL}^{(2-\infty)}[DM]$ in terms of the induced moment ΔQ_t^i on atom i. These long-range contributions have been obtained from the asymptotically corrected[284] (PBE0/AC) molecular charge density, which was calculated with the d-aug-cc-pVTZ Dunning basis[214] on the isolated molecular structure. The molecular charge density was then appropriately partitioned using the Iterated Stockholder Atoms (ISA)[233] partitioning method (**2.5.1.7**), in order to obtain the distributed multipole moments. These multipole moments were evaluated up to rank 4 (hexadecapole) to include the contributions of the atomic anisotropy, such as the lone-pair (dipole) and π-orbital (quadrupole) features, which are important in organic crystals.[74] The atomic polarizabilities ($\alpha_{lm,l'm'}^i$) required for ΔQ_t^i were limited to just the dipolar terms ($l = l' = 1$) and the atomic dispersion coefficients ($C_n^{\iota\kappa}$) were kept isotropic. For his pyridine models Misquitta derived these terms using the Williams-Stone-Misquitta (WSM) polarizability model.[285-287] The C_8 and C_{10} dispersion terms are included for heavy atoms, to ensure the correct distance dependence for the dispersion, as the relative orientations of molecules in crystals sample the dispersion out to infinity. In order to prevent unphysical behavior at small R, the polarization and dispersion terms are damped using a single atom-atom parameter Tang-Toennies damping function.[271] These damping parameters were determined by Misquitta as $\beta_{ind} = 1.25$ a.u. and $\beta_{disp} = 1.67$ a.u.. A bespoke code was written to translate all parameters from CamCASP input to DMACRYS format so they could be used in the in-house systems, with the correct atom labels, numbering and units. These scripts were written using the programming language Python2.7.[31]

3.2.1.2 Short-range terms

For pyridine, Misquitta used the CamCASP and ORIENT programs to refit the short range contributions to the intermolecular potential using the local axis definition in ***Figure 9***. The short-range term was made anisotropic by the inclusion of an orientation-dependent shape function ($\rho^{\iota\kappa}(\Omega_{ik})$). This shape function is a polynomial containing anisotropic coefficients $\rho^{\iota\kappa}$ that have been defined relative to the local axis system (***Figure 9***). As only a limited number of $\rho^{\iota\kappa}(\Omega_{ik})$ terms are available in DMACRYS (***Table 5***), the terms in the original model[38] cannot be simply rotated. A preliminary model was created by employing the distributed density-overlap model to fit the first-order SAPT(DFT) energies computed for almost 4000 pseudo-random dimer configurations. The fit was then refined using an additional 750+ infinite-order SAPT(DFT) calculations and assessed using 250+ infinite-order energies from an independent dataset. The fitting was performed in a hierarchical manner to obtain an atom-atom exponential repulsive model with a pre-exponential (G) and isotropic hardness ($\alpha^{\iota\kappa}$) coefficients.[38]

3.2.1.3 Testing the translated analytical models

The transferring of the DIFF between codes was tested by the calculation of the second virial coefficients (a gas-phase property) and by reproducing the previous ORIENT gas-phase calculations[38] in DMACRYS calculations, where identical dimers were set up within cubic unit cells of lengths 100 Å. Two other non-empirical variants of this model were also tested. DIFF(no-pol) simply omits the polarization term, which is the only many-body contribution included in the DIFF. The effect of the anisotropic repulsion was also tested by refitting the short-range terms to an isotropic model to give DIFF(iso-rep)). In this isotropic model $\rho^{\iota\kappa}(\Omega_{ik}) = 0$ and only the isotropic short-range coefficient $\rho_{00}^{\iota\kappa}$ is included. The non-empirical models were also compared against two empirically fitted isotropic atom-atom *exp-6* potentials (**1.2.2**) with the following functional forms

$$U_{M,N}(R_{MN}, \Omega_{MN}) = \sum_{i \in M, k \in N} U_{EMP}^{ik}(R_{ik}, \Omega_{ik}) = G\exp[-\alpha^{\iota\kappa} R_{ik}] + Q_t^i T_{tu}^{ik} Q_u^k - \frac{C_6^{\iota\kappa}}{R_{ik}^6} \tag{87}$$

where atom i is of type ι (C/H/N) and atom k of type κ. The empirical models use a simpler short-range and dispersion description than the DIFF. The *exp-6* FIT parameters[40] relevant to pyridine were originally fitted to aza-hydrocarbon crystal structures and a limited set of crystal energies[72] whereas the WILL01 set discriminates between different types of N hybridization.[70] The FIT model is combined with the following electrostatic models: the same ISA distributed multipoles as used in DIFF (labeled ISA), the GDMA point charge only component (labelled POINT) and the GDMA electrostatic model (labelled GDMA). The GDMA multipoles are distributed multipoles from the GDMA2[230] analysis of the PBE0/6-31G(d, p) molecular charge density of pyridine. GDMA has often been combined with FIT in crystal structure prediction studies.[19] Consequently, different wave-functions (Ψ) for the GDMA multipoles and the ISA multipoles were used to produce each electrostatic model. The WILL01 potential definition includes foreshortening of the H interaction sites, meaning they are moved slightly into the C-H bond. This to reflect the non-spherical nature of the charge distribution around hydrogen atoms, however, the distributed multipoles for these non-atomic sites could only be obtained using GDMA2.[230]

3.2.2 Simulation Methods

3.2.2.1 2nd virial Coefficients

The virial coefficients derive from the virial expansion of the pressure of a gas (P) in a system (Equation 88). This virial expansion describes deviations from the ideal gas law in terms of a power series in density (ρ). There is a general dependence on temperature (T) and the interaction potential between the bodies in the system.[288] For example, the 2nd virial coefficients ($B_2(T)$) signifies the systems initial deviation from ideality and depends on the pressure due to pairwise additive terms in the interaction potential. The 3rd virial coefficients ($B_3(T)$) depend on non-additive 3-body interactions and so on.

$$\frac{P}{kT} = \rho + B_2(T)\rho^2 + B_3(T)\rho^3 + \cdots \tag{88}$$

where k is the Boltzmann constant. In the DIFF, the polarization term is the only non-additive many body component, therefore, the second virial coefficients are an appropriate test of the intermolecular pair potential in the gas phase, and could be compared with experimental values[169, 289] derived by vapour compressibility measurements for pyridine in the 1950s. These experimental observations gave a heat of vaporization of $35.2\ kJ\ mol^{-1}$ at the boiling temperature of 115.26°C and $40.4\ kJ\ mol^{-1}$ at 25°C. The second virial coefficients have been calculated using ORIENT,[108] by numerical integration of the standard formulation.[290]

$$B_2(T) = -\frac{1}{2}\iint \left[\exp\left(\frac{-V(\mathbf{r},\boldsymbol{\Omega})}{kT}\right) - 1\right] \mathrm{d}\boldsymbol{\Omega}\mathrm{d}^3\mathbf{r} + \frac{\hbar^2}{24(kT)^3}\left[\frac{\langle \mathbf{F}^2\rangle_0}{M} + \sum_{\alpha}\frac{\langle T_\alpha^2\rangle_0}{I_{\alpha\alpha}}\right] \tag{89}$$

The 1st term in the expression can be used to obtain the classical second virial coefficients, where $V(\mathbf{r})$ is the interaction potential energy. Normally, ORIENT includes the 2nd term in Equation 89, the quantum correction to order $\hbar^2$, where $\langle \mathbf{F}^2\rangle_0$ is the mean square force, $\langle T_\alpha^2\rangle_0$ is the mean square torque on one molecule of molecular mass M and a molecular principal moments of inertia $I_{\alpha\alpha}$. However, for this pyridine study, this term was omitted to save a considerable amount of computation time.

3.2.2.2 Lattice Energy Calculations

$$E_{latt} = E_{int} + \Delta E_{intra}^{molecule} \tag{90}$$

Pyridine is held rigid so $\Delta E_{intra}^{molecule} = 0$ and the lattice energy depends only on the atom-atom intermolecular interaction energy. The crystal structures and their lattice energies were modelled using DMACRYS2.2.1.1[40], with lattice summations being carried out to 15 Å followed by a 2 Å splined correction, with the charge-charge, charge-dipole and dipole-dipole electrostatic contributions being evaluated by Ewald summation. Lattice energy minimization used the analytical first derivatives apart from the forces due to the polarization term. As the polarization term is the only non-additive many-body term in the potential, its use in the solid-state poses new challenges (**2.6.1**). The induced moments in a crystal structure need to be solved iteratively to consistency[40] because the polarizing field depends on the polarized moments of all the molecules in the crystal. Consequently, calculating the polarization forces requires taking numerical derivatives of the iterated distributed induced moments. As smooth, numerical forces are needed, minimization with the polarization energy is very computationally demanding. This has been done for the pyridine experimental structures (***Table 6***), where it was confirmed that the including the forces due to polarization ($\mathrm{Pol}_{\mathrm{min}}$) changed the lattice parameters from those minimized without the forces due to polarization (no-$\mathrm{Pol}_{\mathrm{min}}$)by less than ±0.04 Å. Moreover, the geometric changes resulting from including polarization forces caused a negligible further change in the lattice energy.

Structure	FORM I			FORM II		
	no-Pol$_{Min}$	Pol$_{Min}$	*% change*	no-Pol$_{Min}$	Pol$_{Min}$	*% change*
$\boldsymbol{E_{latt}}$	**-64.94**	**-65.02**	***-0.13***	**-66.76**	**-66.46**	***0.44***
$V_{elst}^{(1)}[DM]$	-12.19	-12.18	*0.01*	-16.83	-16.75	*0.45*
$V_{disp}^{(2)}[DM]$	-94.37	-94.33	*0.04*	-97.97	-97.76	*0.22*
V_{SR}	49.23	49.19	*0.09*	52.87	52.62	*0.47*
$V_{POL}^{(2-\infty)}[DM]$	-7.62*	-7.70	*-1.05*	-4.83*	-4.57	*5.27*
Volume / Å^3	406.971	407.031	-0.01	401.685	401.982	-0.07
Density / g cm^{-3}	1.291	1.291	0.02	1.308	1.307	0.08
a / Å	17.066	17.068	-0.01	5.378	5.373	0.10
b / Å	8.672	8.669	0.03	6.713	6.746	-0.49
c / Å	11.000	11.004	-0.04	11.125	11.091	0.31
β / °	90.000	90.000	0.00	90.000	90.000	0.00

Table 6: *Solid-state energy contributions and cell geometries for the protonated (h5-I & h5-II) experimental structures of pyridine (PYRDNA04, PYRDNA05)*[186] *(in kJ mol^{-1}). The structures were minimized using the DIFF, minimizing with (Pol$_{Min}$) and without (no-Pol$_{Min}$) the forces due to polarization. *Single point polarization energies were added after optimization. The lattice energy (E_{latt}) contributions for the experimental structures of Form I & II have been denoted $V_{elst}^{(1)}[DM]$ for the electrostatic, $V_{disp}^{(2)}[DM]$ for the dispersion, V_{SR} for the short-range repulsive term and $V_{POL}^{(2-\infty)}[DM]$ for the polarization.*

A significant effort was required to minimize the lattice energy including polarization forces, and since the calculations in ***Table 6*** show that the cell geometry with polarization forces included is almost identical to that obtained by just adding the polarization energy after minimization, all other pyridine minimizations and second derivative properties calculations did not include the derivatives of the polarization contribution. A single-point polarization energy was added for the DIFF and DIFF(iso-rep). This approximation may only be acceptable because the induced moments are small, as shown by a small induced moment electrostatic potential on the van der Waals surface (***Figure 11***). However, the induced moments do vary with the changes in crystalline environment, with all the four independent molecules in form I having a less symmetric environment, and hence greater polarization, than the single independent molecule in form II (***Figure 11***).

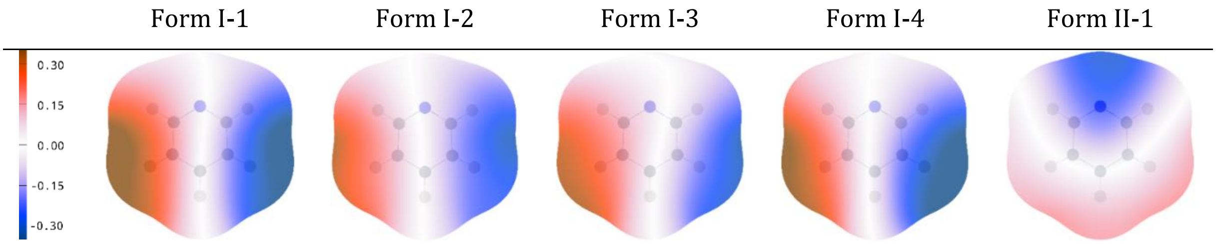


Figure 11: *The additional electrostatic potential on the iso-density surface of* 10^{-3} $electron/bohr^3$ *around each pyridine molecule in the two crystals arising from the induced atomic dipole moments, after optimization with the DIFF. The potential in eV is displayed on the van der Waals surface using ORIENT 4.9.08*[108]*. The lower symmetry environment in* $Z' = 4$ *form I produces a marked difference in the induced moments relative to* $Z' = 1$ *form II and each molecule in form I.*

3.2.2.3 Lattice Dynamics

To investigate the effects of temperature, the energy difference between the hypothetical static crystal at $0K$ and a real crystal was estimated by lattice dynamics with the harmonic approximation (**2.1.1.2**).[11, 183] The Helmholtz free energy (Equation 91)

$$A_{free}(T) = E_{latt} + F_{vib}(T)$$

91

was evaluated at 190 K as this is the temperature at which there is an experimental crystal structure of both form I and II but this is only 40 K below the melting point of pyridine. Hence, the use of the harmonic approximation is unlikely to be realistic, as discussed in **3.4.3**. The effects of pressure were calculated with DMACRYS by optimizing the cell geometry including the PV contribution to the energy, to minimise the lattice enthalpy $H_{latt} = E_{latt} + PV$.

3.2.2.4 Crystal Structure Prediction

The pyridine CSP study follows the CSP workflow detailed in **Chapter 1**, using CrystalPredictor 1.8[35, 36] to generate a million putative $Z' = 1$ crystal structures of pyridine within the 59 most probable space groups. Optimization using the FIT potential and ISA point charges and then clustering generated just over 5000 unique structures. This set was re-minimized with the various non-empirical and empirical force-fields, any minimized structures whose second derivative properties showed that they were not true minima were discarded, and the remaining structures were re-clustered. The most stable structures were analysed in detail for their similarities to each other and the isolated dimer structures. This was done using the similarity tool in Mercury[291] which determines how many molecules (n) of a maximum coordination cluster (15 for crystals, 2 for dimer comparisons) can be matched within a 20% distance in intermolecular atom-atom distances and 20° in interatomic intermolecular angles, and reports the optimum RMSD$_n$ (root mean square deviation of n molecules) of the overlay, ignoring hydrogen or deuterium atoms. A general overview of the CSP process can be found in **1.1.2.2**.

3.3 Results

3.3.1 Gas-Phase Properties

3.3.1.1 Classical 2nd virial coefficients

The classical second virial coefficients (***Figure 12***) differ markedly between the DIFF and empirical models, with the former being closer to the experimental values. The empirical FIT potential underestimates the effect of the intermolecular forces on the gas-phase property by approximately 25%, whereas the DIFF appears to slightly overestimate the effect by around 10%. The slightly over-binding nature of the DIFF is discussed in more detail in **3.4.5**.

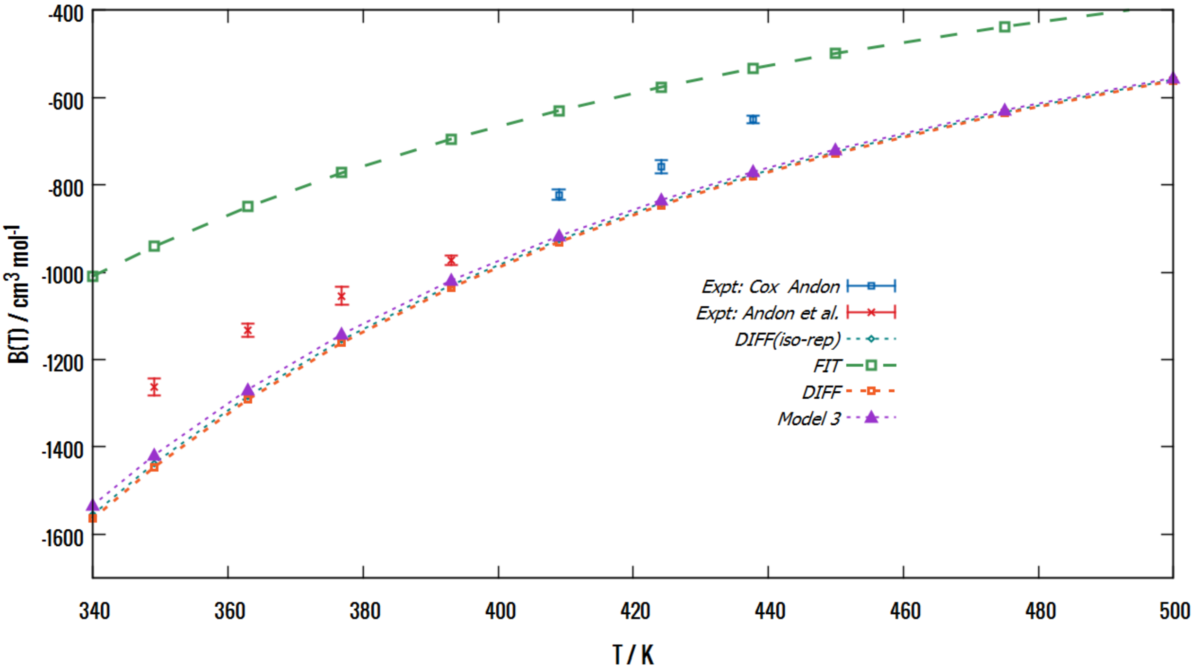


Figure 12: *The second virial coefficients of pyridine calculated from the DIFF and FIT+GDMA potentials, contrasted with the experimental values.*[169, 289] *The values for the DIFF differ from those previously reported*[38] *for Model 3 which were affected by a coding error. The equivalence of DIFF and Model 3 is shown.*

3.3.1.2 Gas-phase minima

In the gas-phase, pyridine's intermolecular potential energy surface has eight distinct minima[38] (***Figure 13***), the most stable minima (Hb1) having two $CH \ldots N$ weak hydrogen bonds (H-bond). Other minima exhibit only one H-bond, while some have distorted T shaped geometries stabilized by $CH \ldots \pi$ interactions, as well as a trio of displaced stacked geometries (***Figure 13***). It is notable that the dimers differ markedly in the dominant contribution to the binding energy; the electrostatic plus polarization contribution is similar to the total binding for the hydrogen bonded Hb1 dimer, but this contribution is less than 1 $kJ\ mol^{-1}$ for the almost iso-energetic stack S1, which has double the dispersion stabilization. There are three distorted T dimers with CH···π interactions and another C-H close to a nitrogen, which are quite similar in energy and structure apart from the relative position of the nitrogen. These T-shaped dimers have all energy components intermediate between the hydrogen-bonded and stacked dimers.

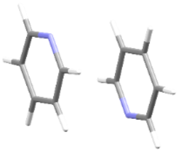
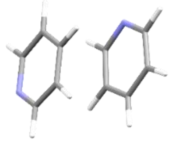
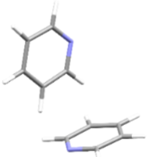
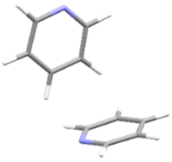
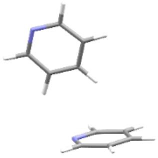
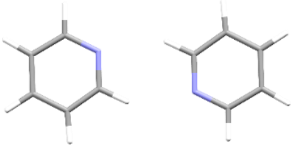
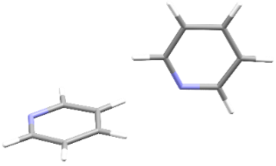
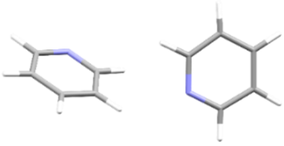
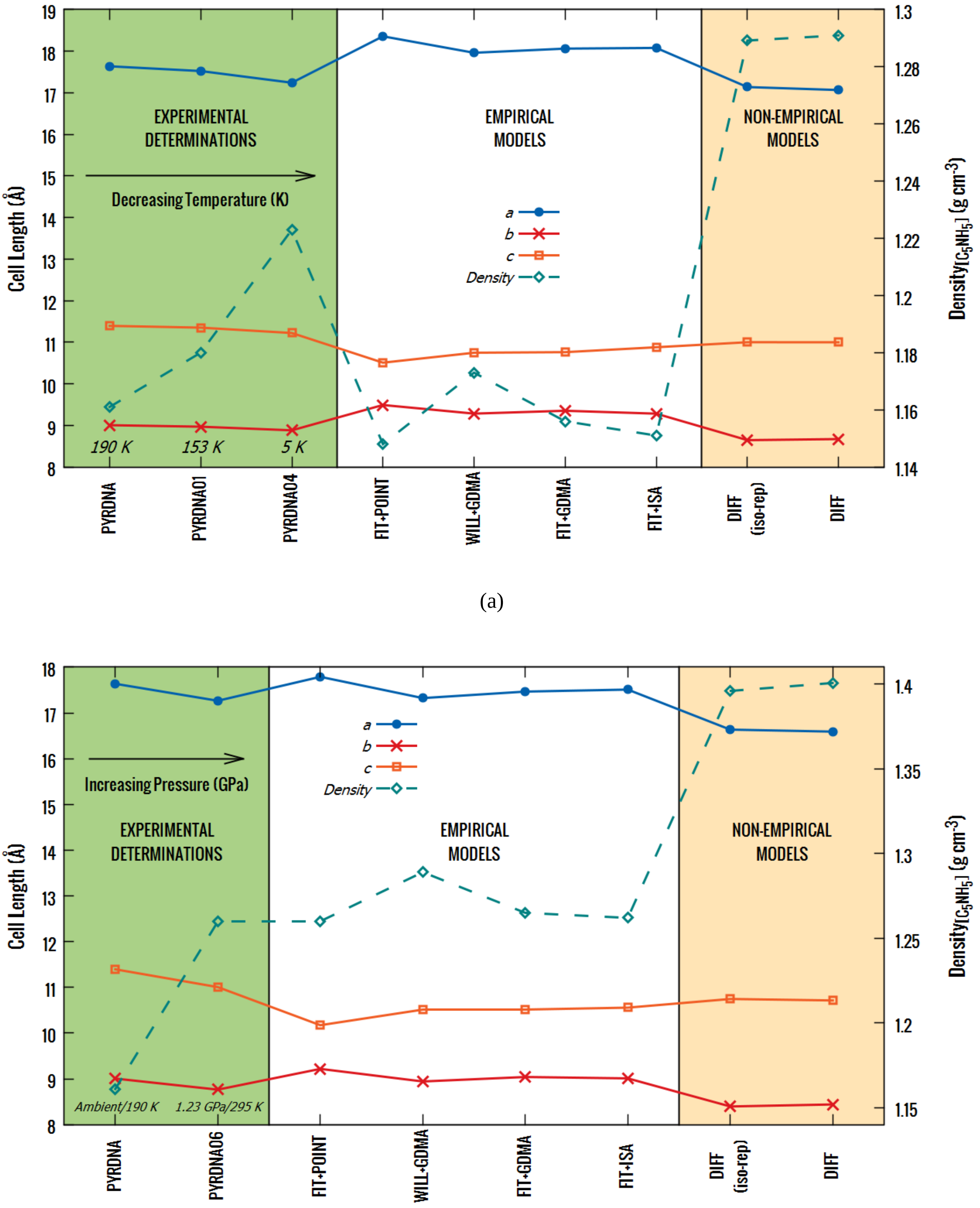
Dimer Label	Structure	Dimer Energy Terms (kJ mol⁻¹)				V_{int}
		$V_{elst}^{(1)}[DM]$	$V_{disp}^{(2)}[DM]$	V_{SR}	$V_{POL}^{(2-\infty)}[DM]$	
S1		-0.87	-28.88	14.85	-0.72	-15.63
S2		-0.29	-29.36	14.98	-0.72	-15.39
T1		-6.45	-19.99	12.26	-0.98	-15.17
T2		-6.47	-19.59	11.79	-1.02	-15.30
bT		-5.84	-19.19	11.38	-1.03	-14.67
Hb1		-12.94	-15.76	14.68	-2.53	-16.56
Hb2		-8.51	-15.72	10.68	-1.68	-15.22
Hb3		-8.23	-15.14	10.61	-1.57	-14.34

Figure 13: *The most stable minima in the pyridine intermolecular potential energy surface (gas phase dimers), with their DIFF energies (*V_{int}*) broken down into components; multipolar electrostatic (*$V_{elst}^{(1)}[DM]$*), damped dispersion (*$V_{disp}^{(2)}[DM]$*), short-range (*V_{SR}*) and damped multipolar polarisation (*$V_{POL}^{(2-\infty)}[DM]$*).*

3.3.2 Reproduction of known crystal structures of pyridine

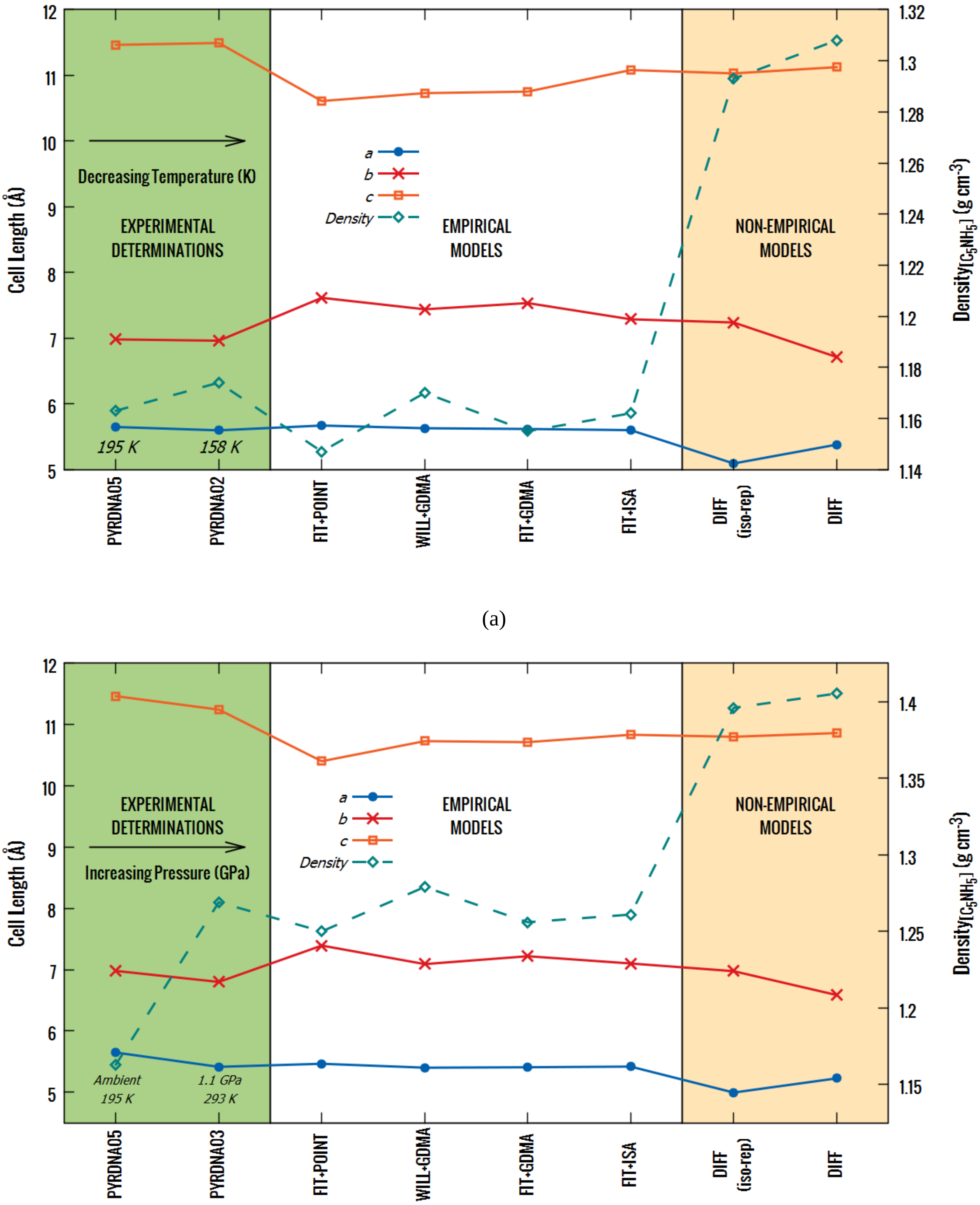
Figure 14 and ***Figure 15*** give comparisons of the cell parameters and densities of cell geometry optimizations with various analytical potentials. In ***Figure 14*** these optimizations are contrasted with the effect of temperature on three experimental determinations of form I between $190K$ and $5K$, and the two determinations of form I at two pressures (but different temperatures). ***Figure 15*** investigates changes in cell geometry as a function of temperature and pressure for form II. Lattice energy minimization (which corresponds to a static crystal geometry at $0K$) using the majority of the potentials gives a satisfactory reproduction of the two polymorphs when compared with the variation of observed structures with temperature. The cell parameters of the non-empirical DIFF are found to better match the $5K$ form $h5$-I structure (PYRDNA04) than the empirical models, while the empirical models that include a distributed multipole electrostatic model (GDMA/ISA) better match the densities at higher temperatures (***Figure 14* (a) & *Figure 15* (a)**). Accessible variations in pressure generally have a larger effect on the crystal structures than temperature changes, but since pressure experiments are usually done at ambient temperature, it is not possible to de-convolute temperature and pressure effects on pyridine which is only solid at ambient temperatures when under pressure. Allowing for these differences in temperatures, minimizing the lattice enthalpy seems to provide a sensible estimate of the structures under modest pressures ($1.1GPa$ ***Figure 14*** **(b)**, $1.23GPa$ ***Figure 15*** **(b)**) for all the intermolecular potentials.



(a)

(b)

Figure 14: *Comparison of experimental and calculated structures of pyridine form I, using different empirical and non-empirical model potentials. (a) form I structures determined at different temperatures vs the geometries obtained from the various models. (b) form I determinations as a function of pressure, against lattice enthalpies calculated including the PV energy at 1.23 GPa (All d5 structure densities have been converted to h5 equivalents).*



(a)

(b)

Figure 15: *Experimental determinations of form II vs the optimized structure computed with different empirical and non-empirical model potentials. (a) form II structures determined at different temperatures vs the geometries obtained from the various models. (b) form II determinations as a function of pressure, against lattice enthalpies calculated including the PV energy at* $1.1 GPa$*, the observed transition pressure*[186] *(All* $d5$ *structure densities have been converted to* $h5$ *equivalents).*

3.3.3 Stability of Observed versus Hypothetical Structures

3.3.3.1 *Crystal Structure Prediction Searches with various models*

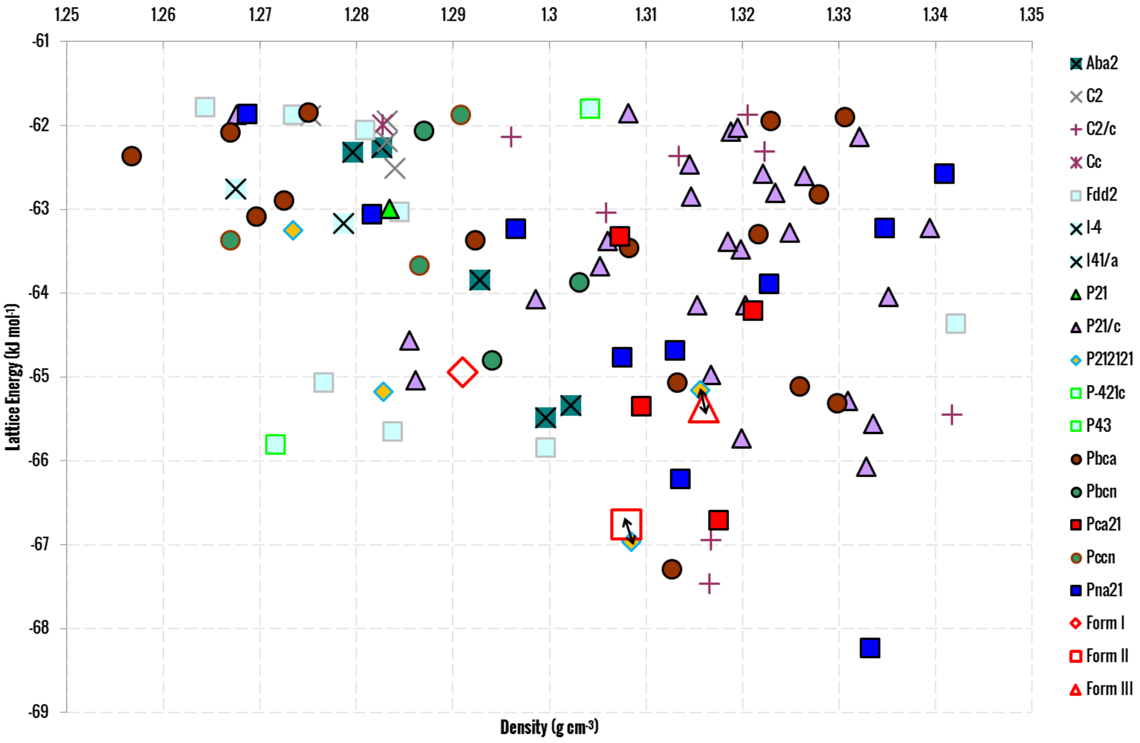
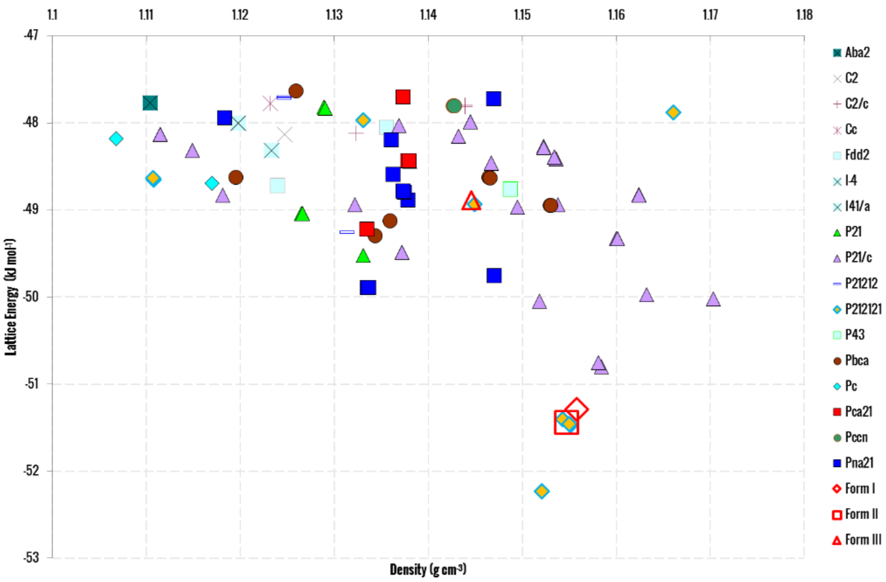


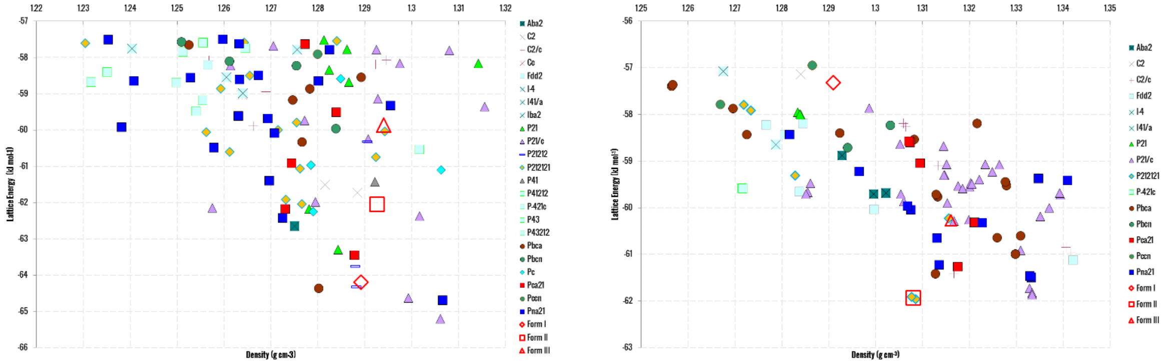
Figure 16: *CSP generated crystal structures of pyridine ranked by lattice energy calculated with the DIFF. Each point represents a mechanically stable crystal structure, classified by its space group. The lattice energy minima obtained by minimizing the experimental structures with a rigid planar molecular structure and the DIFF are shown by open red symbols, and are joined by tie lines to the corresponding CSP generated structure found in this* $Z' = 1$ *search. The above figure shows the 100 most stable structures. The space group, lattice parameters, energy, and the corresponding dimer motif for the 30 most stable structures can be found in the* ***Appendix 3.B****, linked to the structure identifier* $_{pyr}$*#. The complete results of the DIFF CSP can be found in the* ***Appendix CD****.*

As this is a $Z' = 1$ study, the search is not capable of finding form I or any structures with a $Z' > 1$. Regardless, CSP lattice energy landscape with DIFF generated many structures competitive in energy with forms I and II (***Figure 16***), with 52 unique structures within 5 kJ mol^{-1} of the most stable. The lowest energy structure ($_{pyr}$2) is denser and 1.3 $kJ\ mol^{-1}$ more stable than form II. Therefore, this non-empirical model passes the test of predicting the observed structures to be well within an energy range of the global minimum, which is believed to be associated with thermodynamically plausible polymorphs.[7]

Given the experimental uncertainty that the observed structures are the most thermodynamically stable, and the neglect of temperature effects in these models, as discussed further in **3.4**, the DIFF has transferred successfully to the crystalline state. It is therefore worth determining the importance of the different contributions to the relative stability ranking, in light of what parts of the potential energy surface are sampled by the CSP study.



FIT+GDMA



DIFF(iso-rep) DIFF(no-pol)

Figure 17: *The above figure analyses the 100 most stable CSP generated crystal structures of pyridine ranked by lattice energy calculated with the empirical FIT+GDMA potential and approximations to the DIFF. Each point represents a mechanically stable crystal structure, classified by its space group. The corresponding minima for the known polymorphs are labelled. This $Z' = 1$ search is not capable of generating form I. Larger images of each CSP can be found in the* ***Appendix CD****.*

The ranking of the low energy structures varies with force-field (***Figure 16***, ***Figure 17*** & ***Figure 18***). Removing the polarization energy DIFF(no-pol) (***Figure 17***) leads to considerable re-ranking, with form II becoming the most stable, and the energy difference between the two forms increasing to nearly 5 $kJ\ mol^{-1}$. The model fitted with an isotropic repulsion model (DIFF(iso-rep)) produces a very different potential energy landscape. A reversal in the stability order of form I and II was seen, and a different selection of structures (***Figure 17***) that are slightly more stable than the known forms. The empirical FIT+GDMA potential gives the observed form II as very close in energy ($0.14\ kJ\ mol^{-1}$) to form I, but there is also one lower energy structure ($0.8\ kJ\ mol^{-1}$) $_{\text{pyr}}$83, which is significantly different (***Figure 17***). Moreover, the empirical potential produces structures that are less dense than those generated with the non-empirical potentials, but are closer in density to the observed structures (***Figure 14*** & ***Figure 15***).

3.3.3.2 Analysis of contributions to the lattice energy and prominent interactions

Analysis of the lowest energy crystal structures (**Appendix 3.B**) suggests that the coordination clusters are not dominated by the most stable gas-phase dimer structures found in ***Figure 13***. The most stable crystal structure ($_{\text{pyr}}$2) contains the T-shaped dimer T1 and a much distorted version of the singly hydrogen bonded dimer Hb2, whereas the most stable dimer Hb1, occurs in a crystal structure ($_{\text{pyr}}$83) that is $1\ kJ\ mol^{-1}$ less stable. Thus the low energy crystal structures differ in the partitioning between the various energy contributions[292] in a similar manner to the dimers (***Figure 13***), and they differ in the relative contribution from each of the defining dimers that are in van der Waals contact. ***Figure 18*** displays the relative energies of a selection of structures within $4\ kJ\ mol^{-1}$ of the most stable, together with the three known polymorphs of pyridine. The hypothetical structures were selected based on the diversity of the gas-phase-like dimers they contained (**Appendix 3.B**), so as to allow us to clearly explore the effects of the various energy models on the relative lattice energy and its components, most importantly the impact of the non-pairwise additive polarization term. The relative thermodynamic stability order varies significantly as a function of potential (***Figure 18***). There is only a slight re-ranking of the structures in changing between the two empirical repulsion-dispersion models (FIT/WILL+GDMA), but there is a dramatic re-ranking caused by omitting the polarization energy (DIFF(no-pol)) or the short-range anisotropy (DIFF(iso-rep)) in the DIFF approximations. In addition, the global minimum from the DIFF ($_{\text{pyr}}$2) is less stable than forms I and II with the empirical models, and being denser, its stability may be an artefact of the DIFF, as discussed in **3.4.5**. There is also some variation with the electrostatic model used. A re-ranking is observed, caused by using a point charge or distributed multipole representation of the same charge distribution (contrast FIT+POINT with FIT+GDMA). This re-ranking, however, is less severe than that caused by changing the quality of charge distribution and its representation (contrast FIT+GDMA with FIT+ISA). This is unusual, as the use of distributed multipoles rather than the corresponding potential-derived charges usually makes a considerable improvement in CSP studies, particularly for hydrogen-bonded systems.[74]

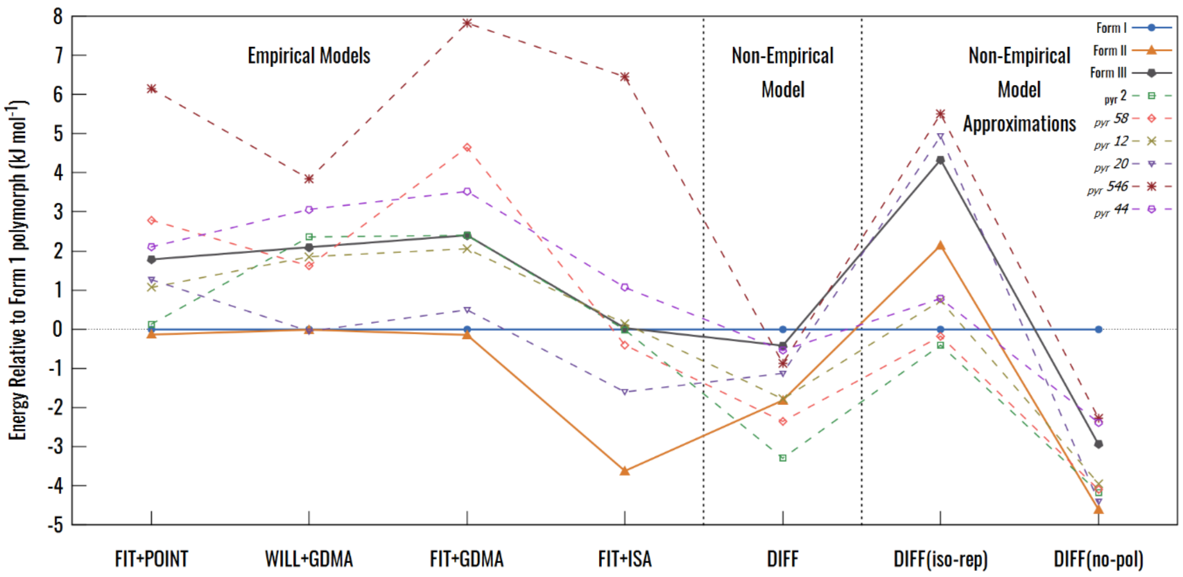


Figure 18: *Relative lattice energies of the observed and selected computer generated crystal structures of pyridine, relative to form I as a function of the intermolecular force-field. The labels for the CSP generated structures* $_{pyr}\#$ *refer to their ranking after the CrystalPredictor*[35, 36] *stage, i.e. with the FIT+POINT model and further details on their structures are given in the* ***Appendix 3.B****.*

$_{pyr}\#$	Form I	Form II	Form III	2(GM)	58	12	20	546	44
Dimer motifs	T2, bT, Hb3	T2, Hb3	None	T1	Hb1	Hb2	S1, T1	T2, Hb3	None
$V_{elst}^{(1)}[DM]$	-12.19	-16.83	-13.21	-11.74	-14.45	-13.58	-13.47	-15.86	-13.22
$V_{disp}^{(2)}[DM]$	-94.37	-97.97	-100.53	-104.13	-99.57	-100.89	-102.50	-95.71	-99.51
V_{SR}	49.23	52.87	53.48	54.37	52.59	53.20	54.24	51.97	53.02
$V_{POL}^{(2-\infty)}[DM]$	-7.62	-4.83	-5.09	-6.74	-5.87	-5.46	-4.34	-6.21	-5.77
Lattice	**-64.94**	**-66.76**	**-65.36**	**-68.24**	**-67.30**	**-66.72**	**-66.07**	**-65.81**	**-65.48**

Table 7: *Solid-state energy contributions (in kJ mol*$^{-1}$*) for DIFF in the observed and the most diverse hypothetical structures.*

The lattice energies of the selected crystal structures vary by less than 4 $kJ\ mol^{-1}$ using the DIFF (***Table 7***), however, the various contributions to this energy vary considerably more. The dominant dispersion contribution varies by over 10 $kJ\ mol^{-1}$ favouring denser structures that contain stacked or T-shaped dimers, whose dimer energies are heavily dominated by the dispersion contribution (***Figure 13***). But this is often partially balanced out by an increase in the short range repulsion term. The electrostatic contribution varies by 5 $kJ\ mol^{-1}$, favouring the experimental forms and other structures containing the $CH \ldots N$ “hydrogen bonds”. The polarization energy is a small contribution, but stabilizes form I ($Z' = 4$) the most, and the difference in polarization energy between the two known polymorphs (2.8 $kJ\ mol^{-1}$) is larger than their total lattice energy difference. Although this difference is accounted for by the lower symmetry within form I allowing larger induced moments (***Figure 11***), the polarization term also stabilizes $_{\text{pyr}}2$, and all other structures that are lower in lattice energy than form II. Hence the polarization contribution is important, even for pyridine and the low energy structures re-rank substantially in the CSP which neglects this structure-dependent non-additive term by using DIFF(no-pol) (***Figure 17***).

3.3.4 Form III

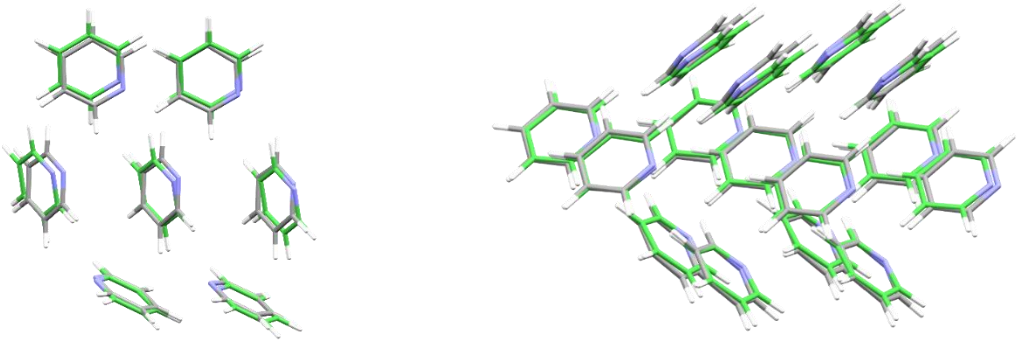

(a) (b)

Figure 19: *(a) Structural overlay of the experimental crystal structures of form II (grey $h5$-II PYRDNA05*[186]*) at 2GPa and the form III of pyridine, refined from experimental powder data, at around 2GPa, showing the optimum overlay of 7/15 molecules, with* $RMSD_7 = 0.364$ *Å. (b) Structural overlay of the crystal structures of the proposed form III (green* $_{\text{pyr}}35$*) calculated at 2GPa and the neutron structure determined at approximately this pressure (grey)*[278] *showing the optimum overlay of 15/15 molecules, with* $RMSD_{15} = 0.212$ *Å.*

On hearing of the occurrence of a new polymorph $d5$-III at around $2GPa$, the crystal energy landscape was recalculated at $2GPa$ using the FIT and DIFF potentials. The pressure-volume (PV) energy term significantly re-ordered the relative stability of the structures (**Appendix 3.B**) as well as increasing the computed density by up to 15%. One structure became significantly more favourable with pressure and was only 1.2 $kJ\ mol^{-1}$ less stable than form II. Furthermore, this structure ($_{\text{pyr}}35$) had 7 coordinating molecules in common with form II, and so it is a very plausible result of a low-barrier transformation (***Figure 19*** (a)), unlike the other low energy hypothetical structures at $2GPa$. Based off its relative energy at $2GPa$ and being the structure in the search with the most plausible degree of rearrangement from the known forms, it seemed likely to be form III. The overlay of this structure ($_{\text{pyr}}35$) with the experimental structure of form III[293] at approximately $2GPa$ is shown in ***Figure 19*** (b). In contrast, the structure of form III could not be identified with the empirical FIT potential. The corresponding structure $_{\text{pyr}}35$ is not amongst the most stable, independent of pressure (***Figure 17*** & **Appendix 3.B**), and is

$3.3\ kJ\ mol^{-1}$ less stable and less dense than form II at $2GPa$. Thus, it would not have been possible to propose a candidate structure for form III using the FIT potential solely from knowledge of its existence at ca. $2GPa$.

3.4 Discussion

The non-empirical DIFF for pyridine gives a reasonable account of the structures of the polymorphs and their stability relative to hypothetical structures. In this instance, its accuracy in the solid-state exceeds that of empirical potentials that had been parametrized to experimental crystal structures. The empirical potentials fail badly in reproducing the second virial coefficients or the behavior of the solid under pressure. This is a major advance, but this study does bring into focus both the advantages and disadvantages of having a genuine pair potential energy surface, which it was hoped would be transferable between phases with the approximation that the polarization was the only significant many-body term for relative energies.

3.4.1 Reproduction of Crystal Structures

For form II, the static lattice or enthalpy ($P = 1.1GPa$ and $2GPa$) structures from the DIFF are in good agreement with the experimental structures, especially given the variation between the structures that have been determined at different temperatures ($5 - 190K$) and pressures ($\leq 1.1GPa$). The DIFF reproduces the $5K$ h5-I structure of pyridine well, but overestimates the density by 5.6%. Perhaps this is not as serious an issue as it may seem as it has recently been noted that a quasi-harmonic modelling of crystalline imidazole using ab initio based potentials gave a 4% increase in molar volume on including just the zero-point energy.[294] Hence, a part of the error in the density at $5K$ could be due to neglecting the effect of the zero-point expansion on the structure. This again highlights the limitations in using the static lattice energy (or enthalpy with a PV term) to model a material's crystal structures. All things considered, in ***Figure 14*** & ***Figure 15*** it appears that the DIFF gives a reasonable estimate of the experimentally inaccessible static crystal structure. The empirically fitted potentials can only be expected to model structures at similar temperatures and pressures as they have been fitted to, since they implicitly include some averaging over zero-point energy effects, thermal expansion and possibly pressure effects.

3.4.2 How accurate are the lattice energies?

The lattice energy landscape generated for the most mechanically stable structures using the DIFF **(*Figure 16*)** and FIT+GDMA (***Figure 17*)**, shows marked differences. The differences in lattice energies for any one structure are on the order of $15\ kJ\ mol^{-1}$ (***Table 8***) between the DIFF and empirical potentials. This emphasizes how many approximations in both the force-field and simulation methods are partially absorbed by empirically fitting the potential parameters to experimental data similar to that being computationally modelled. For example, a hypothetical heat of sublimation, $\Delta H_{sub} \approx 48.7\ kJ\ mol^{-1}$ at $298K$ (66° above pyridine's normal melting point) can be estimated using the heat of fusion $\Delta H_{fus} = 8.28\ kJ\ mol^{-1}$ (presumably for form I[295]) and derived heat of vapourisation,[169] $\Delta H^{0}_{vap} = 40.4\ kJ\ mol^{-1}$ at $298K$. This heat of sublimation estimate is quite close to the lattice energies calculated using the empirical potentials ($E_{latt} \sim -51\ kJ\ mol^{-1}$ for FIT+GDMA), with a few $kJ\ mol^{-1}$ variations between using the alternative empirical WILL potential parameters, or changing the electrostatic model (***Table 8***). This results from the parameter fitting assuming that $\Delta H_{sub} \approx -E_{latt}$ and highlights the absorption of temperature effects in these potentials. This approximate relationship has sometimes been improved by the common "2RT" thermal correction[270] of $-5\ kJ\ mol^{-1}$, but recent work with lattice dynamics show this is often a poor estimate,[296] and that it can vary for small molecular crystals by up to $-11\ kJ\ mol^{-1}$.[297]

Structure	FIT+POINT	WILL+GDMA	FIT+GDMA	FIT+ISA	DIFF(iso-rep)	DIFF	DIFF(no-pol)
FORM I	-48.71	-49.20	-51.29	-49.66	-64.19	-64.94	-57.32
FORM II	-48.85	-49.22	-51.44	-53.28	-62.05	-66.76	-61.93
FORM III	-46.93	-47.54	-48.89	-49.63	-59.86	-65.36	-60.27
$_{pyr}2$ (GM)	-48.58	-46.84	-48.89	-49.67	-64.60	-68.24	-61.50

Table 8: *Lattice energies (in kJ mol^{-1}) of the DIFF global minimum ($_{pyr}2$), the experimental forms I, II (PYRDNA04, PYRDNA05),[186] and III for each analytical intermolecular model.*

3.4.3 Lattice Dynamics for Zero-point Energy & Thermal Effects

	FORM I	FORM II	FORM III	$_{pyr}2$ (GM)
$\boldsymbol{E_{latt}}$	**-64.939**	**-66.758**	**-65.359**	**-68.237**
$h5$ Zero-Point E	2.027	2.163	2.146	2.114
$F_{vib}(190\ K)$	-6.274	-5.561	-5.716	-6.282
$\boldsymbol{h5\ A_{free}(190\ K)}$	**-71.213**	**-72.319**	**-71.075**	**-74.520**
$d5$ ZPE	1.912	2.039	2.020	1.991
$F_{vib}(190\ K)$	-6.773	-6.050	-6.215	-6.783
$\boldsymbol{d5\ A_{free}(190\ K)}$	**-71.712**	**-72.808**	**-71.574**	**-75.020**

Table 9: *The free energy at $T = 190K,\ P = 0GPa$ calculated using lattice dynamics and the DIFF for the observed polymorphs and the global minimum (GM) in the CSP study in* ***Figure 16****. Energies in $kJ\ mol^{-1}$.*

The lattice energies of the crystals are only an approximation to the thermal stability neglecting the isotope-dependent zero-point energy and thermal contributions. The free energy estimates in ***Table 9*** show that including the zero-point energy and thermal corrections estimated by lattice dynamics, reverses the order of stability of $h5$-forms I and III and significantly reduces the energy difference between forms I and II by over $0.7\ kJ\ mol^{-1}$. While, these harmonic, rigid-molecule lattice-dynamic estimates are approximate, they emphasize that both the zero-point energy and thermal corrections are significant, and affected by deuteration. The effect of deuteration is virtually identical for every polymorph within this fixed rigid-molecule harmonic approximation. For pyridine, the rigid-body harmonic lattice dynamical estimate of the thermal correction (the vibrational free energy (F_{vib})) is $-6\ kJ\ mol^{-1}$ at $190K$ (***Table 9***), subsequently, using the lattice energy from the DIFF (***Table 8***), and allowing for a 7% error from three body dispersion, the zero-point energy and heat capacity improves the agreement; $\Delta H_{sub} = -(1-0.07)E_{latt} + F_{vib} = 65 - 4.5 - 6 = 55.5\ kJ\ mol^{-1}$ at $190K$, which compared with the estimated $\Delta H_{sub} \approx$ 48.7 kJ mol^{-1} at $298K$. Accordingly, the lattice energies given by the DIFF are reasonable.

Most organic crystals have a significant thermal expansion, which can be very anisotropic and dependent on the specific crystal structure. Recent quasi-harmonic periodic electronic structure methods,[297] and empirical potential estimates for a large dataset of organic crystals[298] show that the effects of thermal expansion are thermodynamically significant, producing an underestimation of heat capacities at high temperatures.[296, 299] Moreover, a divergence between harmonic-approximation phonon-modes and those in the crystal as modelled by Molecular Dynamics (MD) has been noted at quite low temperatures for imidazole and 5-azauracil.[188] For pyridine, which is a liquid at room temperature, the molecular motions are likely to be large in amplitude at the temperatures of the majority of experimental measurements, and therefore should be more realistically modelled by a finite-temperature MD simulation. What ***Table 9*** demonstrates, is that both zero-point and thermal contributions are significant for the relative stability of the crystal structures, with the common observation that their inclusion reduces the energy differences between structures.[300]

3.4.4 The advantages of a non-empirical force-field

The success of the DIFF relative to potentials specifically developed for modelling the organic state derives from having a functional form based on the theory of intermolecular forces and not parameterization from experiment. The diversity of gas-phase dimers motifs in the low energy hypothetical and experimental crystal structures emphasizes that a solid-state CSP samples the intermolecular potential energy surface extensively. A single dimer structure only samples the potential around one configuration, whereas a single crystal structure samples a wider range of relative orientations and close contacts from all the molecules in the nearest neighbour coordination sphere, which is typically fourteen molecules. The lattice summation, in lattice energy calculations, samples the potential over a larger distance than a gas-phase dimer. Additionally, the closest contacts in crystals sample higher up the repulsive wall because of the attractive forces from the second, and higher-order coordination spheres, even at ambient pressure. The dispersion contribution is about 150 % of the total lattice energy in the low energy crystals (***Table 7***). Hence, the theoretically based distance-dependence of the long range forces is particularly important in the solid-state. Using the DIFF dispersion model with a split into C_6, C_8 and C_{10} is more theoretically justified than using an empirical C_6 only potential where the higher-order dispersion coefficients have been absorbed in the parametrization. For empirical models, a small error in C_6 could result in a large overall error due to the lattice summation. A similar advantage in the distance dependence is seen when the long range electrostatic terms are given by distributed multipoles over atomic point charges, though the main advantage of the higher atomic multipole moments is a more truthful modelling of the orientation dependence of the electrostatic forces due to lone pair and π electron density. These non-spherical features in the charge distribution also determine the anisotropy in the repulsion, as the short-range anisotropic terms are dependent on the shape of the atom. The slightly closer contacts in solid-state and diversity of contact configurations means that the relative energies of the crystal structures are very sensitive to the anisotropy in the repulsion, as shown by a considerable re-ranking of the crystal structures when anisotropy coefficients are not explicitly in the potential (***Figure 17*** & ***Figure 18***). The successful identification of form III as a structure that becomes relatively more stable under pressure is clear evidence of the importance of using a non-empirical anisotropic potential. It is not surprising that parameterizing an oversimplified functional form of the repulsive wall by fitting to ambient pressure crystal structures fails to properly extrapolate the intermolecular interactions of the closer contacts sampled at moderate pressure. Applying pressure changes the intermolecular contact distances more than changing temperature and so probes interactions higher up the repulsive wall.[301, 302] High-pressure recrystallization is a versatile route to generating new polymorphs; with structural properties modifying significantly with pressures around and above $1\ GPa$.[303-305] Therefore, more realistic, theory-based, non-empirical intermolecular potentials will be very important for more dependable exploration of structure-property relationships in organic crystals under pressure.

3.4.5 Accuracy of the pyridine DIFF

3.4.5.1 Possible errors in the pair-potential

The second virial coefficients show that the DIFF systematically over-estimates the intermolecular interactions (***Figure 12***). It is believed that these errors stem not from problems with the fitting procedure described by Misquitta and Stone[38] but from errors in the reference SAPT(DFT) interaction energies to which the DIFF was fit. A deeper analysis of some of these problems can be found in **Chapter 5**, where the omission of certain terms and further details on the development process of non-empirical potentials are discussed. For the pyridine DIFF there are two main likely causes for the over-binding of the SAPT(DFT) interaction energies:

1. Interaction energy contributions from third to infinite order in the intermolecular interaction operator are approximated by the δ_{int}^{HF}energy[265, 266] (**2.6.3**). This energy includes mainly higher order induction and exchange-induction contributions, which are known to be important for capturing the effects of hydrogen bonding in water. δ_{int}^{HF} has been included in the DIFF, however, this may not be suitable for configurations of pyridine in which the binding is primarily or dominantly from the dispersion interaction,[227, 306, 307] as this correction may lead to an overestimation of the binding.
2. The choice of asymptotic correction used for calculating the molecular properties, made by Misquitta and Stone,[38] may not be suitable for a strongly anisotropic (in shape) molecule such as pyridine. Consequently, a different asymptotic correction is used for TNB in **Chapter 5**.

The first problem is a potentially serious one as the δ_{int}^{HF} term is needed for systems with hydrogen bonds[178], but is known to be inappropriate for dispersion-bound dimers as it leads to over-binding.[308, 309] This term is sometimes excluded entirely[106, 310] if the primary binding is from the dispersion energy as is the case for the benzene crystal. But it is not yet clear what is to be done for a system such as pyridine which exhibits both hydrogen-bonding as well as dispersive interactions. Work is currently underway by Misquitta, using CCSD(T) dimer calculations to better understand both problems, as it is essential for modelling pharmaceuticals where most crystal structures are a balance between hydrogen bonding, π-π stacking and other dispersion interactions and conformational changes.

3.4.5.2 Non-additive many-body terms in condensed-phases

While a pairwise additive intermolecular potential is sufficient for a complete potential energy surface for the isolated dimer, it only accounts for the dominant contributions in condensed phases. The only explicitly non-pairwise term included in the DIFF is the polarization term, which leads to a net attractive many-body contribution to the lattice energy. Yet in a weakly-bound molecular crystal like pyridine, where most of the attractive forces arises from the two-body dispersion energy (***Table 7***), the many-body dispersion energy may be expected to be large and repulsive. For example, it has been shown[103 311] that the non-additive three-body dispersion contributes repulsively to the benzene crystal by as much as 7-14% of the total lattice energy. A similar contribution is expected to arise in the pyridine crystal. In particular, the dense, strongly dispersion bound global minimum structure ($_{\text{pyr}}2$) will be relatively destabilized by the three-body dispersion. This could change the landscape in ***Figure 16*** significantly and result in the observed polymorphs becoming the new global minima.

3.5 Conclusions

This investigation has shown the advantage of using non-empirical potentials for CSP, particularly for form III which was obtained under pressure, where the repulsive wall is sampled in regions not used in fitting empirical potentials. However, it also shows how sensitive the relative thermodynamics are to the modelling assumptions. For example, $_{\mathrm{pyr}}2$ is predicted to be more stable than form II, both at the level of the lattice energy (***Figure 16***) and taking into account thermal effects (***Table 9***), though the difference is considerably reduced on application of pressure (**Appendix 3.B**) and its stability may be because of the neglect of the many-body dispersion term. On the other hand, what would be the consequence of a confident calculation that $_{\mathrm{pyr}}2$ or any other structure was more thermodynamically stable than the observed polymorphs at accessible temperatures and pressures? The prediction of undiscovered more thermodynamically stable phases is a major justification[30] for the development and testing[22] of CSP methods, as this determines the risk of disappearing polymorphs,[312] helps justify devising experiments for finding new polymorphs[313] and can generally assist in the development of specialty organic materials. When polymorphs switch relative stability with temperature or pressure, it may often not be as obvious as a solid-state phase transformation but only apparent by recrystallization experiments in the presence of seeds of both polymorphs. It is difficult to obtain, let alone structurally characterize, the most stable crystalline forms of molecules that are liquid or gas under ambient conditions,[314] let alone change the experimental conditions sufficiently to vary the kinetics (already implicated in the isotopic polymorphism of pyridine), at low temperatures or at pressure[315] to compete with the nucleation of the known forms. The similarities between the known polymorphs of pyridine (***Figure 10*** & ***Figure 19***) and the contrast to the different dimers in the coordination sphere in $_{\mathrm{pyr}}2$ (***Table 7***) emphasize that the observed structures may be kinetically favoured but not necessarily thermodynamically favourable. Pyridine could well have alternative thermodynamically competitive polymorphs that have not yet been found. The DIFF could be used for simulating liquid-phase pyridine using MD, adding confidence to interpreting experimental data,[316, 317] as recent simulations show that this is sensitive to the anisotropy of the electrostatic model.[281] Such simulations could suggest why the observed forms crystallize, by revealing a link between the liquid and solid state structures via the most readily formed nucleus. Non-empirical potentials have considerable advantages over periodic electronic structure methods, or other advanced methods of calculating lattice energies of molecular crystals[130] in that they are relatively inexpensive and can be evaluated sufficiently readily to be used in CSP and for estimating the effect of temperature. Empirical potentials have the advantages of simpler functional forms, and that errors in the functional form, transferability assumptions and in the method of simulation, such as neglect of zero-point and thermal effects, are partially absorbed into the potential. Conversely, these advantages imply uncertainties in extrapolating to other conditions as required for solid-state phase diagrams,[318] and empirical potentials cannot be expected to transfer to the gas-phase, as demonstrated by the second virial coefficients (***Figure 12***). In contrast, the DIFF and other non-empirical models[79, 96] have the advantage that the approximations used are known and can be built upon. Although, the polarization energy is challenging to include in any force-field, this is being tackled in the development of next generation force-fields and simulation codes.[319, 320,321] The improved realism of intermolecular force-fields has to go alongside the further development of simulation methods and codes to include both zero-point effects and realistic temperature dependent dynamics. There is clearly a conceptual advantage in using non-empirical intermolecular force-fields where the approximations are known and controlled and the interaction energy can be decomposed into contributions that are based off rigorous chemical theories. The current potentials are not definitive, but can act as a strong framework for the development of force-fields for organic molecules.

This chapter has shown that a rigid non-empirical distributed intermolecular force-field (DIFF), derived from the theory of intermolecular forces and SAPT(DFT) calculations using the CamCASP and ORIENT programs, can be used for modelling the solid-state of pyridine with a realism that exceeds that of the transferable empirical potentials that have been previously used to model the crystalline state. The following chapter shall investigate the rigid approximation made in this thesis and how the molecular conformation affects the molecular charge density and in turn the associated electrostatic properties and intermolecular interactions. This approximation seems to have been suitable for pyridine but does it extend to more flexible molecules? Furthermore, this investigation emphasizes the importance of CSP as a comprehensive way of testing the robustness of intermolecular potentials. The DIFF was particularly effective for studying the effects of pressure on the relative stability of structures, as shown by the identification of form III. Nevertheless, empirical potentials do have the advantage of absorbing many errors, including approximations in the simulation methods, and being more readily implemented in existing codes due to their simplified functional forms. Contrasting these novel CamCASP-derived potentials, and other models driven by the theory of intermolecular forces, with more approximate models, should produce a hierarchy of force-fields that can be used intelligently with molecule-specific knowledge on the effects of various approximations.

Appendix 3.A – Potential Parameters

The DMACRYS input files with axis definition files, can be found in the Appendix CD. Further details on the formatting of the input files can be found in the DMACRYS manual.[59]

1. DIFF DMACRYS potential input file

This input file is identical to its ORIENT counterparts save for formatting and unit conversions.

2. Iterated Stockholder Atoms – Distributed Multipole Moments (ISA-DMA)

The ISA-DMA multipolar moments are given DMACRYS input file format. These detail the atom labels, atom types and multipole moment ranks used for TNB's electrostatic model. The atomic positions can also be found in this input file, adjacent to the corresponding atom label.

3. Iterated Stockholder Atoms – Distributed Polarizabilities

The ISA-DMA derived $Rank = 1$ distributed polarizabilities in DMACRYS input file format calculated in the molecule-fixed axis.

4. Local axes system

The atom fixed axes system using pyridines atom labels. The .res structure file corresponding to each experimental structure can also be found in the Appendix CD.

Potential Input File

```
! Pair-Potential: Atom-Atom Potential
! Sites:
!      H1 H2 H3 N C1 C2 C3

Induction  damping factor 1.25
Dispersion damping factor 1.67

H1 H1 rho alpha       C6
   00  00    0     0.409815E+01    0.196501E+01    0.440183E+01
   00  10    1    -0.110862E+00
   00  11c   1     0.248000E-01
   10  00    1    -0.110862E+00
   11c 00    1     0.248000E-01
END
H1 H2 rho alpha       C6
   00  00    0     0.393462E+01    0.196342E+01    0.406704E+01
   00  10    1    -0.461940E-01
   00  11c   1    -0.751400E-02
   10  00    1    -0.110862E+00
   11c 00    1     0.248000E-01
END
H1 H3 rho alpha       C6
   00  00    0     0.403103E+01    0.190116E+01    0.373377E+01
```

```
         00  10    1     -0.120796E+00
         00  11c   1      0.978500E-02
         10  00    1     -0.110862E+00
         11c 00    1      0.248000E-01
      END
      H1 N rho alpha       C6      C8      C10
         00  00    0     0.497509E+01    0.177257E+01    0.102929E+02    0.251339E+03
0.538412E+03
         00  10    1      0.104283E+00
         00  11c   1     -0.181740E-01
         10  00    1     -0.110862E+00
         11c 00    1      0.248000E-01
         00  20    2      0.125383E+00
         00  22c   2     -0.163580E+00
      END
      H1 C1 rho alpha       C6      C8      C10
         00  00    0     0.478680E+01    0.183408E+01    0.767587E+01    0.770406E+02
0.544620E+04
         00  10    1      0.130845E+00
         00  11c   1      0.335830E-01
         10  00    1     -0.110862E+00
         11c 00    1      0.248000E-01
         00  20    2     -0.772060E-01
         00  22c   2     -0.113724E+00
      END
      H1 C2 rho alpha       C6      C8      C10
         00  00    0     0.510776E+01    0.167851E+01    0.130303E+02    0.726110E+02
0.462274E+04
         00  10    1      0.120027E+00
         00  11c   1     -0.206020E-01
         10  00    1     -0.110862E+00
         11c 00    1      0.248000E-01
         00  20    2     -0.606040E-01
         00  22c   2     -0.137913E+00
      END
      H1 C3 rho alpha       C6      C8      C10
         00  00    0     0.495024E+01    0.173860E+01    0.957761E+01    0.164475E+03
0.498458E+04
         00  10    1      0.922750E-01
         00  11c   1      0.112700E-01
         10  00    1     -0.110862E+00
         11c 00    1      0.248000E-01
         00  20    2     -0.537020E-01
         00  21c   2     -0.303000E-02
         00  22c   2     -0.124746E+00
      END
      H2 H2 rho alpha       C6
         00  00    0     0.388760E+01    0.202354E+01    0.411769E+01
         00  10    1     -0.461940E-01
         00  11c   1     -0.751400E-02
```

```
         10  00    1     -0.461940E-01
         11c 00    1     -0.751400E-02
      END
      H2 H3 rho alpha        C6
         00  00    0     0.397726E+01    0.193503E+01    0.327252E+01
         00  10    1     -0.120796E+00
         00  11c   1      0.978500E-02
         10  00    1     -0.461940E-01
         11c 00    1     -0.751400E-02
      END
      H2 N rho alpha        C6       C8       C10
         00  00    0     0.484838E+01    0.188781E+01    0.849477E+01    0.245134E+03
0.496401E+03
         00  10    1      0.104283E+00
         00  11c   1     -0.181740E-01
         10  00    1     -0.461940E-01
         11c 00    1     -0.751400E-02
         00  20    2      0.125383E+00
         00  22c   2     -0.163580E+00
      END
      H2 C1 rho alpha        C6       C8       C10
         00  00    0     0.468908E+01    0.188031E+01    0.636298E+01    0.665131E+02
0.430509E+04
         00  10    1      0.130845E+00
         00  11c   1      0.335830E-01
         10  00    1     -0.461940E-01
         11c 00    1     -0.751400E-02
         00  20    2     -0.772060E-01
         00  22c   2     -0.113724E+00
      END
      H2 C2 rho alpha        C6       C8       C10
         00  00    0     0.490780E+01    0.171564E+01    0.873624E+01    0.636031E+02
0.406208E+04
         00  10    1      0.120027E+00
         00  11c   1     -0.206020E-01
         10  00    1     -0.461940E-01
         11c 00    1     -0.751400E-02
         00  20    2     -0.606040E-01
         00  22c   2     -0.137913E+00
      END
      H2 C3 rho alpha        C6       C8       C10
         00  00    0     0.481746E+01    0.187761E+01    0.796230E+01    0.177627E+03
0.406056E+04
         00  10    1      0.922750E-01
         00  11c   1      0.112700E-01
         10  00    1     -0.461940E-01
         11c 00    1     -0.751400E-02
         00  20    2     -0.537020E-01
         00  21c   2     -0.303000E-02
         00  22c   2     -0.124746E+00
```

```
      END
      H3 H3 rho alpha       C6
         00  00    0     0.404245E+01    0.207565E+01    0.176919E+01
         00  10    1     -0.120796E+00
         00  11c   1      0.978500E-02
         10  00    1     -0.120796E+00
         11c 00    1      0.978500E-02
      END
      H3 N rho alpha       C6      C8      C10
         00  00    0     0.493769E+01    0.181110E+01    0.724358E+01    0.180762E+03
0.498610E+03
         00  10    1      0.104283E+00
         00  11c   1     -0.181740E-01
         10  00    1     -0.120796E+00
         11c 00    1      0.978500E-02
         00  20    2      0.125383E+00
         00  22c   2     -0.163580E+00
      END
      H3 C1 rho alpha       C6      C8      C10
         00  00    0     0.483091E+01    0.187809E+01    0.567625E+01    0.620263E+02
0.384651E+04
         00  10    1      0.130845E+00
         00  11c   1      0.335830E-01
         10  00    1     -0.120796E+00
         11c 00    1      0.978500E-02
         00  20    2     -0.772060E-01
         00  22c   2     -0.113724E+00
      END
      H3 C2 rho alpha       C6      C8      C10
         00  00    0     0.502446E+01    0.172752E+01    0.738145E+01    0.582632E+02
0.349286E+04
         00  10    1      0.120027E+00
         00  11c   1     -0.206020E-01
         10  00    1     -0.120796E+00
         11c 00    1      0.978500E-02
         00  20    2     -0.606040E-01
         00  22c   2     -0.137913E+00
      END
      H3 C3 rho alpha       C6      C8      C10
         00  00    0     0.494214E+01    0.184553E+01    0.678423E+01    0.113445E+03
0.346944E+04
         00  10    1      0.922750E-01
         00  11c   1      0.112700E-01
         10  00    1     -0.120796E+00
         11c 00    1      0.978500E-02
         00  20    2     -0.537020E-01
         00  21c   2     -0.303000E-02
         00  22c   2     -0.124746E+00
      END
      N N rho alpha       C6      C8      C10
```

```
         00  00    0     0.587069E+01    0.168134E+01    0.305702E+02    0.834432E+03
0.165808E+05
         00  10    1      0.104283E+00
         00  11c   1     -0.181740E-01
         10  00    1      0.104283E+00
         11c 00    1     -0.181740E-01
         00  20    2      0.125383E+00
         00  22c   2     -0.163580E+00
         20  00    2      0.125383E+00
         22c 00    2     -0.163580E+00
      END
      N C1 rho alpha       C6      C8      C10
         00  00    0     0.579316E+01    0.174043E+01    0.244635E+02    0.553381E+03
0.223420E+05
         00  10    1      0.130845E+00
         00  11c   1      0.335830E-01
         10  00    1      0.104283E+00
         11c 00    1     -0.181740E-01
         00  20    2     -0.772060E-01
         00  22c   2     -0.113724E+00
         20  00    2      0.125383E+00
         22c 00    2     -0.163580E+00
      END
      N C2 rho alpha       C6      C8      C10
         00  00    0     0.580618E+01    0.178418E+01    0.339916E+02    0.645135E+03
0.214618E+05
         00  10    1      0.120027E+00
         00  11c   1     -0.206020E-01
         10  00    1      0.104283E+00
         11c 00    1     -0.181740E-01
         00  20    2     -0.606040E-01
         00  22c   2     -0.137913E+00
         20  00    2      0.125383E+00
         22c 00    2     -0.163580E+00
      END
      N C3 rho alpha       C6      C8      C10
         00  00    0     0.582263E+01    0.183870E+01    0.239745E+02    0.896858E+03
0.276031E+05
         00  10    1      0.922750E-01
         00  11c   1      0.112700E-01
         10  00    1      0.104283E+00
         11c 00    1     -0.181740E-01
         00  20    2     -0.537020E-01
         00  21c   2     -0.303000E-02
         00  22c   2     -0.124746E+00
         20  00    2      0.125383E+00
         22c 00    2     -0.163580E+00
      END
      C1 C1 rho alpha       C6      C8      C10
         00  00    0     0.573571E+01    0.180125E+01    0.109484E+02    0.375201E+03
0.221470E+05
```

```
         00  10    1      0.130845E+00
         00  11c   1      0.335830E-01
         10  00    1      0.130845E+00
         11c 00    1      0.335830E-01
         00  20    2     -0.772060E-01
         00  22c   2     -0.113724E+00
         20  00    2     -0.772060E-01
         22c 00    2     -0.113724E+00
      END
      C1 C2 rho alpha       C6      C8      C10
         00  00    0     0.584264E+01    0.169205E+01    0.134900E+02    0.378918E+03
0.210985E+05
         00  10    1      0.120027E+00
         00  11c   1     -0.206020E-01
         10  00    1      0.130845E+00
         11c 00    1      0.335830E-01
         00  20    2     -0.606040E-01
         00  22c   2     -0.137913E+00
         20  00    2     -0.772060E-01
         22c 00    2     -0.113724E+00
      END
      C1 C3 rho alpha       C6      C8      C10
         00  00    0     0.579373E+01    0.178477E+01    0.168231E+02    0.613483E+03
0.304373E+05
         00  10    1      0.922750E-01
         00  11c   1      0.112700E-01
         10  00    1      0.130845E+00
         11c 00    1      0.335830E-01
         00  20    2     -0.537020E-01
         00  21c   2     -0.303000E-02
         00  22c   2     -0.124746E+00
         20  00    2     -0.772060E-01
         22c 00    2     -0.113724E+00
      END
      C2 C2 rho alpha       C6      C8      C10
         00  00    0     0.604840E+01    0.164913E+01    0.292550E+02    0.441478E+03
0.303622E+05
         00  10    1      0.120027E+00
         00  11c   1     -0.206020E-01
         10  00    1      0.120027E+00
         11c 00    1     -0.206020E-01
         00  20    2     -0.606040E-01
         00  22c   2     -0.137913E+00
         20  00    2     -0.606040E-01
         22c 00    2     -0.137913E+00
      END
      C2 C3 rho alpha       C6      C8      C10
         00  00    0     0.589174E+01    0.166545E+01    0.249404E+02    0.718721E+03
0.328551E+05
         00  10    1      0.922750E-01
```

```
         00  11c   1      0.112700E-01
         10  00    1      0.120027E+00
         11c 00    1     -0.206020E-01
         00  20    2     -0.537020E-01
         00  21c   2     -0.303000E-02
         00  22c   2     -0.124746E+00
         20  00    2     -0.606040E-01
         22c 00    2     -0.137913E+00
      END
      C3 C3 rho alpha       C6      C8      C10
         00  00    0     0.592659E+01    0.173539E+01    0.235277E+02    0.983470E+03
0.445211E+05
         00  10    1      0.922750E-01
         00  11c   1      0.112700E-01
         10  00    1      0.922750E-01
         11c 00    1      0.112700E-01
         00  20    2     -0.537020E-01
         00  21c   2     -0.303000E-02
         00  22c   2     -0.124746E+00
         20  00    2     -0.537020E-01
         21c 00    2     -0.303000E-02
         22c 00    2     -0.124746E+00
      END
```

Distributed Multipole Moments Input File

```
! Multipole moments for pyridine
! Based on DF-type : ISA

  H1    -3.87454677     2.40829326     0.00000000     Type   H1      Rank   4
       0.061471
       0.000000      -0.050749       0.036166
      -0.033021       0.000000       0.000000      -0.046058       0.050787
       0.000000       0.192664      -0.108250       0.000000       0.000000
                      0.000098       0.062916
      -0.185895       0.000000       0.000000       0.090360      -0.212359
                      0.000000       0.000000      -0.009016       0.017231

  H2    -4.05745524    -2.27382980     0.00000000     Type   H2      Rank   4
       0.139293
       0.000000      -0.035408      -0.018515
      -0.014446       0.000000       0.000000      -0.038619      -0.057252
       0.000000       0.206783       0.106450       0.000000       0.000000
                     -0.013135      -0.058056
      -0.213268       0.000000       0.000000       0.120076       0.161027
                      0.000000       0.000000      -0.001913       0.001330

  H3     0.00000000    -4.70080300     0.00000000     Type   H3      Rank   4
```

```
     0.097072
     0.000000       0.000000      -0.049607
    -0.027536       0.000000       0.000000       0.070457       0.000000
     0.000000      -0.000003       0.231686       0.000000       0.000000
                   -0.000002       0.042166
    -0.205392       0.000000       0.000000      -0.227994       0.000001
                    0.000000       0.000000       0.024111      -0.000007

H4     4.05745524    -2.27382980     0.00000000     Type   H2      Rank   4
     0.139293
     0.000000       0.035413      -0.018518
    -0.014447       0.000000       0.000000      -0.038618       0.057251
     0.000000      -0.206778       0.106446       0.000000       0.000000
                    0.013136      -0.058050
    -0.213274       0.000000       0.000000       0.120080      -0.161035
                    0.000000       0.000000      -0.001911      -0.001315

H5     3.87454677     2.40829326     0.00000000     Type   H1      Rank   4
     0.061469
     0.000000       0.050756       0.036171
    -0.033021       0.000000       0.000000      -0.046058      -0.050784
     0.000000      -0.192662      -0.108244       0.000000       0.000000
                   -0.000097       0.062914
    -0.185896       0.000000       0.000000       0.090361       0.212361
                    0.000000       0.000000      -0.009015      -0.017239

 N     0.00000000     2.61319624     0.00000000     Type    N      Rank   4
    -0.449971
     0.000000       0.000001      -0.244265
     0.045510       0.000000       0.000000       0.316527       0.000011
     0.000000       0.000032      -0.504088       0.000000       0.000000
                    0.000004       1.475723
    -0.382475       0.000000       0.000000       0.396937       0.000017
                    0.000000       0.000000      -0.831627      -0.000005

C1    -2.14372406     1.30476509     0.00000000     Type   C1      Rank   4
     0.228761
     0.000000       0.130985       0.026047
    -0.129155       0.000000       0.000000       0.019084       0.064717
     0.000000       0.408435       0.163515       0.000000       0.000000
                   -0.058246       0.913013
    -1.652743       0.000000       0.000000      -0.054676       0.500720
                    0.000000       0.000000      -0.031698      -0.726274

C2    -2.24974336    -1.31486189     0.00000000     Type   C2      Rank   4
    -0.268306
     0.000000       0.053328       0.000501
    -0.145870       0.000000       0.000000       0.043185       0.136925
     0.000000       0.173628       0.017677       0.000000       0.000000
```

```
                  -0.180072      -0.354243
    -0.910322       0.000000       0.000000       0.042759      -0.104278
                    0.000000       0.000000      -0.107168       0.564049

C3     0.00000000    -2.65300899     0.00000000     Type   C3      Rank   4
     0.030478
     0.000000       0.000002       0.062223
    -0.103883       0.000000       0.000000      -0.037043      -0.000025
     0.000000       0.000072       0.289346       0.000000       0.000000
                    0.000034       0.842877
    -1.365969       0.000000       0.000000       0.368386      -0.000090
                    0.000000       0.000000       1.137529       0.000040

C4     2.24974336    -1.31486189     0.00000000     Type   C2      Rank   4
    -0.268316
     0.000000      -0.053314       0.000497
    -0.145815       0.000000       0.000000       0.043213      -0.136935
     0.000000      -0.173576       0.017727       0.000000       0.000000
                    0.180015      -0.354040
    -0.910243       0.000000       0.000000       0.042858       0.104270
                    0.000000       0.000000      -0.107146      -0.564034

C5     2.14372406     1.30476509     0.00000000     Type   C1      Rank   4
     0.228755
     0.000000      -0.130972       0.026056
    -0.129109       0.000000       0.000000       0.019114      -0.064710
     0.000000      -0.408369       0.163502       0.000000       0.000000
                    0.058243       0.912796
    -1.652630       0.000000       0.000000      -0.054614      -0.500771
                    0.000000       0.000000      -0.031678       0.726285
```

Distributed Polarizabilities Input File

```
! Polarizabilities

 H1  H1
  0        0        0        0
  0        2.41744 -0.21103  0
  0       -0.21103  1.30295  0
  0        0        0        2.59121

 H2  H2
  0        0        0        0
  0        2.57775  0.04114  0
  0        0.04114  1.17700  0
  0        0        0        2.69257
```

```
H3  H3
 0        0        0        0
 0        2.27716 -0.01568  0
 0       -0.01568  1.34126  0
 0        0        0        2.70659

H4  H4
 0        0        0        0
 0        2.57775  0.04114  0
 0        0.04114  1.17700  0
 0        0        0        2.69257

H5  H5
 0        0        0        0
 0        2.41744 -0.21103  0
 0       -0.21103  1.30295  0
 0        0        0        2.59121

N  N
 0        0        0        0
 0        6.75163  0.01380  0
 0        0.01380  11.6518  0
 0        0        0        6.22215

C1  C1
 0        0        0        0
 0        9.35259  0.37846  0
 0        0.37846  12.8105  0
 0        0        0        3.87911

C2  C2
 0        0        0        0
 0        8.45343 -0.48931  0
 0       -0.48931  13.1125  0
 0        0        0        4.18607

C3  C3
 0        0        0        0
 0        8.94908 -0.02629  0
 0       -0.02629  12.9489  0
 0        0        0        4.30903

C4  C4
 0        0        0        0
 0        8.45343 -0.48931  0
 0       -0.48931  13.1125  0
 0        0        0        4.18607

C5  C5
```

```
  0        0        0        0
  0        9.35259  0.37846  0
  0        0.37846  12.8105  0
  0        0        0        3.87911
```

Local Axes

```
Axes
  N   z from C3 to N   x from N  to C1
  C1  z from C1 to H1  x from C1 to C2
  H1  z from C1 to H1  x from C1 to C2
  C5  z from C5 to H5  x from C5 to C4
  H5  z from C5 to H5  x from C5 to C4
  C2  z from C2 to H2  x from C2 to C3
  H2  z from C2 to H2  x from C2 to C3
  C4  z from C4 to H4  x from C4 to C3
  H4  z from C4 to H4  x from C4 to C3
  C3  z from C3 to H3  x from C3 to C4
  H3  z from C3 to H3  x from C3 to C4
End
```

Appendix 3.B – Crystal Structure Data

1. *Form III prediction*

Analysis of the relative stabilities after re-minimizing the PES under $2GPa$. The structural details of 87 structures re-minimized can be found in the Appendix CD.

2. *Relative stabilities*

Relative stabilities of hypothetical and observed *xminopt* structures from the DIFF CSP of pyridine. The full rankings can be found in the Appendix CD.

FORM III PREDICTION: CALCULATIONS UNDER PRESSURE with DIFF and FIT+GDMA

Structure ($_{pyr}$#)	Lattice Energy	Volume	Density	Space-Group	a	b	c	α	β	γ	# molecules	RMSD$_{15}$
2	43.19	350.68	1.50	Pna21	6.34	10.49	5.27	90.0	90.0	90.0	1 out of 15	0.00
27	43.61	352.02	1.49	P21/c	5.17	10.19	7.13	90.0	69.5	90.0	4 out of 15	0.93
FORM II	44.39	359.42	1.46	P212121	6.52	5.14	10.71	90.0	90.0	90.0	15 out of 15	0.00
8	44.63	708.64	1.48	Pbca	9.21	8.26	9.31	90.0	90.0	90.0	2 out of 15	0.71
6	45.18	353.63	1.49	P21/c	6.39	5.31	10.42	90.0	90.0	90.0	1 out of 15	0.00
9	45.40	355.70	1.48	Pna21	6.82	5.50	9.49	90.0	90.0	90.0	1 out of 15	0.00
978	45.43	710.12	1.48	C2/c	6.53	12.64	9.27	90.0	68.2	90.0	3 out of 15	0.93
35 ≈ FORM III	45.48	356.07	1.48	P212121	12.84	5.26	5.27	90.0	90.0	90.0	7 out of 15	0.38
20	45.56	352.19	1.49	P21/c	6.89	5.46	10.44	90.0	63.9	90.0	3 out of 15	0.34
424	45.70	1407.73	1.49	Fdd2	8.63	18.74	8.70	90.0	90.0	90.0	1 out of 15	0.00
2039	45.87	702.80	1.50	C2/c	8.67	8.71	10.11	90.0	66.9	90.0	1 out of 15	0.00
7	46.32	355.94	1.48	Pca21	10.35	5.34	6.44	90.0	90.0	90.0	1 out of 15	0.00
307	46.55	347.06	1.51	P21/c	9.91	5.36	6.84	90.0	72.5	90.0	1 out of 15	0.00
18	46.75	356.29	1.47	Pna21	8.48	8.02	5.24	90.0	90.0	90.0	5 out of 15	0.28
72	46.85	352.95	1.49	Pna21	6.70	5.34	9.87	90.0	90.0	90.0	3 out of 15	0.23
124	46.88	349.75	1.50	P21/c	6.33	5.88	10.47	90.0	64.0	90.0	3 out of 15	0.53
256	47.09	701.51	1.50	Pbca	9.28	9.37	8.07	90.0	90.0	90.0	2 out of 15	0.74
104	47.10	349.70	1.50	Pna21	6.78	10.17	5.07	90.0	90.0	90.0	3 out of 15	0.16
37	47.11	355.78	1.48	P21/c	5.50	7.83	9.55	90.0	59.9	90.0	3 out of 15	0.37
75	47.12	354.73	1.48	Pca21	6.56	5.39	10.02	90.0	90.0	90.0	3 out of 15	0.27

Table 10: *Comparison of lattice energies* ($in\ kJ\ mol^{-1}$)*, crystal structures in the 20 lowest energy structures (in order of stability) and the experimental form II for the non-empirical DIFF potential under* $2\ GPa$*. These stable structures were analyzed in detail for their similarities to the form II (*$h5$*-II PYRDNA05)*[186] *under* $2\ GPa$ *of pressure.*

Structure (pyr#)	Energy	Volume	Density	Space-Group	a	b	c	α	β	γ	# molecules	RMSD$_{15}$
2183	75.22	401.31	1.31	P21212	6.85	10.88	5.39	90.00	90.00	90.00	5 out of 15	0.43
37	75.31	394.48	1.33	P21/c	5.89	7.00	11.29	90.00	57.97	90.00	3 out of 15	0.52
32	75.32	394.50	1.33	P21/c	5.89	7.00	11.18	90.00	58.85	90.00	3 out of 15	0.52
34	75.93	400.99	1.31	P41	7.05	5.30	10.74	90.00	90.00	90.00	15 out of 15	0.00
3	75.94	401.69	1.31	P212121	10.24	5.47	7.17	90.00	90.00	90.00	1 out of 15	0.00
FORM II	75.95	401.00	1.31	P212121	7.05	5.30	10.74	90.00	90.00	90.00	15 out of 15	0.00
130	75.95	401.71	1.31	Pna21	10.24	7.17	5.47	90.00	90.00	90.00	1 out of 15	0.00
3506	75.96	401.01	1.31	P212121	7.05	10.75	5.30	90.00	90.00	90.00	15 out of 15	0.00
7534	75.98	401.74	1.31	P212121	7.17	5.47	10.24	90.00	90.00	90.00	1 out of 15	0.00
20	76.18	400.03	1.31	P21/c	7.34	5.62	10.74	90.00	64.41	90.00	3 out of 15	0.86
90	76.18	400.03	1.31	P21/c	7.34	5.63	10.74	90.00	64.41	90.00	3 out of 15	0.86
27	76.36	397.81	1.32	P21/c	5.48	10.96	7.14	90.00	67.92	90.00	4 out of 15	1.01
14	76.61	197.51	1.33	P21	7.05	5.53	5.44	90.00	68.79	90.00	3 out of 15	0.26
74	76.70	395.16	1.33	P21212	5.53	10.60	7.04	73.04	90.00	90.00	3 out of 15	0.27
5981	76.75	792.80	1.33	Aba2	10.88	10.67	7.23	109.11	90.00	90.00	4 out of 15	1.04
57	76.76	792.83	1.33	Aba2	10.88	10.67	7.23	70.88	90.00	90.00	4 out of 15	1.04
52	76.78	393.69	1.33	P21/c	5.98	7.13	10.94	90.00	57.58	90.00	4 out of 15	0.90
49	77.00	397.34	1.32	P21/c	5.37	9.15	8.94	90.00	64.69	90.00	1 out of 15	0.00
9552	77.15	794.82	1.32	P43212	14.34	10.24	5.41	90.00	90.00	90.00	3 out of 15	0.41
961	77.16	794.86	1.32	P43212	10.24	14.35	5.41	90.00	90.00	90.00	3 out of 15	0.41
8	77.70	795.94	1.32	Pbca	9.34	8.80	9.68	90.00	90.00	90.00	2 out of 15	0.90
863	78.06	801.29	1.31	P43212	5.58	10.29	13.96	90.00	90.00	90.00	3 out of 15	0.25
24	78.09	399.34	1.32	P21/c	6.94	5.61	11.15	90.00	67.02	90.00	1 out of 15	0.00
1854	78.12	1602.04	1.31	I41/a	7.07	10.98	20.64	90.00	90.00	90.04	1 out of 15	0.00
18	78.17	402.41	1.31	Pna21	8.77	8.35	5.49	90.00	90.00	90.00	5 out of 15	0.28
39	78.62	401.82	1.31	P21/c	10.37	5.37	7.48	90.00	74.73	90.00	1 out of 15	0.00
33	78.77	804.27	1.31	Pbca	5.40	10.87	13.69	90.00	90.00	90.00	2 out of 15	0.28
1144	78.77	804.27	1.31	Pnma	10.87	13.69	5.40	90.00	90.00	90.00	2 out of 15	0.28
29	78.83	402.17	1.31	P21/c	5.72	6.79	10.35	90.00	90.45	90.00	1 out of 15	0.00
46	78.88	400.94	1.31	Pca21	10.47	6.85	5.59	90.00	90.00	90.00	1 out of 15	0.00
1963	78.90	801.93	1.31	Cmc21	17.24	10.47	5.59	90.00	90.00	127.40	1 out of 15	0.00
2	79.05	401.57	1.31	Pna21	6.87	10.59	5.52	90.00	90.00	90.00	3 out of 15	0.80
533	79.09	805.84	1.30	Pccn	5.52	10.29	14.17	90.00	90.00	90.00	1 out of 15	0.00
139	79.17	199.77	1.32	P-1	6.68	5.71	6.36	77.27	65.59	65.01	4 out of 15	1.12
35	79.26	403.18	1.30	P212121	13.27	5.51	5.51	90.00	90.00	90.00	7 out of 15	0.46

Table 11: *Comparison of Lattice Energies, Crystal structures in the 35 lowest energy structures (in order of stability) and the experimental form II for the empirical FIT+GDMA potential at a pressure of 2 GPa. These stable structures were analyzed in detail for their similarities to the experimental form II (**h5**-II PYRDNA05)[186] modelled at 2 GPa of pressure using the same methods as above but the empirical FIT+GDMA potential.*

STABILITY OF OBSERVED VERSUS HYPOTHETICAL STRUCTURES

Structure ($_{\text{pyr}}$#)	Lattice	Volume	Density	Space-Group	a	b	c	α	β	γ	Dimers
2	-68.24	394.05	1.3333	Pna21	6.69	10.86	5.42	90.00	90.00	90.00	T1
594	-67.46	798.12	1.3166	C2/c	7.13	12.94	9.43	90.00	66.63	90.00	Hb1
58	-67.30	800.42	1.3128	Pbca	6.82	10.43	11.26	90.00	90.00	90.00	Hb1
1 ≈ FORM II	-66.96	401.53	1.3085	P212121	6.72	5.38	11.10	90.00	90.00	90.00	T2, Hb3
978	-66.94	798.03	1.3167	C2/c	7.12	12.94	9.42	90.00	66.83	90.00	Hb1
FORM II	-66.76	401.69	1.3080	P212121	5.38	6.71	11.13	90.00	90.00	90.00	T2, Hb3
12	-66.72	398.76	1.3176	Pca21	6.99	5.59	10.21	90.00	90.00	90.00	Hb2
9	-66.22	399.98	1.3136	Pna21	7.25	5.62	9.82	90.00	90.00	90.00	Hb2
20	-66.07	394.19	1.3328	P21/c	7.10	5.68	10.76	90.00	65.31	90.00	S1, T1
428	-65.85	1616.93	1.2997	Fdd2	12.86	17.30	7.27	90.00	90.00	90.00	Hb1
546	-65.81	826.30	1.2717	P-421c	12.26	12.26	5.50	90.00	90.00	90.00	T2, Hb3
6	-65.73	398.06	1.3199	P21/c	6.68	5.55	10.73	90.00	90.01	90.00	T1
88	-65.66	1637.03	1.2838	Fdd2	10.23	13.93	11.49	90.00	90.00	90.00	None
597	-65.56	393.99	1.3335	P21/c	7.12	5.69	10.73	90.00	114.87	90.00	S1, T1
44	-65.48	808.53	1.2996	Aba2	13.30	10.99	5.53	90.00	90.00	90.00	None
2039	-65.45	783.20	1.3417	C2/c	9.04	9.04	10.53	90.00	65.56	90.00	S1
7	-65.36	401.19	1.3096	Pca21	10.65	5.59	6.74	90.00	90.00	90.00	T1
2018	-65.34	806.96	1.3022	Aba2	13.66	10.81	5.47	90.00	90.00	90.00	None
559	-65.32	790.11	1.3299	Pbca	8.96	10.10	8.73	90.00	90.00	90.00	S1
27	-65.29	394.77	1.3309	P21/c	5.42	10.36	7.64	90.00	67.12	90.00	Hb1, Hb2, S1
4	-65.17	409.58	1.2828	P212121	5.54	11.14	6.64	90.00	90.00	90.00	T1
35 ≈ FORM III	-65.16	399.37	1.3156	P212121	13.19	5.50	5.51	90.00	90.00	90.00	None
8	-65.12	792.47	1.326	Pbca	9.54	8.63	9.63	90.00	90.00	90.00	Hb3, S1
91	-65.07	1646.10	1.2767	Fdd2	14.23	12.20	9.49	90.00	90.00	90.00	bT
437	-65.07	800.10	1.3133	Pbca	6.64	12.72	9.48	90.00	90.00	90.00	Hb1, Hb3, S1, T2
917	-65.04	408.51	1.2861	P21/c	8.59	8.22	5.94	90.00	77.17	90.00	Hb1, S1
37	-64.97	399.02	1.3167	P21/c	5.64	8.47	9.63	90.00	60.17	90.00	Hb1, S1
FORM I	-64.94	406.97	1.2910	Pna21	17.07	8.67	11.00	90.00	90.00	90.00	T2, bT, Hb3
85	-64.81	811.96	1.2941	PBCN	10.21	11.28	7.05	90.00	90.00	90.00	Hb1
41	-64.77	401.80	1.3076	Pna21	7.08	8.86	6.40	90.00	90.00	90.00	Hb2
18	-64.69	400.10	1.3131	Pna21	9.10	8.14	5.40	90.00	90.00	90.00	None
162	-64.56	408.71	1.2855	P21/c	5.86	8.26	9.37	90.00	64.36	90.00	Hb1, S1

***Table 12:** Comparison of lattice energies (in $kJ\ mol^{-1}$), crystal structures and the occurrence of the gas phase dimer structures (**Figure 13**) in the 30 lowest energy structures and the experimental forms I and II (h5-I PYRDNA04, h5-II PYRDNA05)[186] for the DIFF potential. These stable structures were analyzed in detail for their similarities to each other and the isolated dimer structures.*

4 The influence of nitro-group conformations and charge distributions on organic energetics[168]

4.1 Introduction and overview

The rigid molecule approximation was successful in determining the polymorphs of pyridine (**Chapter 3**), but how well does this approximation extend to more flexible molecules like energetic materials? What is the influence of conformation on the electrostatic properties of organic energetics? This chapter investigates the influence of variations in the charge distribution of NO_2 groups within the crystalline polymorphs of energetic materials. Previously suggested correlations are used to see how this may affect their explosive properties. The Iterated Stockholder Atoms (ISA) partitioning of high quality (PBE0/aug-cc-p-VTZ) charge distributions is used to examine the approximations that can be made in modelling polymorphs and their physical properties, using RDX, TNT, TNB and HNB as exemplars. This is done by examining the variations in the NO_2 charge distribution depending on the neighbouring atoms, the rest of the molecule, and also the NO_2 torsion angle within the possible variations found in observed crystal structures. Proposed correlations between the molecular electrostatic properties, such as trigger-bond potential or maxima in the electrostatic potential, and impact sensitivity, originally proposed for comparisons between different molecules, are used to determine the influence of the changes in conformation that occur on crystallization. This requires investigating the variations in NO_2 torsion observed in experimental crystal structures by examining the Cambridge Structural Database (CSD), and determining how this is linked to the conformational energy, and the charge and dipole magnitude on each atom. The effect of changes in charge density with conformation can be examined by comparing the electrostatic potential around the molecule and the lattice energy. These are calculated using the distributed multipoles calculated for the crystalline conformation and using multipole moments analytically rotated from the optimized molecular structure. Assuming that the charge distribution is not specific to the conformation appears adequate for proposing plausible crystal structures in a CSP study but too inaccurate for modelling the relative lattice energies. This establishes the basis of transferability for realistic non-empirical model intermolecular potentials for simulating energetic materials. One such potential is developed in **Chapter 5** for TNB, where the use of a rigid conformation for CSP seems a sensible approximation (**Chapter 6**). The implications of the transferability of the NO_2 charge distribution for future modelling energetic materials are further discussed in **Chapter 7**.

4.1.1 Energetic properties

Energetic materials (**1.4**), which decompose explosively under various external stimuli, are intrinsically difficult to study experimentally, and hence the design of new materials with desirable property combinations such as low sensitivity but high detonation performance can benefit from accurate molecular modelling.[137, 138, 322] Explosive properties are sensitive to the arrangement of molecules in the crystal, and hence to the polymorph formed.[6, 137, 236, 323] Consequently, Crystal Structure Prediction (CSP) methods may be used to determine whether an energetic molecule can crystallize in a dense structure with desirable energetic properties, helping to focus synthetic efforts. There have been extensive quantum mechanical studies[324] seeking to find correlations with the experimental impact sensitivity of the materials[137, 138, 146, 237, 325-333] but these are on the intrinsic sensitivity or stability of the isolated molecule as opposed to the measured sensitivity of the condensed phase.[324] It has been recognized that the crystal structure will also affect energetic properties, as the intermolecular interactions in the solid are linked to dissociation processes essential to detonation, such as defect formation. Correlations have been found between the heat of fusion of the crystal and the N-N bond dissociation energy[334] or crystalline void space,[335] but the prediction of impact sensitivity and other processes is clearly a complex combination of molecular and crystalline properties.[336-338] In particular, the crystalline conformation can differ from the isolated molecule conformation, vary between polymorphs and with temperature and pressure. The changes in NO_2-torsion angles that can occur on crystallization are examined and their effect on the molecular charge distribution and associated electrostatic properties.

Computing energetic material properties requires accurately modelling regions of the repulsive wall that are not adequately sampled by empirical intermolecular potentials that have been fitted to crystals at ambient conditions. Thus developing non-empirical methods for modelling the intermolecular forces between energetic molecules, suitable for molecular dynamics (MD) simulations, has been the subject of a considerable body of recent research.[22, 339, 340] However, such methodologies have usually treated the molecules as rigid. A distributed multipole model of the molecular charge distribution provides a method of both examining the transferability of the local charge distribution as well as a model for calculating the electrostatic contribution to the intermolecular interaction energy. This chapter only examines the conformation dependence of the electrostatic term. However, other contributions to the intermolecular energy, such as the polarization, dispersion and anisotropic short-range repulsive terms are very closely related to the molecular charge density, and therefore are expected to show similar transferability properties. The electrostatic term (**2.5**) tends to be the most dominant, orientation-dependent determinant of intermolecular interactions between molecules in van der Waals contact. It is arguably[341] the dominant contribution in interactions which are often described by crystal engineers as hydrogen bonding, $\pi \cdots \pi$

stacking, $N\cdots O$ interactions, etc. Accurate atom-atom non-empirical models for intermolecular forces derived from the molecular charge density have been isotropic and tended to focus on rigid molecules,[164] (**1.4.2.2**) because of the difficulty of modelling the changes in charge distribution with conformation.

4.1.2 The influence of conformation on impact sensitivity

There are currently only a few reliable methods for predicting the detonation pressures and velocities of CHNO energetics,[137] and there has been less progress made towards predicting the sensitivity of an explosive compound to the different stimuli that can initiate detonation.[342] This reflects the intricacies of detonation, for example the explosive molecule LLM-119 is sensitive to impact but moderately insensitive to friction and electrical sparks.[343] A widely studied and important measure of sensitivity is impact sensitivity, which is the vulnerability of the material to explosion after sudden compression due to impact.[324] This is a complex process, dominated by the chemistry of the chemical reaction that takes place,[322] chemical bonding and molecular interactions within the crystal. The impact sensitivity of an energetic material is normally determined using a drop-weight test,[344] being inversely proportional to $h_{50\%}$, the height at which 50% of the experiments produce a reaction on dropping a weight onto the material.

The impact sensitivity of a material is heavily dependent on the particle size,[137] the shape and hardness of the crystals, the roughness of its surfaces, the purity and the presence of lattice defects, as well as the intrinsic differences in polymorph packing,[6, 345-347] the surrounding temperature during experiment and, most importantly, the weight used. Hence, the observed impact sensitivities ($h_{50\%}$) for TNT vary from around 100 cm to over 250 cm.[236, 323] The experimental impact sensitivities obtained in 1990 by Wilson and Bliss[323] are used as they are a consistent set of experiments under consistent conditions, for TNB, HNB, and TNT, though these do not have separate values for the different polymorphs. RDX is a comparatively new explosive and its impact sensitivities were measured later.[342, 348]

Correlations between the molecular electrostatic potential (a molecular property) and the impact sensitivity (a crystal property) of CHNO energetic materials were investigated decades ago,[146, 326-329] based on the crude electrostatic models[237, 326, 330-333] available at the time. One approach focuses on the charge distribution at the trigger bonds (the bond likely to break during detonation).[344, 349] The trigger bonds are typically X-NO_2 functional groups, and the strengths of C-NO_2 and N-NO_2 bonds are found to be inversely proportional to the magnitudes of the positive electrostatic potentials in the C-N and N-N inter-nuclear regions (V_{mid}).[237, 331, 350, 351] The electrostatic potential at the longest C-NO_2 bond is often used[237, 333, 352] as calculated by the equation first proposed by Owens in 1985[333]

$$V_{mid} = \frac{Q_C}{0.5R} + \frac{Q_N}{0.5R}$$

92

where R is the longest bond length between C and N in the C-NO_2 bonds, while Q_C and Q_N are the atomic charges. The above equation is based on the assumption the longest X-NO_2 bond will be the first bond to break during the initiation process. The initial correlations between V_{mid} and impact sensitivity computed using the Mulliken point charges[240] (**2.5.1.1**) with a HF/STO-3G level of theory[333] on each atom were only observed within a specific group of poly-nitroaromatics, which at least partly reflects the limitations of the partitioned charges in early ab initio calculations. An alternative is to incorporate the response of the other trigger bonds in the molecule. The averaged V_{mid}

$$V_{mid,avg} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Q_X^i}{0.5R_i} + \frac{Q_N^i}{0.5R_i}\right)$$

93

has been be used[236] to derive empirical correlation models for reproducing impact sensitivities of CHNO explosives. A more general but complex correlation has been suggested between electrostatic potential surface of the isolated molecule and crystal impact sensitivities, with the sensitivity of the explosive to impact being exponentially related to the average positive and negative potential across the molecular electrostatic potential surface[236]

$$h_{50\%} = a_1 + a_2 \exp\left[-\left(a_3\left|\bar{V}_S^+ - |\bar{V}_S^-|\right|\right)\right]$$

94

where a_n are best fit parameters, and $\bar{V}_S^+$ and $\bar{V}_S^-$ are the average positive and negative electrostatic potentials on the iso-surface. This is an example of the more intricate correlations that have been investigated. However, the goal of finding an adequate predictive correlation of impact sensitivity with molecular properties across all chemical families of explosives is still elusive. This could well be because sensitivity to detonation is not limited to a molecular property, but also the crystal environment. Now that CSP could, in principle, be used in the design of new explosives, the likely crystallization environment of a molecule prior to its synthesis is no longer unknown. Given the empirical correlations of molecular properties with impact sensitivity have focused on electrostatic properties yet ignored the conformational change on crystallization, it is appropriate to ask whether these changes in NO_2 torsion angle would affect the impact sensitivity.

The main focus of this chapter is on the electrostatic potential surface maximum (V_{max}) and minimum (V_{min}) as many studies have focused on the correlation of impact sensitivity with these molecular properties.[137, 138, 324, 336, [349-351, 353-356] The electrostatic maxima and minima in molecules have been previously used in a method of estimating the likelihood of co-crystal formation,[357] i.e. as a simple and approximate surrogate for likely relative crystal energies. Analysis of the electrostatic potential of isolated TNT and CL-20 molecules and their hetero/homo-dimers[138] was able to correlate their electrostatic potential maxima (V_{max}), minima (V_{min}) and the range of the minima and maxima (V_{tot}), to justify co-crystal formation and observed sensitivities to impact. Though not explicitly discussed in the paper,[138] it showed that the change in conformation of CL-20 and TNT from pure crystal to co-crystal changes the charge distribution of the isolated molecules and thus the electrostatic potential, which may help rationalise the reduced impact sensitivity of the co-crystal.

In our study, the modelling of the NO_2 group charge distributions and the energetic crystals is aided by the ISA analysis of the molecular charge distributions (**2.5.1.7**).[231, 232] The atomic multipole moments of the nitro groups can be examined for both transferability between molecules and conformations, and used in models for impact sensitivity. The main molecules of focus are the nitro explosives 1-3-5 trinitrobenzene (TNB), hexanitrobenezene (HNB), trinitrotoluene (TNT) and 1,3,5-trinitroperhydro-1,3,5-triazine (RDX) whose structurally characterized polymorphs are shown in ***Figure 20***. From the variation in nitro torsion angles (and axial/equatorial conformation in RDX) between polymorphs it is clear that the crystalline structure has a significant effect on the molecular conformation. This, in turn, will affect the charge distribution by changing the orbital overlaps in π and σ bonding. In contrast to the previous work used in proposing correlations between charge distribution and impact sensitivity, this thesis uses state-of-the art electrostatic models to examine the role of conformational change on the charge distributions of NO_2 groups. High quality electrostatic models are required to reduce the artefacts associated with poor methods of partitioning inaccurate molecular charge distributions. The Iterated Stockholder Atoms (ISA)[231] partitioned distributed multipoles is used to determine whether changes in torsion angle due to crystallization affects some proposed correlations with impact sensitivity. Whether this change is important to the molecular modelling of nitro-energetics is also investigated, specifically for crystal structure prediction (CSP) studies and the development of non-empirical anisotropic atom-atom intermolecular potentials.

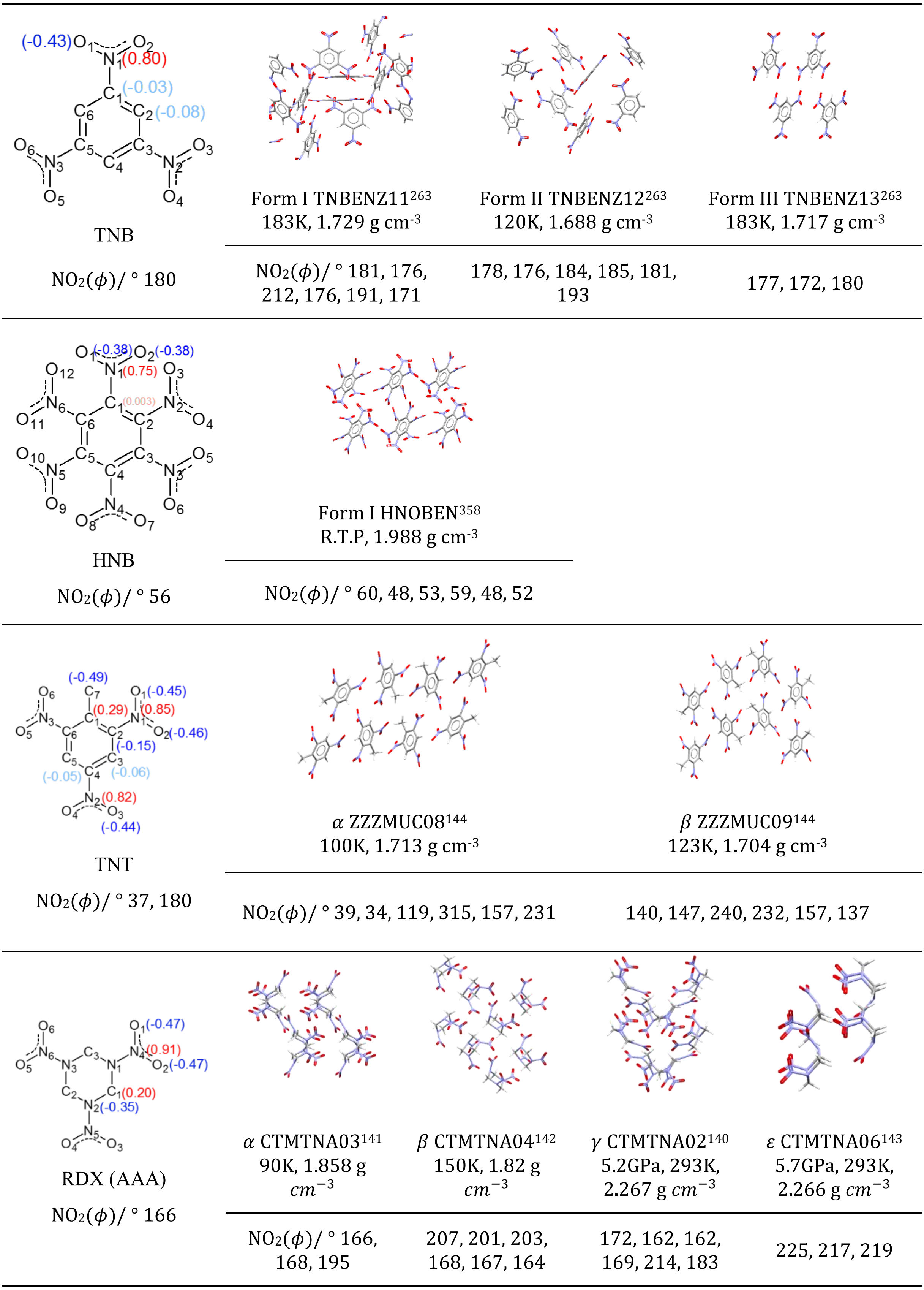


Figure 20: *The crystal structures of the energetic polymorphic crystals TNB*[263]*, HNB*[358]*, TNT*[144] *and RDX*[140-143]*. The form, Cambridge Structural Database (CSD) reference code, determination conditions (at ambient pressure unless stated otherwise), density and NO_2 torsion angles (ϕ°) are given below each crystal structure. These were obtained from their experimental Cambridge Structural Database*[359-361] *entries, and visual representations of the*

structures constructed using CCDC Mercury 3.6.[291, 362] The Iterated Stockholder Atoms (ISA) atomic charges for the optimized structures, computed at the PBE0/aug-cc-pVTZ level, are given in parenthesis in red or blue for positively or negatively charge nuclei. The NO_2 torsion angles for the optimized molecules are given below the structural diagrams.

4.2 Methodology

4.2.1 The electrostatic model

The static isolated molecular structures of RDX, TNT, TNB and HNB were optimized using the Perdew-Burke-Ernzerhof (PBE) general gradient approximation combined with a portion of exact exchange to form the PBE0 functional[203-205] and the aug-cc-pVTZ Dunning basis[214] using the Gaussian 09 program.[235] This level of theory was also used for conformational analysis (unless stated otherwise). In addition, for calculating the molecular charge densities in the experimental crystal conformations, hydrogen atom positions were corrected to standard bond lengths ($C_{aromatic} - H = 1.08$Å, $C_{sp3} - H = 1.06$Å)[363] to account for the systematic error in X-ray structure determinations. The distributed multipoles (Q_{lm}^{i}) of each conformation were calculated in the molecular axis frame (***Figure 21***) and derived using the one of the more recent iterated stockholder atoms algorithms implemented in CamCASP 6.0[107], the ISA-A algorithm.[233, 273] The main basis used an aug-cc-pVTZ basis, the auxiliary basis used Cartesian GTOs of aug-cc-pVTZ size and the atomic basis used an aug-cc-pVQZ basis set. The distributed multipole expansion is evaluated to hexadecapole level (Rank $l = 4$) to include the electrostatic effects due to lone-pair and π atomic anisotropy.

4.2.2 Electrostatic potential surface and structural analysis

Using the various sets of distributed multipole moments, the electrostatic potential was computed and mapped onto an iso-density surface of 10^{-3} $electron/bohr^3$ using CamCASP 6.0,[107] which was visualized on the isolated molecular structures using the ORIENT 4.9.08[108] program. This iso-density surface was chosen as the majority of electrostatic potential studies on energetic materials and their sensitivities use this surface, following the suggestion by Bader[239, 364] as that it can include more than 95% of a molecule's true electronic density. This surface has recently been used to define the molecular volume for use in estimating the density of energetic materials.[365] The electrostatic potential on this iso-density surface is approximately 3Å from the nearest nucleus in the molecule and so gives the electrostatic potential sampled by the non-hydrogenic nuclei that would be in van der Waals contact with the molecule in the crystal. Lattice energies were calculated using a combination of the ISA distributed multipoles, to obtain the electrostatic energy, and the *exp-6* FIT potential, for all other contributions using

DMACRYS.[40] Crystal structure analysis and visual representations used the analysis software of the Cambridge Crystallographic Data Centre, mainly Mercury 3.6.[291, 362] Root Mean Square Deviation ($RMSD_n$) calculations, the optimum overlay of n ($n \leq 15$ for crystal structures, $n = 1$ for molecular conformation comparisons) molecules in two crystal structures excluding the hydrogen atoms, were also done using Mercury.

4.2.3 Surveying the Cambridge Structural Database

To test the transferability of the multipole moments with changes in the NO_2 torsion angle, Z matrices of the structures in ***Figure 20*** were created and the NO_2 functional groups were rotated through 180° in 20° increments, while the rest of the molecule was held rigid at the optimized structure. The extent to which the nitro-group torsion angles can change in the condensed phase was investigated using histograms of the frequency a $O{-}N{-}X{-}C$ torsion angle (rounded to nearest degree) occurred in the relevant organic crystal structures found in the Cambridge Structural Database (CSD).[366] The histograms were constructed for the four unique chemical environments found in the four energetic molecules, with the most commonly observed environment being a $C{-}NO_2$ on an aromatic ring with two adjacent hydrogen atoms (***Table 13***).

Distribution curves illustrating the occurrence of each $O{-}N{-}X{-}C$ torsion in a specific chemical environment (for example, an NO_2 adjacent to two hydrogens or a methyl group in an aromatic ring) were generated using an in-house python program[366] for determining the possible torsion angles that could be adopted in crystal structures based on the torsion angle statistics of all 870,000+ molecules in the Cambridge Structural Database (CSD).[366] The specific NO_2 fragments used are as follows

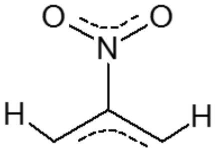
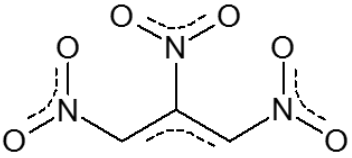
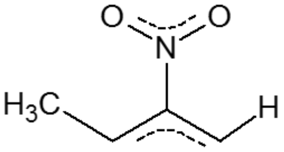
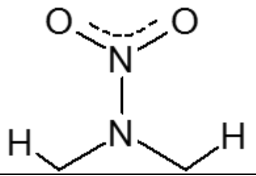
			
aromatic C-NO_2 adjacent hydrogens (TNB, para-NO_2 TNT)	aromatic C-NO_2 adjacent NO_2 (HNB)	aromatic C-NO_2 adjacent a hydrogen and a methyl group (ortho-NO_2 TNT)	aliphatic N-NO_2 adjacent hydrogens (RDX)

Table 13: *The NO_2 fragments that were searched for within the CSD. The occurrence of each torsion angle (rounded to the nearest degree) was plotted in the conformational histograms (**Figure 25** &* **Appendix 4.A** ***Figure 27****).*

4.2.4 Analytical Rotation

The ORIENT[108] program can transform the ISA[231, 233] distributed multipoles of a molecule (originally calculated in the molecule-fixed axis) into the atomic local axes of the molecule as defined in ***Figure 21***. This is done by analytically rotating the global multipoles into the local axis frame using Equation 95. $Q_{lk}^{(L)}$ is the multipole moment in the local axis, $Q_{lm}^{(G)}$ the multipole moment in the molecule-fixed axis and $D_{mk}^{l}(\Omega)$ the transformation tensor used to analytically rotate the multipoles. The molecular geometry and the user defined local axes must first be defined before the multipole moments calculated in the global axis are then analytically rotated into this user defined local axis.

$$Q_{lk}^{(L)} = \sum_{m} Q_{lm}^{(G)} D_{mk}^{l}(\Omega) \tag{95}$$

4.2.4.1 Method

The ISA distributed multipoles can be analytically rotated from the molecular axes into the atomic local axes of the molecule, with the z-axis along the X-N plane (along the bond) and the x-axis in to X-N=O plane (perpendicular to the z-axis but in the plane of the molecule) (***Figure 21***). Analytically rotating the local axis-defined multipoles to a different X-NO_2 torsion angle provides an electrostatic model that includes the geometric effects of conformational change but does not explicitly account for changes in the molecular charge distribution. Provided that the non-NO_2 torsion angles, bond lengths and angles are identical, it is possible to analytically rotate the multipoles to represent the geometric change in torsion angles, using ORIENT 4.9.08;[108] a methodology previously used to study the transferability of electrostatic models between polypeptides.[367]

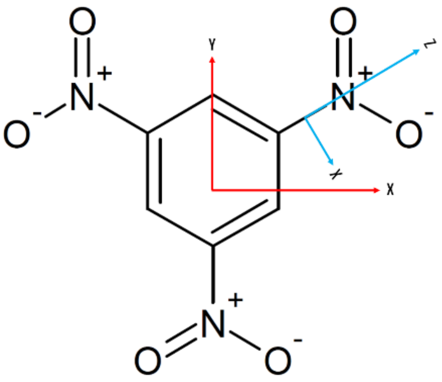


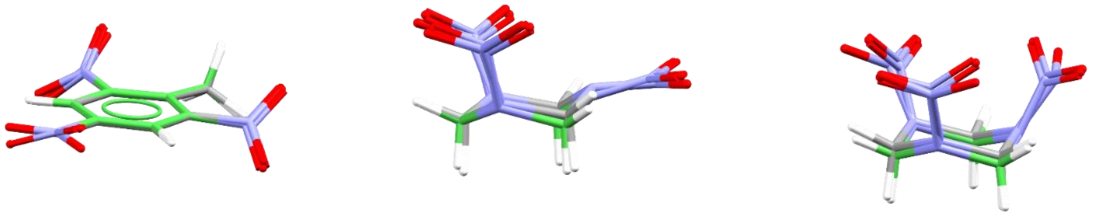
Figure 21: *An example of the local atomic axis (blue) definitions used for comparing and analytically rotating the NO_2 multipole moments from the molecule fixed axis (red) using ORIENT to be used in lattice energy minimizations.*

In order to focus solely on the effects of changing NO_2 torsion angles and compare experimentally observed and optimized conformations, the experimentally observed NO_2 torsion angles were used while all the bond lengths and angles and non-NO_2 torsion angles were kept rigid at their optimized values. These structures and charge distributions calculated in this conformation are referred to as the *optexptNO₂*. In addition, multipole moments calculated in the optimized conformation were analytically rotated into the *optexptNO₂* conformation. Charge distributions obtained this way are referred to as *anarot* (analytically-rotated). The electrostatic properties obtained from the analytically rotated distributed multipoles can then be contrasted with those acquired from calculating the charge distribution in the *optexptNO₂* conformation, which also accounts for the rearrangement of charge within the molecule, e.g. from changes in π conjugation. Analytical rotation can only be applied to the nitro-aromatic molecules, as the aromatic ring is rigid, and undergoes negligible changes when the molecules are optimized and the nitro-group rotate. Only the NO_2 torsion angles change between the optimized and observed conformations (***Figure 22***). However, for nitramines, like RDX, there is a notable difference in the conformation of the aliphatic ring in the experimental crystalline conformers and the optimized conformation. The difference results in a change in non-NO_2 torsion angles upon optimization. Thus, changing only the NO_2 torsion angle in the optimized structure using the standard Z-matrix definition (the *optexptNO₂* methodology) gives a structure that is different from the experimental structure. Hence, it is not possible, let alone appropriate, to test transferability by analytically rotating the multipoles of RDX using the *anarot* method (***Figure 22*, Appendix 4.B *Table 17***).

4.2.4.2 Differences in optimized and experimental conformations of RDX contrasted with TNT

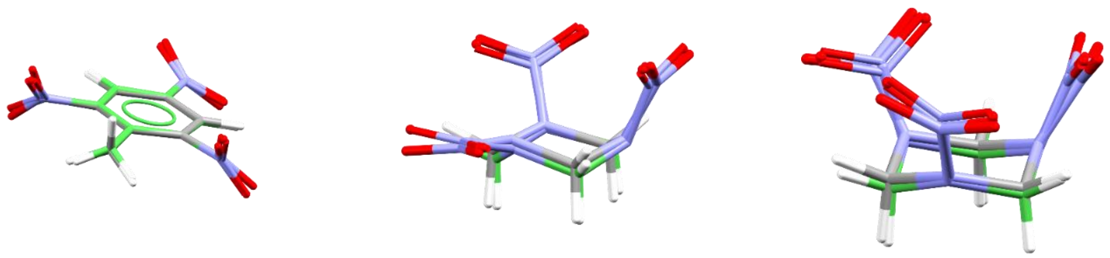
Overlays of experimental conformations, the optimized conformations, and the conformations with only the NO_2 angles rotated (*optexptNO₂*) are shown in ***Figure 22***, to illustrate the changes that accompany nitro-group rotation. In the AAE conformation of RDX, the equatorial NO_2 moves more into the plane of the ring, as the aliphatic ring is more open for the isolated molecule, possibly due to repulsion from lone pairs on the ring sp^3 nitrogen atoms, that are not being counteracted by crystal packing forces. A similar change in the aliphatic ring conformation and nitrogen atoms in the axial NO_2 groups is seen for the AAA conformation, therefore, it is not possible to consistently analytically rotate the nitro-group multipole moments. While these changes may not look severe, one can see by a comparison of differences in cell parameters, $RMSD_{15}$ values and lattice energies between using the experimental and optimized molecular conformations in **Appendix 4.B *Table 17*** that these differences result in very different minima. In contrast, the aromatic ring in the optimized and experimental conformations of the nitro-aromatic TNT is unchanged, with only small methyl group rotation (***Figure 22***), meaning analytical rotation is feasible. **Appendix 4.B** shows that although there is a significant difference in the nitro groups, which does have a major effect on the crystal packing, changing the nitro- torsion angles to the experimental values can result in a good reproduction of the structure.

optimized (grey) vs experimental (green)



TNT, $RMSD_1 = 0.199$ Å   RDX (AAE), $RMSD_1 = 0.182$ Å   RDX (AAA), $RMSD_1 = 0.379$ Å

optexptNO₂ (grey) vs experimental (green)



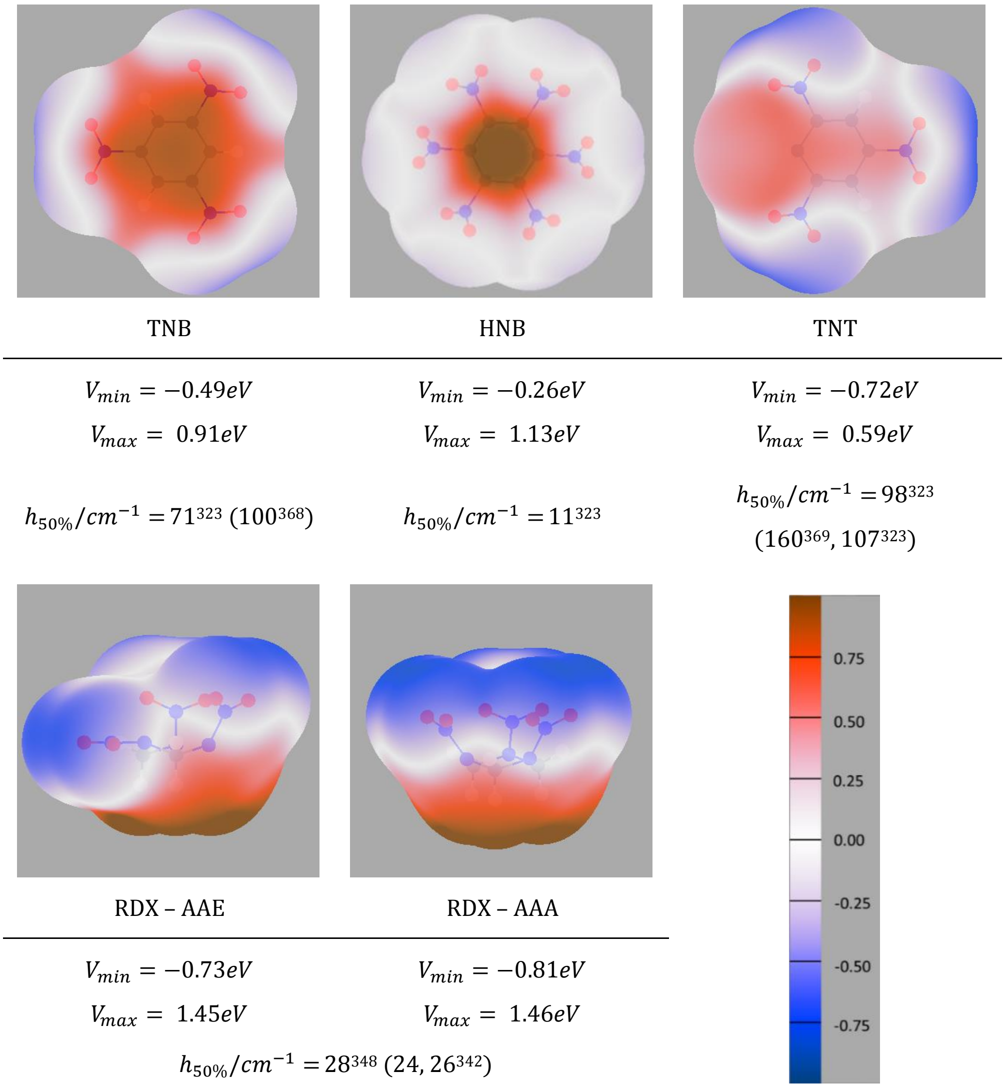
TNT, $RMSD_1 = 0.152$ Å   RDX (AAE), $RMSD_1 = 0.185$ Å   RDX (AAA), $RMSD_1 = 0.287$ Å

Figure 22: *Overlay of experimental crystalline conformation (green) of TNT form I (molecule 1/ZZZMUC08*[144]*), RDX* α *(AAE/CTMTNA03*[141]*) and* ε *(AAA/CTMTNA06*[143]*) with their PBE0/aug-cc-pTVZ optimized structures (green). The row below is the overlays of the experimental conformations (green) and the optimized conformations with experimental* NO_2 *torsion angles, optexptNO*$_2$ *(grey).*

4.3 Results

4.3.1 Molecular electrostatic properties and correlation with impact sensitivity

The electrostatic potential around the static optimized conformations of the four energetic molecules (***Figure 23***) clearly shows that the electrostatic term will greatly influence intermolecular interactions, hence explosive properties. Potential extrema (V_{max} and V_{min}) are influenced by molecular symmetry and not necessarily centralized around the NO_2 groups, in contrast to the measure of the local NO_2 charge distribution V_{mid}. The AAE and AAA crystalline conformations of RDX optimize to very different geometries, therefore, are conformational polymorphs.[147] The AAE conformation, with one equatorial NO_2 group, is the lower symmetry, higher energy conformation of RDX and occurs in the most stable α polymorph and one of the two independent molecules in the high pressure γ form. It has a different surface shape and potential minimum (V_{min}) but the potential maximum for this bowl-shaped molecule is virtually unaffected. The nitramine RDX has a significantly larger gradient in the electrostatic potential around the molecule than any of the flatter nitro-aromatics.



TNB	HNB	TNT
$V_{min} = -0.49 eV$	$V_{min} = -0.26 eV$	$V_{min} = -0.72 eV$
$V_{max} = 0.91 eV$	$V_{max} = 1.13 eV$	$V_{max} = 0.59 eV$
$h_{50\%}/cm^{-1} = 71^{323}\ (100^{368})$	$h_{50\%}/cm^{-1} = 11^{323}$	$h_{50\%}/cm^{-1} = 98^{323}$ $(160^{369}, 107^{323})$

RDX – AAE	RDX – AAA
$V_{min} = -0.73 eV$	$V_{min} = -0.81 eV$
$V_{max} = 1.45 eV$	$V_{max} = 1.46 eV$

$h_{50\%}/cm^{-1} = 28^{348}\ (24, 26^{342})$

Figure 23: *The electrostatic potential (eV) computed using the ISA distributed multipole analysis (ISA-DMA) on the iso-density surface of 10*$^{-3}$ $electron/bohr^3$ *around the static optimized isolated molecular structures of TNB, HNB, TNT and RDX. Note that for some of the above molecules the maximum,* V_{max} *is greater than the potential scale, which is +1 eV (red) to* -1 *(blue) eV.*

The variations in charges on the X and N atoms in the X-NO_2 groups of the optimized conformations are contrasted with X-NO_2 groups in the experimentally observed condensed phase conformations that have the most unique structures (***Table 14***). This clearly shows that the nitrogen atoms in the N-NO_2 groups in RDX are more positively charged ($0.882 - 0.945\ e$) than any nitrogen in the aromatic C-NO_2 groups ($0.749 - 0.852\ e$) and even within the C-NO_2 groups, the charge distributions are far from transferable (as seen in the atomic charges in ***Figure 20***). Changes in conformation result in variations in the trigger-bond parameter V_{mid} (both longest and average) that can be comparable to the conformational differences between molecules (***Table 14***). The values of V_{mid} differ more between the molecules, but also vary between the different crystalline conformations.

Molecule	TNB (TNBENZ)			HNB	TNT			RDX			
Polymorph	13 (III)	11 (I)	12 (II)	HNOBEN	ZZZMUC08 (α)			CTMTNA03 (α)		CTMTNA06 (ε)	
Molecule	Opt	Mol 1	Mol 2	Opt	Opt	Mol 1	Mol 2	Opt	Expt	Opt	Expt
$Q_X min$ $Q_X max$ $/eV$	−0.0289 −0.0278	−0.0145 −0.00378	−0.0213 −0.0203	0.000305 0.00266	−0.154 −0.0481	−0.0938 0.0303	−0.0873 0.0105	−0.352 −0.329	−0.350 −0.318	−0.355 −0.353	−0.383 −0.373
$Q_N min$ $Q_N max$ $/eV$	0.797 0.799	0.796 0.810	0.796 0.798	0.749 0.750	0.820 0.852	0.774 0.811	0.777 0.811	0.906 0.943	0.882 0.902	0.907 0.908	0.928 0.945
$V_{mid}/eV^{\dagger}$	1.044	1.093	1.054	1.016	0.948	0.999	0.981	0.819	0.759	0.789	0.783
$V_{mid}/eV^{\ddagger}$	1.045	1.093	1.054	1.016	1.051	1.098	1.074	0.852	0.857	0.792	0.836
$V_{mid,avg}$ $/eV$	1.044	1.071	1.052	1.016	0.982	1.021	1.007	0.828	0.792	0.790	0.816
V_{min}/eV	−0.49	−0.60	−0.49	−0.26	−0.72	−0.79	−0.74	−0.73	−0.84	−0.81	−1.00
V_{max}/eV	0.91	0.76	0.91	1.13	0.59	0.59	0.61	1.45	1.63	1.46	1.57
$h_{50\%}/cm^{-1}$	71[323] (100[368])			11[323]	98[323] (160[369], 107[323])			28[348] (24, 26[342])			

Table 14: *A comparison of electrostatic properties of the 4 explosives and experimental impact sensitivities ($h_{50\%}$). The optimized isolated structures and the most stable experimental conformations within the polymorphs were used. In RDX this is the AAE conformation in CTMTNA03, and the AAA conformation in CTMTNA06. The range of Q_X (X is either C or N, the atom connected to the nitro-group) and Q_N is the min and max values from the 3NO₂ groups (6 for HNB). The impact sensitivity used is given with other impact sensitivity determinations in brackets; for RDX the polymorph used is not specified, thus it is probable that the impact sensitivity of the ε form is very different from literature values, if the experiments used the most stable α form. $^{\dagger}V_{mid}$ of the longest bond. ‡The trigger bond with the most positive V_{mid}.*

In comparing the molecular electrostatic properties V_{max} and V_{mid} against experimental impact sensitivity (***Figure 24***), one observes that conformational changes due to crystallization does affect these properties. There is a notable spread in V_{max} and V_{mid} of the optimized and observed conformations in the nitrobenzenes TNB and TNT. This will significantly affect any correlation with impact sensitivity for the nitrobenzenes. Furthermore, for RDX, the spread of values for either V_{max} and V_{mid} is not substantially larger than for TNB despite RDX having much larger conformational changes (***Figure 20***, ***Figure 22***).

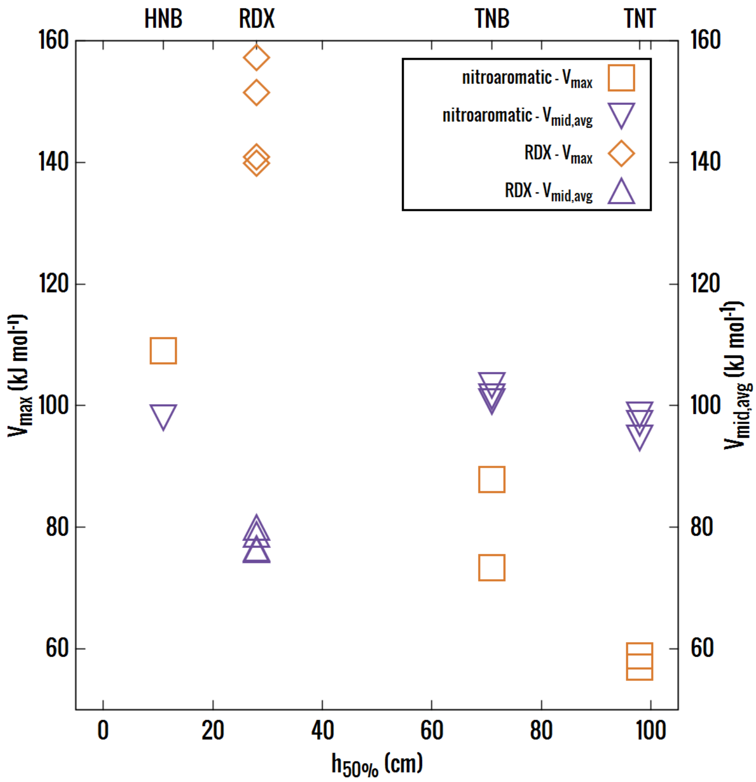


Figure 24: *The calculated trigger-bond potential,* $V_{mid,avg}$ *and molecular electrostatic potential surface maximum,* V_{max} *plotted against the experimentally observed impact sensitivities* ($h_{50\%}$) *for the three nitro-aromatic explosives.*[323] *RDX has a different symbol as it is from a different chemical family, nitramines, and its observed* $h_{50\%}$ *has been obtained from a different source.*[348] *The variations between the optimized and the most different crystalline conformations; HNB(opt ≈ expt), TNB(opt ≈ form III, molecule 1 in form I, molecule 2 in form II) TNT(*opt, α*) and RDX(opt(AAA),opt (AAE), α(AAE) & ε(AAA)) are shown. The atomic charge variations, other electrostatic properties (including* $V_{mid,longest}, V_{min}$*) are given in* ***Appendix 4.B****, along with the plotted values.*

4.3.2 Modelling the electrostatic contribution to lattice energies

4.3.2.1 The effects of NO_2 conformation on molecular charge density

How much variation is likely within the crystalline state and how does this affect atomic charges? Analysis of the NO_2 torsion angles in the crystal structures in the Cambridge Structural Database (CSD) shows that an aromatic NO_2 group adjacent to two hydrogen substituents is almost always planar, to within 20°, which corresponds to a fairly small change in conformational energy, as shown for the para-nitro group in TNT in ***Figure 25***. There is a large barrier of about 25 $kJ\ mol^{-1}$ for rotating the nitro group, presumably from changes in π electron conjugation. The atomic charge on C4 (the carbon bound to the para-NO_2) rises, while those on the neighbouring C3 and C5 fall. It was found that the atomic charge on N2 is mainly responsible for the charge balance. Additionally, the dipole moments on the O atoms change significantly with rotation. A similar picture is seen for the NO_2 groups in TNB (**Appendix 4.A**), which are also adjacent to aromatic $C\text{-}H$ groups. The picture is very different for the ortho nitro-groups in TNT (***Figure 25***), where a large range of torsion angles are observed for the NO_2 between a methyl and hydrogen substituents on an aromatic ring. Here the planar conformation is the most commonly observed in the crystal structures, but this corresponds to a maximum in the intramolecular energy. There is a known preference for planar, extended conformations of molecules in the crystalline state as this often allows a denser packing.[370] The height of the energy barrier is probably overestimated because the methyl group being fixed in the calculation and hence sterically clashing with the NO_2 group, whereas in the crystal structure the methyl could rotate out of the way. However, one should note that the crystallographic angle in the CSD is an average over the torsional librations of the methyl and nitro-groups. In the ortho case the charges and dipoles change on many more atoms, though again the dipoles change most on the nitro-oxygen atoms. While this effect is probably exaggerated by the steric clash, it highlights the effects of changing π conjugation between the nitro groups and the aromatic ring.

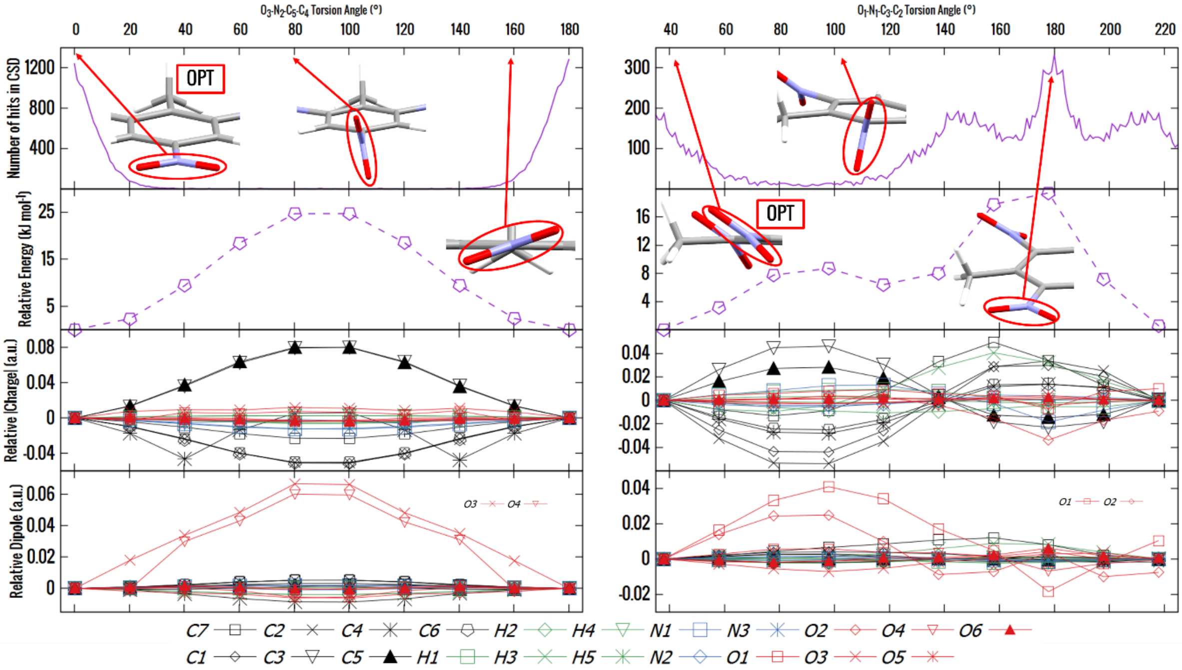
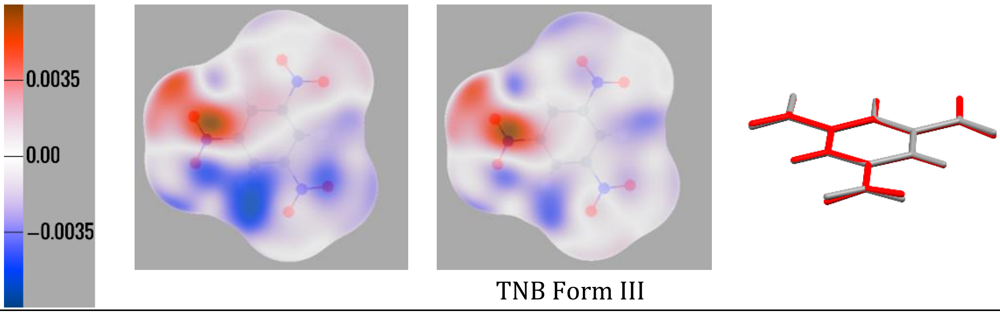


Figure 25: *Conformational behaviour of the para- (left) and ortho- (right) nitro groups in TNT. The behaviours as a function of torsion angle is given for (top to bottom) the distribution of the observed angles in the CSD for each environment; the change in PBE0/aug-cc-pVTZ energy relative to the optimized molecule as the nitro group torsion angle is changed; and the relative changes in the magnitude of ISA charge and dipole on each atom. The oxygen atoms that undergo the most significant changes in dipole moment have been indicated, and the optimized NO_2 torsion angles are also illustrated, alongside other more notable NO_2 torsion angles.*

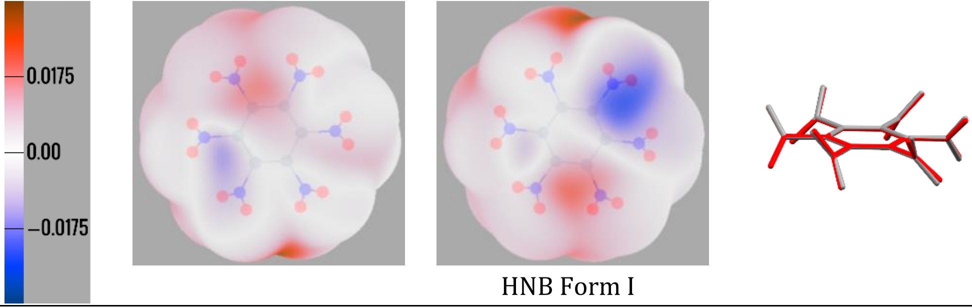
There is far sparser CSD data for an aromatic NO_2 between two other NO_2 groups (HNB), suggesting that any angle could be observed (**Appendix 4.A**). Nonetheless, the height of the conformational energy barrier is so great that without allowing for a concerted and highly correlated change in the other NO_2 torsion angles the range of angles seen in other crystalline conformations would not occur in HNB. There is even less CSD data for N-NO_2 groups (**Appendix 4.A**) as seen in the analysis of the AAE conformer of RDX. It was found that there were significant variations in some of the atomic charges and dipoles for the ring nitrogen and the attached NO_2 atoms. Overall, from ***Figure 25***, it is clear that the nitro group can adopt a range of conformations, as a compromise between the intramolecular steric hindrance and the crystals packing forces, and that these changes in conformation do affect the nitro-group charge distribution, resulting in very different charge distributions for various NO_2 environments.

The difference maps in ***Figure 26*** show the extent to which the changes in charge distribution, from the rearrangement of the valence electrons, affect the electrostatic potential around each nitro-aromatic molecule. The difference between the potential calculated for *optexptNO₂* and *anarot* multipole moments is shown. The differences are very dependent on the extent to which the torsion angles differ between the optimized and experimental structures; ranging from being negligible for TNB, to underestimating V_{min} by nearly $0.3\ eV$ for TNT.



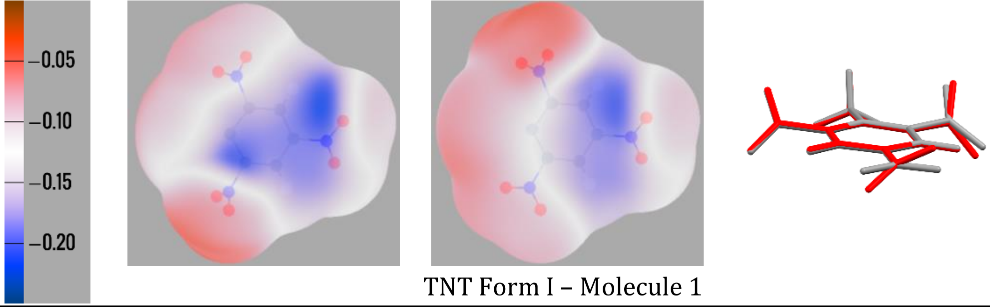
TNB Form III

$\Delta V_{min} = -0.0061eV, \Delta V_{max} = 0.0070eV$ $RMSD_1 = 0.067$ Å



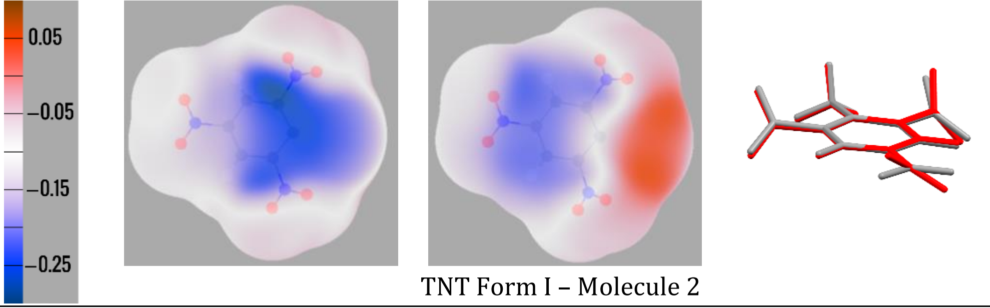
HNB Form I

$\Delta V_{min} = -0.0235eV, \Delta V_{max} = 0.0332eV$ $RMSD_1 = 0.057$ Å



TNT Form I – Molecule 1

$\Delta V_{min} = -0.22eV, \Delta V_{max} = -0.05eV$ $RMSD_1 = 0.258$ Å



TNT Form I – Molecule 2

$\Delta V_{min} = -0.28eV, \Delta V_{max} = 0.05eV$ $RMSD_1 = 0.251$ Å

Figure 26: *The difference in electrostatic potential for molecular charge distributions calculated with the* NO_2 *groups in their experimentally observed torsion angles (optexptNO*$_2$*) and those analytically rotated (anarot) viewed from above and below. The potential is mapped onto an iso-density surface of* 10^{-3} $electron/bohr^3$ *in eV with a varying scale. The* $RMSD_1$ *(in Å) for the overlays of the optimized (grey) and optimized with experimental* NO_2 *torsion angles (optexptNO*$_2$*) (red) molecular conformations, for each molecule is included for comparison.*

4.3.2.2 The effects of NO₂ conformation on lattice energies

Table 15 illustrates how analytically rotating the atomic multipole moments affects the intermolecular lattice energy (E_{latt}). For TNT, which has the largest NO_2 torsion angle difference between experimental and optimized structures, there is a significant change in the lattice energies of TNT (***Table 15***). However, the effect of *anarot* multipoles on the minimized lattice parameters and hence potential energy surface minima is negligible. The small change in the lattice energy of TNB reflects the very small differences in the torsion angles upon crystallization. The lattice energy differences are mainly due to variation in the electrostatic contributions ($V_{elst}^{(1)}[DM]$) as changes in the dispersion-repulsion terms ($V_{disp-rep}$) arise only from the change in the lattice structure. Even though the relative lattice energies are not well reproduced when rotated multipole moments are employed, the low RMSD_{15} values indicate that the crystal geometries differ insignificantly. This is also illustrated by the overlays in ***Figure 22.*** Hence the *anarot* distributed multipole model, which assumes that the atomic charge distributions can be analytically rotated with changes in the NO_2 torsion angle, reproduces the experimental polymorph structures satisfactorily enough to be used in CSP for initial analysis of the potential energy landscape.

	TNT			TNB			HNB		
REFCODE	ZZZMUC08			TNBENZ13			HNOBEN		
	E_{latt}	$V_{elst}^{(1)}[DM]$	$V_{disp-rep}$	E_{latt}	$V_{elst}^{(1)}[DM]$	$V_{disp-rep}$	E_{latt}	$V_{elst}^{(1)}[DM]$	$V_{disp-rep}$
$E_{optexptNO2-ISA}$ $/kJ\ mol^{-1}$	-121.0	-46.4	-74.6	-104.7	-34.4	-70.3	-150.1	-45.1	-105.1
$E_{anarot-\mathrm{ISA}}$ $/kJ\ mol^{-1}$	-123.2	-48.3	-74.9	-104.4	-34.2	-70.2	-150.8	-46.0	-104.9
$\Delta\mathrm{E}\ /\ kJ\ mol^{-1}$	2.177	1.820	0.367	-0.334	-0.262	-0.073	0.710	0.910	-0.200
%↑ in a	0.270			-0.015			0.060		
%↑ in b	0.259			-0.244			-0.050		
%↑ in c	-0.333			0.190			0.058		
$RMSD_{15}/\text{Å}$ (*optexptNO2*)	0.246			0.166			0.235		
$RMSD_{15}/\text{Å}$ (*anarot*)	0.218			0.162			0.237		

Table 15: *A comparison of the effect of ISA multipoles moments calculated in the optexptNO$_2$ conformation and those analytically rotated into the optexptNO$_2$ conformation (anarot) using ORIENT. Each experimental crystal structure has been minimized holding the molecules rigid in their optextNO$_2$ observed conformation, using the FIT potential*[371] *and the defined set of atomic multipole moments to calculate the intermolecular forces. The changes in each intermolecular energy contribution and % change in cell parameters are compared. The RMSD$_{15}$ comparisons are with the experimental crystal structure. More details on the effect of conformation on lattice energy minimizations, including using the optimized molecular conformation and the experimental conformation can be found in the* ***Appendix 4.B****.*

4.4 Discussion

The crystalline and isolated (optimized) conformations of most molecular explosive materials were found to differ significantly. This provides the challenge of evaluating the thermodynamic stability of their polymorphs even at ambient conditions[372] let alone at the high temperatures and pressures sampled during detonation. Here we have highlighted the effects of the changes in the NO_2 torsion angles that can occur upon crystallization; how these can differ between polymorphs and its importance in accurate predictive modelling.

4.4.1 The influence of conformation on electrostatic properties

The experimental polymorphs studied show a range of crystalline conformations that are fairly typical of those seen in all the other nitro-aromatic and nitramines in the Cambridge Structural Database (CSD). When the adjacent functional groups on the aromatic ring (or $C\text{-}N\text{-}NO_2$ chain, e.g. RDX) are hydrogen atoms, the NO_2 torsion angles vary by about $\pm 20°$ and the nitro groups are relatively planar (***Figure 25*, Appendix 4.A *Figure 27*** (a)). On the other hand, when the adjacent functional groups are larger, any angle may be observed (***Figure 25***, **Appendix 4.A *Figure 27*** (b)). In the isolated molecule, the interactions with the specific neighbouring substituents can have a major effect on the charge distribution and torsion angle of a nitro group. However, in the condensed phase, correlated changes in substituent conformation are more affected by crystal packing. Comparing NO_2 torsions in ***Figure 20***, the nitro-aromatics show differing degrees of conformational adjustment due to the packing forces within the crystal lattice. HNB has only one crystal structure, but the six nitro groups have different torsion angles ranging between 48-60°, losing its molecular symmetry and reflecting the compromises between the intramolecular steric forces and the adaption to intermolecular interactions to give a dense crystal. The TNT polymorphs have rather similar crystalline conformations as they are polytypic, comprising of different stacks of comparable layers of molecules.[6, 373] Nonetheless, the two independent molecules in each polymorph show variations in the torsion angles of the ortho-NO_2 groups that sterically interact with the methyl, and the para-NO_2 groups are not coplanar. The optimized TNB molecule is planar, but the four crystalline conformations in the two polymorphs have torsion angles that are up to 32° from planar. Thus, crystal packing forces can easily overcome intramolecular stabilization from conjugation, shifting nitro groups from their preferred planar conformation.

Using an improved novel method of partitioning the isolated molecular charge distribution (ISA) and a high level of theory (PBE0/aug-cc-VTZ) to obtain realistic atomic charges and higher order multipole moments for an X-NO_2 group shows that the NO_2 charge density differs considerably between N-NO_2 and C-NO_2 (***Figure 20***, ***Table 14***). Moreover, the charge density varies significantly with the neighbouring functional groups, (e.g. ortho and para groups in TNT, ***Figure 20***), and also with the NO_2 torsion angle. This is to be expected as the atomic charges, in particular, reflect the changes in valence electron distribution, which also determine the molecular reactivity. This investigation shows that the published empirical correlations of local electrostatic properties around the NO_2 "trigger-bond", such as V_{mid} ,or the molecular property V_{max} with the crystal property, impact sensitivity ($h_{50\%}$).[137, 356, 374] will be significantly affected by the neglect of the conformational change (***Figure 24***, ***Table 14***).

The packing arrangement of the reactive groups in the crystal will also strongly affect the energetic properties. The early correlation between detonation properties and density led to the development of MOLPAK,[20] one of the first Crystal Structure Prediction (CSP) codes, which sought to establish the possible density of a crystal from the molecular structure. Since then CSP methods have evolved to help predict possible polymorphs, and help support their structural characterization from powder X-ray diffraction, and other analytical techniques. Whilst these techniques are being extensively developed for the pharmaceutical industry,[30, 91] the extension to energetic materials requires fundamentally different modelling of specific functional groups under the extreme conditions unique to explosives. For example, periodic density functional calculations on crystalline nitroguanidine required the development of a specific dispersion correction.[375]

4.4.2 The influence of conformation on interaction energies

The care required in CSP studies of nitro-aromatic explosives has been established, to ensure that the large variations in possible NO_2 torsion angles, shown by the CSD distributions (***Figure 25*** and **Appendix 4.A *Figure 27***) are effectively considered. It is the balance between the molecular cost for conformational change (ΔE_{intra}) and the improved intermolecular lattice energy (E_{int}) from a denser packing, which determine the total lattice energy ($E_{latt} = \Delta E_{intra} + E_{int}$). Ab initio calculations on the molecule, the Ψ_M approach (**Chapter 2**) can be used to obtain ΔE_{intra} and analyzed to obtain the atomic multipoles, polarizabilities and dispersion coefficients that determine the long range intermolecular forces. In this chapter, the ISA analysis of Ψ_M provides an anisotropic intermolecular potential with the atomic multipole moments being used to calculate the electrostatic contribution to the intermolecular energy. It seems that cell geometries can be obtained relatively accurately and very quickly, without recalculating the charge density as the nitro groups rotate (***Table 15***), by analytically rotating the multipole moments. This suggests that the *anarot* methodology could be used to determine the range of low energy crystal structures that could be adopted by a nitro-aromatic energetic molecule within the Ψ_M approach to Crystal Structure Prediction,[376] and thus could improve the prediction of crystal packing from the molecular diagram of new energetic molecules.[365] This method is suitable as an initial approximation of the potential energy landscape but the relative energy differences in ***Table 15*** are large in comparison with the lattice energy differences between polymorphs, which are usually much less than 5 $kJ\ mol^{-1}$ for small molecules.[7] The conclusion that analytical rotation does not reproduce relative electrostatic energies well was also reached in the study on the transferability of electrostatic models between polypeptides.[367] ***Table 15*** shows when doing ab initio calculations on the molecule (Ψ_M), it is necessary to use an accurate molecular charge density for the refinement of CSP generated crystal structures, or a sufficiently accurate periodic electronic structure calculation, in order to determine the relative lattice energies reliably and accurately enough to know which of the generated crystal structures are thermodynamically plausible polymorphs.[376]

4.5 Conclusion

The charge distributions of the nitro groups in TNB, TNT, HNB and RDX vary considerably depending on the bonded and neighbouring functional groups, and also with the conformational differences between their polymorphs and their isolated molecular structures. The rigorous ISA partitioning of high quality molecular charge density allows us to obtain the distributed multipole representation of the atomic charge density, which has minimal sensitivity to basis set and hence artefacts of conformational change are minimized. Using ISA partitioning this investigation has shown that variations in NO_2 conformations, which occur in crystal structures cause sufficient changes in the molecular charge density to affect previously proposed empirical correlations between impact sensitivity and molecular electrostatic properties. This chapter demonstrates that it is a poor assumption that the molecular charge density will not be significantly changed when placed in the crystalline environment.

The condensed phase torsion-angles observed for nitro-groups are a subtle balance between intra- and intermolecular interactions. The charge distributions of the nitro-groups, and some neighbouring atoms, can change significantly with the NO_2 torsion angle (***Figure 25*** & ***Figure 27***), reflecting the change in the bonding and molecular charge distribution. This change is sufficiently large that distributed multipole models, which form the basis of non-empirical model intermolecular potentials, should be calculated for each conformation if the potential is to be accurate enough to model lattice energies to the accuracy needed to predict polymorph relative stability. However, analytically rotating the multipoles to reflect the change in aromatic NO_2 torsion angles is a reasonable approximation that may be useful in performing fast CSP calculations. This chapter utilizes old and new computational methodologies in a novel way to establish the necessary assumptions about the transferability of charge distribution in nitro-aromatic explosives. These assumptions will be instrumental in producing reliable non-empirical potentials suitable for modelling possible polymorphs, particularly those crystallized under pressure.[199] In the next section, **Chapter 5**, the lessons learned here, with regards to the importance of an accurate description of the molecular charge density, shall be carried forward to develop a rigid non-empirical distributed intermolecular force-field for trinitrobenzene.

Appendix 4.A

3. *Effects of nitro-group rotation on electrostatic properties*

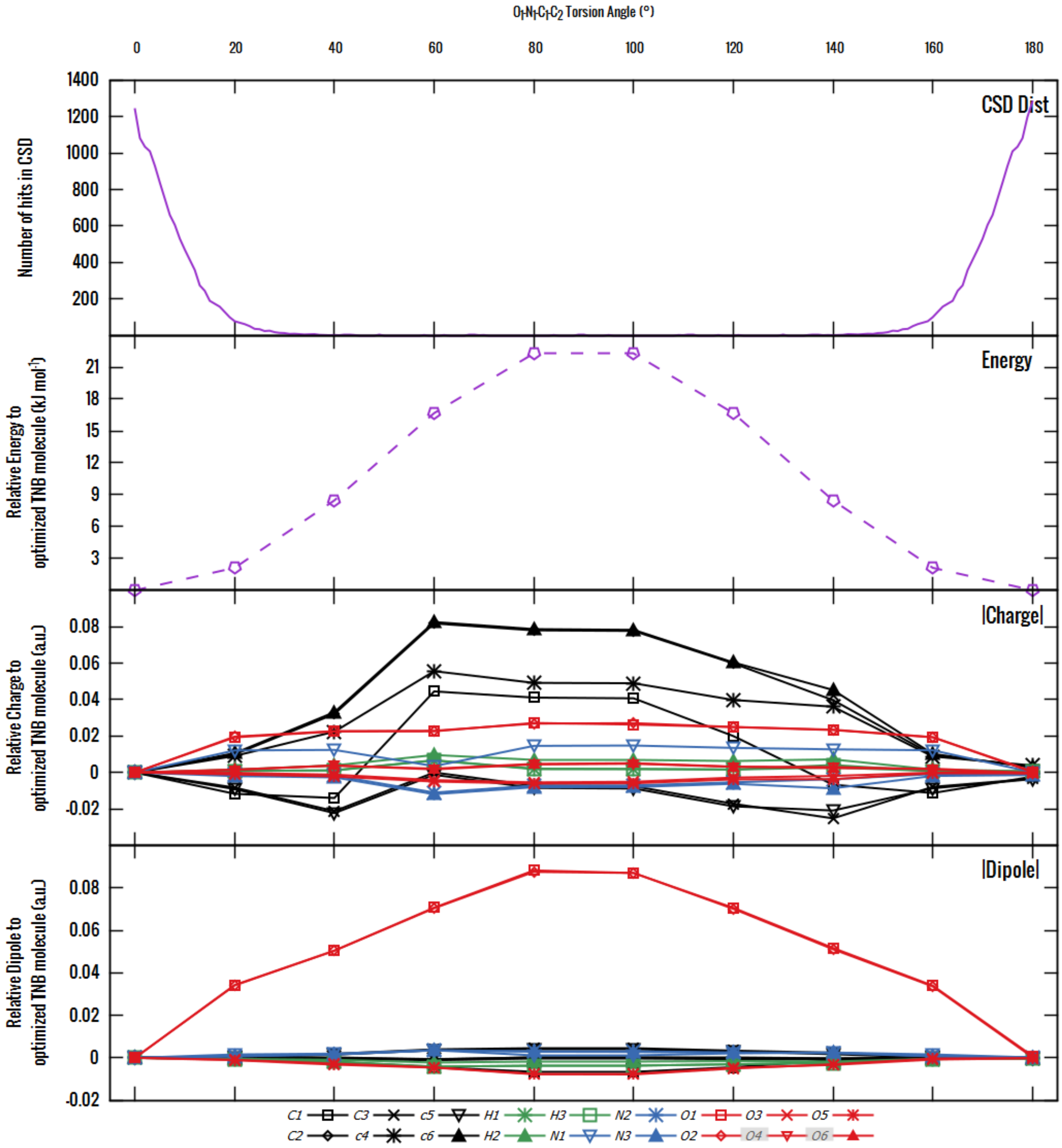
The relative changes in intramolecular energy and multipole moments with torsion angle for the remaining energetic molecules from ***Figure 20***.

4. *The effects of electrostatic model on the lattice*

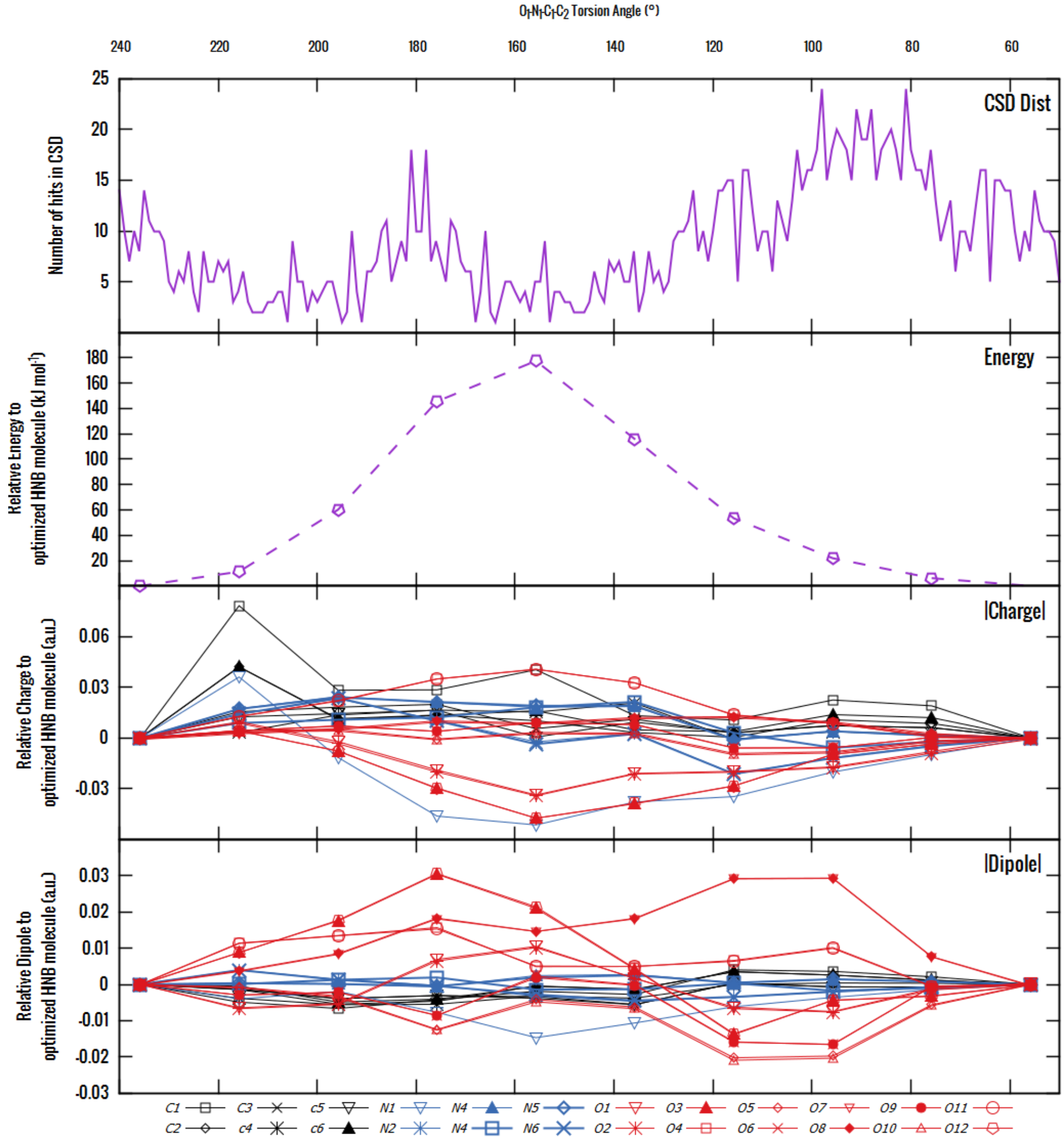
Analysis of how rotating the multipole moments and changes in conformation can affect the lattice energy and minimized cell geometries.

EFFECTS OF ROTATING NITRO-GROUPS IN TNB, HNB AND RDX (AAE)

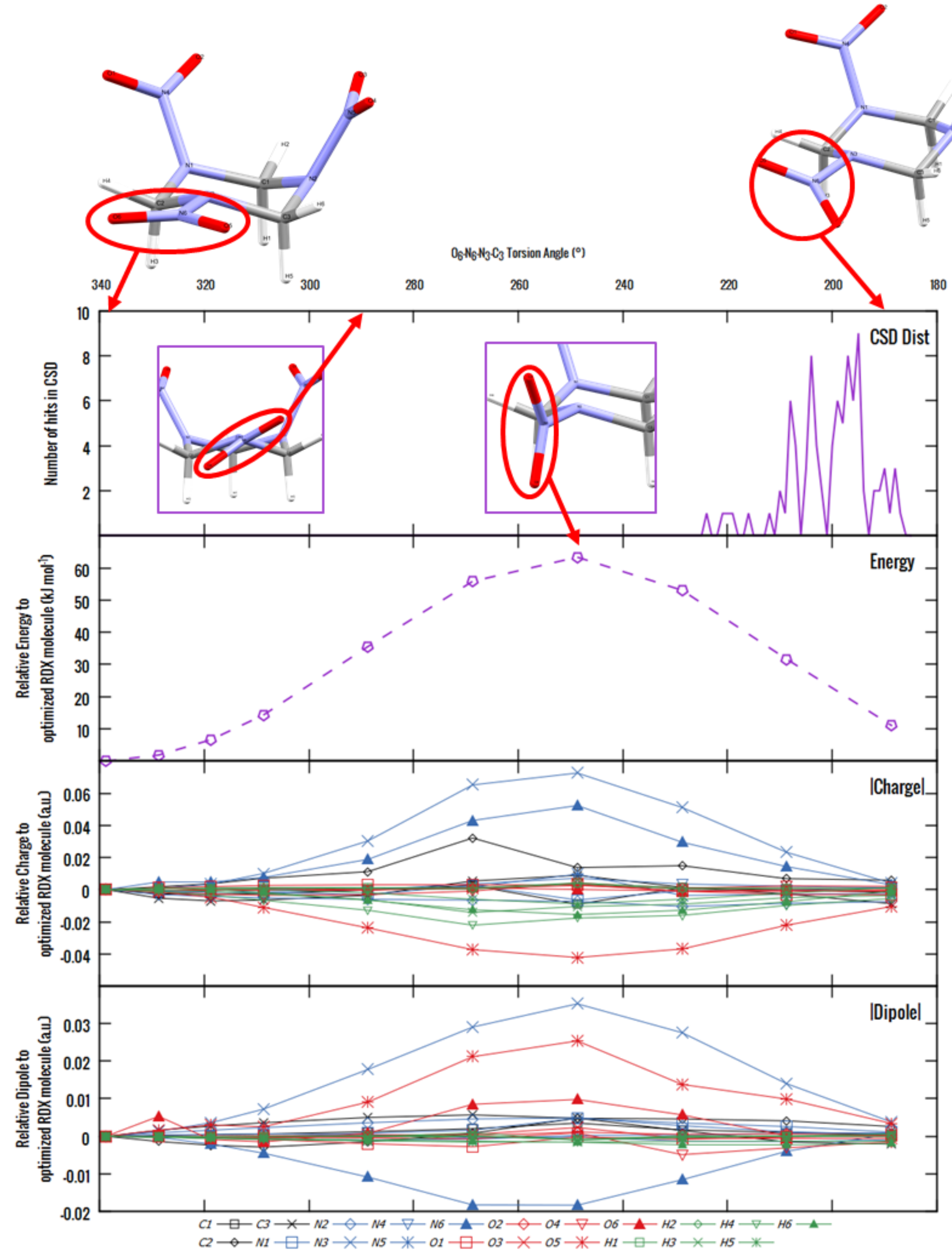
The relative changes of the charge and dipole moment magnitudes with NO_2 torsion angle are compared with Cambridge Structure Database (CSD) distributions and the intramolecular energy relative to the optimized isolated molecular structure for the molecules where this data is not given in this chapter.



(a) TNB


O1N1C1C2 Torsion Angle (°)
240
220
200
180
160
140
120
100
80
60
25
20
15
10
5
Number of hits in CSD
CSD Dist
180
160
140
120
100
80
60
40
20
Relative Energy to optimized HNB molecule (kJ mol-1)
Energy
0.06
0.03
0
-0.03
Relative Charge to optimized HNB molecule (a.u.)
|Charge|
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
Relative Dipole to optimized HNB molecule (a.u.)
|Dipole|
C1
C2
C3
c4
c5
c6
N1
N2
N4
N4
N5
N6
O1
O2
O3
O4
O5
O6
O7
O8
O9
O10
O11
O12

(b) HNB



(c) RDX-*AAE*

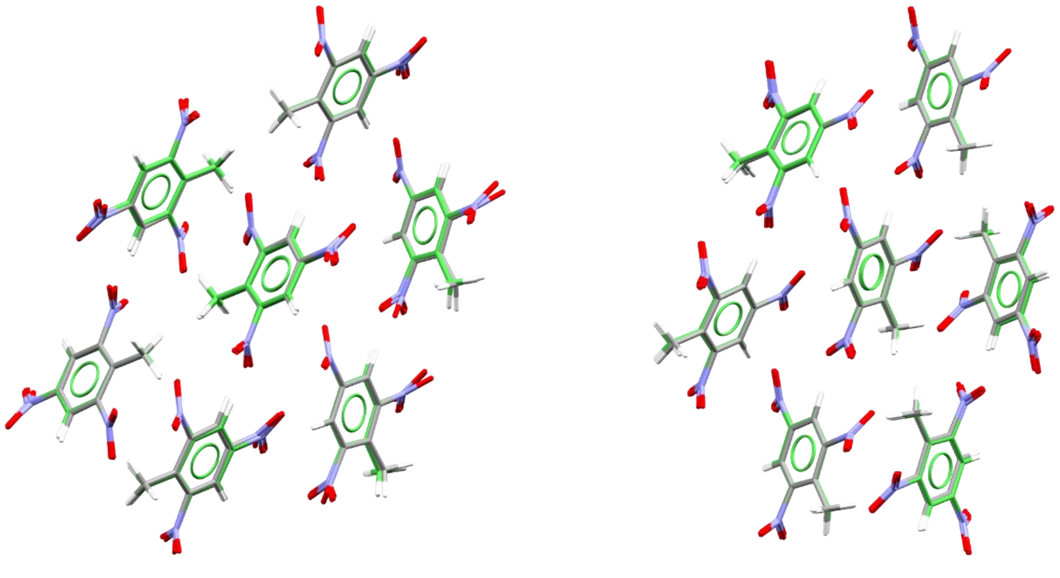
Figure 27: *Conformational behaviour of TNB (a), HNB (b) and RDX-AAE (c) as their NO$_2$ torsion angles are changed. Only the equatorial NO$_2$ in RDX (c) is rotated. The behaviours as a function of angle is given for (top to bottom) the CSD distribution of observed angles for this group; the change in PBE0/aug-cc-VTZ energy relative to the optimized molecule as only the nitro group torsion angle is changed; and the variation in the ISA charge and dipole on each atom.*

Str (REFCODE)	TNT (ZZZMUC08 (Form I))				
	expt	xminexpt	xminopt	xminoptexptNO2	xminanarot
$a/Å$	14.911	14.982	14.209	14.854	14.814
$b/Å$	6.034	6.079	7.397	6.135	6.119
$c/Å$	20.882	20.774	19.386	21.078	21.148
$n(RMSD_{15})/Å$	~	15 (0.064)	3 (0.683)	15 (0.246)	14 (0.214)
E_{latt} $(kJ\ mol^{-1})$		-117.6	-100.4	-121	-123.2
Str (REFCODE)	TNB (TNBENZ13 (Form III))				
	expt	xminexpt	xminopt	xminoptexptNO2	xminanarot
$a/Å$	12.896	12.833	12.751	12.779	12.781
$b/Å$	5.723	5.871	5.603	5.743	5.757
$c/Å$	11.287	11.038	11.667	11.414	11.393
$n(RMSD_{15})/Å$	~	15 (0.154)	15 (0.245)	15 (0.166)	15 (0.162)
E_{latt} $(kJ\ mol^{-1})$		-112.6	-111.8	-104.7	-104.4
Str (REFCODE)	HNB (HNOBEN (Form I))				
	expt	xminexpt	xminopt	xminoptexptNO2	xminanarot
$a/Å$	13.22	12.822	12.960	12.768	12.760
$b/Å$	9.13	8.978	8.792	8.812	8.817
$c/Å$	9.68	9.442	9.517	9.558	9.553
$n(RMSD_{15})/Å$	~	15 (0.170)	15 (0.214)	15 (0.235)	15 (0.237)
E_{latt} $(kJ\ mol^{-1})$		-146.1	-144.6	-150.1	-150.8

Table 16: *The expt (experimental structure), xminexpt (DMACRYS*[59] *minimized crystal structure, with all molecules held rigid in their experimentally observed conformations), xminopt (DMACRYS minimized crystal structure, with all molecules held rigid in their gas-phase optimized conformation), xminoptexptNO*$_2$ *(DMACRYS minimized optimized structure with experimental NO*$_2$ *torsions using* ψ *computed after rotation) and the xminanarot (DMACRYS minimized optimized structure with experimental NO*$_2$ *torsions using analytically rotated multipole moments) crystal structures and intermolecular lattice energies of TNT, TNB and HNB. The intermolecular lattice energy (*E_{latt}*) estimates above are calculated using a distributed multipole electrostatic force-field derived from the molecular (PBE0/aug-cc-pVTZ) wave-function and empirical repulsion-dispersion model; the empirical FIT potential*[371] *& ISA*[233] *(PBE0/aug-cc-pVTZ). The* $RMSD_{15}$ *values have been calculated against each respective experimental determination.*

Structure	RDX					
REFCODE	CTMTNA03 (Form α)			CTMTNA06 (Form ε)		
	Expt	xminexpt	xminopt	Expt	xminexpt	xminopt
$a/\text{Å}$	11.420	11.635	11.540	7.032	7.371	7.715
$b/\text{Å}$	10.586	10.615	10.561	10.530	11.113	12.120
$c/\text{Å}$	13.140	13.426	13.787	8.791	9.673	9.194
$n(RMSD_{15})/\text{Å}$	~	15 (0.169)	15 (0.447)	~	15 (0.431)	7 (0.857)
$E_{latt}\ (kJ\ mol^{-1})$	~	-121.1	-121.1	~	-118.9	-93.4

Table 17: *The expt (experimental structure), xminexpt (DMACRYS[59] minimized crystal structure, with all molecules held rigid in their experimentally observed conformations) and xminopt (DMACRYS minimized crystal structure, with all molecules held rigid in their gas-phase optimized conformation) crystal structures and intermolecular lattice energies of RDX. The intermolecular lattice energy (E_{latt}) estimates above are calculated using a distributed multipole electrostatic force-field derived from the molecular (PBE0/aug-cc-pVTZ) wave-function and empirical repulsion-dispersion model; the empirical FIT potential[371] & ISA[233] (PBE0/aug-cc-pVTZ). This highlights how slight differences in the experimentally observed and optimized conformations of both polymorphs of RDX can result in large changes in intermolecular energy. $RMSD_{15}$ against each respective experimental determination.*



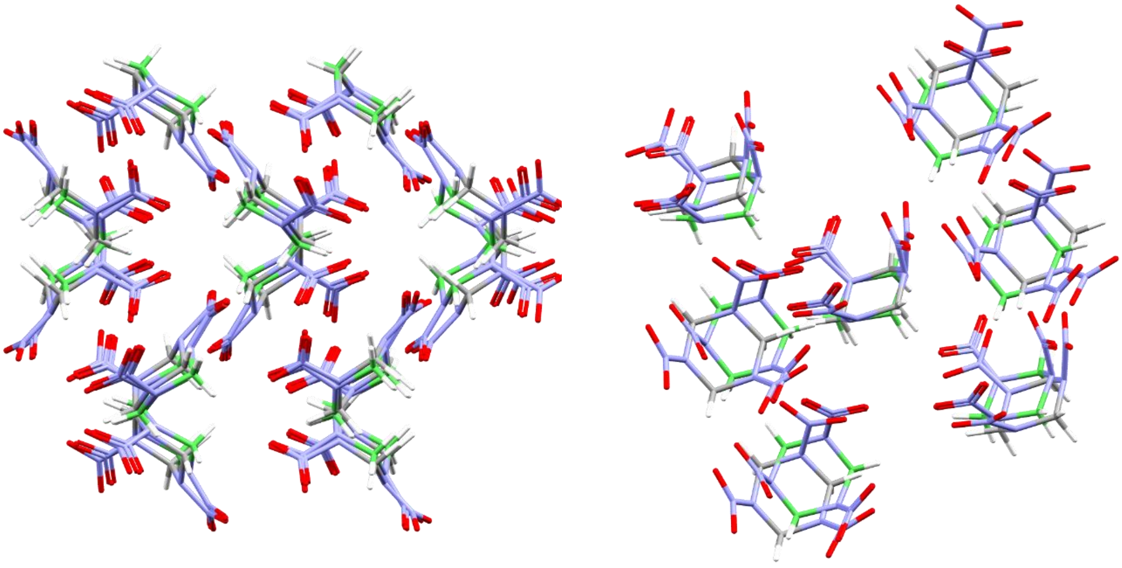
(a) (b)

Figure 28: *Overlay of the experimental ZZZMUC08[144] structure (green) and their DMACRYS minimized structures using the molecular conformations and distributed multipole moments:*

(a) calculated for the PBE0/aug-cc-pVTZ optimized structure after rotating all NO_2 groups into their experimental observed torsion angles, *optexptNO₂* (grey). $RMSD_{15} = 0.246\text{Å}$

(b) calculated for the PBE0/aug-cc-pVTZ optimized structure and then analytically rotated (*anarot*) into the experimental observed NO_2 torsion angles (grey). $RMSD_{14} = 0.214Å$

Figure 28 shows that even for TNT, which has the greatest differences in optimized and observed NO_2 torsion angles, analytical rotation of the atomic multipoles into the experimental NO_2 torsion angles and recalculation of the wave-function for each NO_2 torsion angle results in very small structural differences. The only significant differences are seen in the intermolecular energies.



(a) (b)

Figure 29: *Overlay of the experimental structures of RDX (green), CTMTNA03*[141] *(a) and CTMTNA06*[143] *(b) and their DMACRYS minimized structures using the molecular conformations and distributed multipole moments:*

(a) calculated for the PBE0/aug-cc-pVTZ optimized AAE conformation in the crystal, *xminopt* (grey). $RMSD_{15} = 0.447Å$

(b) calculated for the PBE0/aug-cc-pVTZ optimized AAA conformation in the crystal, *xminopt* (grey). The *xminopt* crystal minimizes to a very different minima and energy compared to the experimental lattice (*xminexpt*), with only 7 molecules out of 15 matching and a high RMSD value. $RMSD_7 = 0.857Å$.

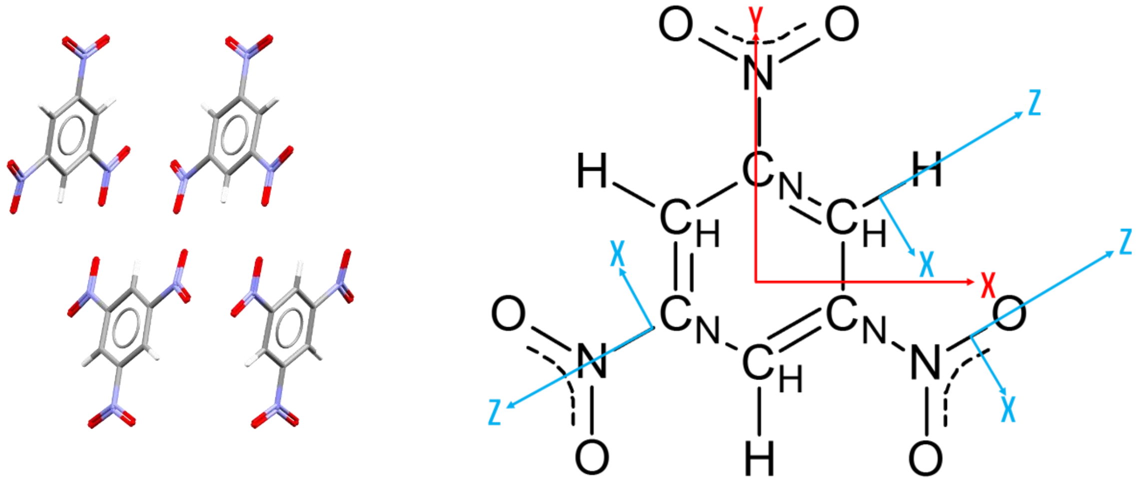
5 Non-empirical force-field development for trinitrobenzene (TNB)

This chapter pioneers the development of non-empirical force-fields for larger, non-spherical molecules. It aims to establish how far the methods previously used for smaller molecules such as pyridine can be applied to larger molecules, to provide potentials suitable for use in crystal structure prediction (CSP). The potential development process Misquitta used for pyridine[38] has been further built upon and adapted for the larger more challenging trinitrobenzene (TNB). Using the general workflow described in ***Figure 2*** in **Chapter 1**, we aim to illustrate how to develop and test a non-empirical anisotropic atom-atom intermolecular force-field for CSP from scratch. The various stages of the model building process shall be discussed in depth and one shall see how the non-empirical potentials are tailored to each molecule. The products of this investigation are multiple non-empirical force-fields. Two approximate models are created from the 1st order SAPT(DFT) dimer interaction energies dubbed Model0-aniso and Model0-iso, which respectively have anisotropic and isotropic short-range descriptions. The complete and final force-field for TNB, a Distributed Intermolecular Force-Field (DIFF) is created, which aims to model the genuine potential energy surface of TNB to SAPT(DFT) quality. In the force-field development process some transient potentials dubbed ModelD, ModelR and ModelA are also created by fitting to the 2nd order SAPT(DFT) interaction energies of different sets of dimer configurations. These are described in detail in **5.6.2.1**.

5.1 Trinitrobenzene (TNB)

Trinitrobenzene (TNB) is a comparably small energetic molecule but a large system for high quality molecular calculations. Energetic materials are intrinsically hard to study in the lab due to safety concerns but can also be difficult to model accurately and dependably because of their size. Due to the scaling of the underlying electronic structure calculations with system size, TNB stretches the current codes and computational resources available. As a result, adaptions have been made to the Misquitta's previous workflow[38] and various novel codes have been tested. TNB (***Figure 30***) has been chosen as a modest sized molecule of 18 atoms, 108 electrons, whose molecular structure could be assumed to be rigid as it does not vary drastically in its three polymorphs (**Chapter 4**). The high symmetry of TNB's isolated molecule has the advantages that the unique atoms types are O, N, H, C_N and C_H.

Accordingly, the static isolated molecular structure is obtained by optimizing the experimental structure TNBENZ13[263] using the Gaussian09 program[235], the PBE0 functional[203-205] (the Perdue-Burke-Emzerhof (PBE) general gradient approximation combined with a portion of exact exchange) and the d-aug-cc-pVTZ Dunning basis,[214] and kept rigid in all simulations. An identical calculation was carried out on pyridine (**Chapter 3**), and the molecular and local axis were defined in a similar manner. In order to avoid any future complications or re-fitting (which had to be done for pyridine (**Chapter 3**)), it is important to correctly define the axis systems. The molecule-fixed and local axis definitions will determine the multipole moments, atomic polarizabilities and anisotropic repulsion terms. The definitions will also help us later determine which terms on which atoms are most important (**5.4.3.1**).



(a) (b)

Figure 30: *(a) The experimentally observed form III of trinitrobenzene (TNBENZ13).*[263] *The other 2 polymorphs of TNB can be found in* ***Chapter 6 Figure 44****. (b) The molecule-fixed axis (red) and local axis (blue) definitions used for TNB.*

The energetic material TNB poses the challenge of accurately modelling a large non-spherical organic molecule with a limited amount of experimental data for reference. Challenges like this stresses the importance of constructing a sound model from scratch based on the theory of intermolecular forces[164] (**Chapter 2**) and generating further theoretical data to test the force-field.

5.2 The intermolecular interaction energy and its components

The non-empirical distributed intermolecular force-fields (DIFF) were developed using a combination of programs: CamCASP[107], Psi4[228], DMACRYS[59] and ORIENT.[108] Using the theory of intermolecular forces introduced and defined in **Chapter 2**, and the axis definitions and atomic types in ***Figure 30***, the total interaction energy can be represented as a non-empirical potential, whose long-range components are represented using the distributed multipole expansion[164] and the short-range components represented as an exponential, with the following functional form:

$$U_{M,N}(R_{MN},\Omega_{MN}) = V_{SR} + V_{LR}$$

$$\sum_{i\in M,k\in N} U_{INT}^{ik}(R_{ik},\Omega_{ik}) = G\exp\left[-\alpha^{\iota\kappa}\left(R_{ik} - \rho^{\iota\kappa}(\Omega_{ik})\right)\right] + Q_t^i T_{tu}^{ik} Q_u^k + \sum_{n=0}^{2} f_{2n+6}\left(\beta_{disp}, R_{ik}\right)\frac{C_{2n+6}^{\iota\kappa}}{R_{ik}^{2n+6}}$$

$$+f_n\left(\beta_{pol}, R_{ik}\right)(\Delta Q_t^\iota T_{tu}^{\iota\kappa} Q_u^\kappa + \Delta Q_t^\kappa T_{tu}^{\kappa\iota} Q_u^\iota)/2$$

96

With regards to tensor notation, the long-range multipolar electrostatic potential energy $V_{elst}^{(1)}[DM]$ can be given as $Q_t^i T_{tu}^{ik} Q_u^k$ where Q_t^i is a multipole moment on atomic site i of rank t, using the notation $t = 00{,}10{,}11c, \ldots$ to describe the multipole moment and its related axes[164]. T_{tu}^{ik} is the transformation tensor[164] determined by our axes definitions (***Figure 30***). The same notation (tu) used for multipole components is used for the polarizabilities and transformation tensors. The Tang-Toennies damping function f_n is defined in **2.8** & **5.3.4**. The remaining parameters are described and contextualised in detail in **Chapter 2** and below.

5.3 Calculating long-range terms from molecular properties

After ab initio calculations of the molecular wave-function (Ψ_M) on the isolated static molecule using a quantum chemical program like Psi4[228], NWCHEM[377] or Dalton[378], one can employ CAMCASP[107] to calculate the following molecular properties:

- Distributed multipole moments (**2.5**) – these are then used to obtain the electrostatic multipolar energy.
- Distributed atomic (local) polarizabilities (**2.6.1**) – can be calculated using the Williams-Stone-Misquitta algorithm,[285] however, we use the ISA-Pol method[200] to obtain the polarization energy
- Distributed two-centred dispersion coefficients (**2.7**) – used in a damped Buckingham potential to obtain the dispersion energy.

For TNB a combination of Psi4, NWCHEM and CAMCASP was used. As previously described in **Chapter 2**, the intermolecular interaction energy can be separated into long-range contributions, when there is no overlap of the molecular wave-functions Ψ, and short-range contributions, when two Ψ overlap. The long-range multipolar expansion is a power series of 1/R, where R is the centre-of-mass separation of two sites and the long-range energy V_{LR} portion of the force-field is

$$V_{LR} = \sum_{n=0}^{2} f_{2n+6}\left(\beta_{disp} R_{ik}\right)\frac{C_{2n+6}^{\iota\kappa}}{R_{ik}^{2n+6}} + \text{electrostatic } (R_{ik}, Q_t^i\ l \leq 4,\text{ISA}, \Psi) + \text{polarization } (R_{ik}, Q_t^\iota, \alpha_{tt'}^\iota, \beta_{pol})$$

97

The long-range components are the electrostatic, polarization and dispersion contributions to the intermolecular interaction energy.

5.3.1 The electrostatic model from multipole moments

A concise description of the long-range electrostatic contribution requires a sound partitioning of the molecular charge density (calculated with a high quality wave-function) into distributed multipole moments (**2.5**). The charge density of the isolated TNB molecule was calculated using an asymptotically corrected PBE0 hybrid functional and the augA-Sadlej basis set.[379] PBE0 is an asymptotically corrected hybrid functional, therefore, the asymptotic correction for each molecule must be calculated when calculating the molecular charge density. The asymptotic shift (Δ) can be defined as summation of the ionization potential (IP) of the molecule and the energy of the highest occupied molecular orbital (HOMO) eigenvalue (ϵ_{HOMO})

$$\Delta = IP + \epsilon_{HOMO}$$

98

The ϵ_{HOMO} can be obtained after calculating the DFT energy of the neutral molecule, while the ionization potential can either be obtained by finding the difference between the DFT energy of the neutral molecule and its cation or from gas-phase data in the NIST standard reference database.[380] The asymptotic shift (Δ) for TNB was calculated using the experimental IP of 0.4028 a.u.[380] and the energy of the highest occupied molecular orbital (HOMO) eigenvalue (ϵ_{HOMO}) resulting in $\Delta = 0.0539\ \mathrm{a.u.}$. For TNB we move away from the asymptotic correction used for pyridine (the GRAC)[381] to one which can better model anisotropic molecules like TNB. Accordingly, the CS00 (Casida-Salahub) asymptotic correction[382] was used in NWChem[377] to perform a DFT calculation and obtain the molecular wave-function (Ψ_M) for each conformation. The Iterated Stockholder Atoms (ISA) multipoles (evaluated to hexadecapole level (Rank $l = 4$)) are then calculated in CamCASP relative to the molecular axis frame (***Figure 30***) and derived using the most recent iterated stockholder atoms algorithm implemented in CamCASP 6.0,[107] the ISA-A algorithm.[273] The ISA-A method develops on the previous BS-ISA method (**2.5.1.7**) in CamCASP 5.9, which was used for pyridine.[38] ISA-A includes a method for re-fitting the atomic electron density tails,[233] by including another independent atomic basis set, dubbed the AtomAux-Basis. Therefore, the ISA-A algorithm in CamCASP 6.0 contains 4 basis:

- A main basis – to calculate Ψ_M, for TNB the augA-Sadlej is used
- An auxiliary basis – for re-fitting the atomic electron density tails. This basis can take both Cartesian and Spherical Gaussian Type Orbitals (GTOs). We use Cartesian GTOs of aug-cc-pVTZ size.

- An atomic basis – to determine the ISA atomic expansions. An aug-cc-pVQZ basis set is used in calculating the atomic expansions, and is independent from the refitting & molecular wave-function calculations. However, the spherical coordinate system for the GTOs must always be used.
- An ISA-Basis – which is the s-function part of the atomic basis (aug-cc-pVQZ) that is used to obtain the ISA shape-function expansions.

The multipole moments have been evaluated to Rank $l = 4$ in order to obtain an accurate electrostatic model. Dipoles are crucial in describing lone-pair features, and quadrupoles to correctly describe π-orbitals in organic crystals. The computational cost of calculating moments above hexadecapole currently outweighs the improvement in quality of electrostatic model. ***Table 18*** shows the effect of symmetry on the multipole moments up to quadrupole. Other terms are zero due to TNB's D_{3h} symmetry. One can see the importance of higher order moments, in the electrostatic potential maps of the higher order moments on each atom type in TNB (***Figure 31***). ***Figure 31*** highlights the anisotropy of the molecule, the dominant terms, the directionality of each moment, and the importance of going beyond the use of point charge models in force-field development. The chemistry of the molecule is reflected in ***Table 18*** & ***Figure 31***, emphasising the influence and magnitude of the dipole, quadrupole etc. multipole moments. The most dominant higher order multipole moment in ***Table 18*** is observable in the atomic electrostatic maps in ***Figure 31***. For example, the quadrupolar moment Q_{22c} is the most dominant anisotropic term for oxygen and we see in ***Figure 31***, that the most dominant term is a quadrupolar one with the correct sign. Likewise the most prominent anisotropic term on nitrogen is the Q_{11s} dipole moment; also reflected in ***Figure 31***. The inspecting the sign and magnitude of the moment in and ***Figure 31*** provides a good crosscheck.

In the Global Axis	Charge	Dipole		Quadrupole		
	Q_{00}	Q_{11c}	Q_{11s}	Q_{20}	Q_{22c}	Q_{22s}
O	-0.435	0.004	-0.012	-0.006	0.325	0.049
N	0.810	0.036	-0.064	-0.028	0.005	0.009
C_N	-0.033	-0.009	0.016	-0.173	-0.032	-0.055
C_H	-0.081	0.033	0.057	-0.025	-0.001	0.002
H	0.179	-0.023	-0.039	-0.027	-0.011	0.020

Table 18: *The symmetry unique multipoles moments for each atom type in TNB in atomic units up to Rank $l = 2$. The dipole term Q_{10} and quadrupole terms Q_{21c} & Q_{21s} are zero by symmetry. The complete multipole moments (up to Rank = 4) input file can be found in* ***Appendix 5.A*** *and the* ***Appendix CD***. ***Appendix 5.A*** *provides the*

distributed multipole moments ORIENT input file for TNB in atomic units while atom co-ordinates of the rigid optimized structure are given in Bohr.

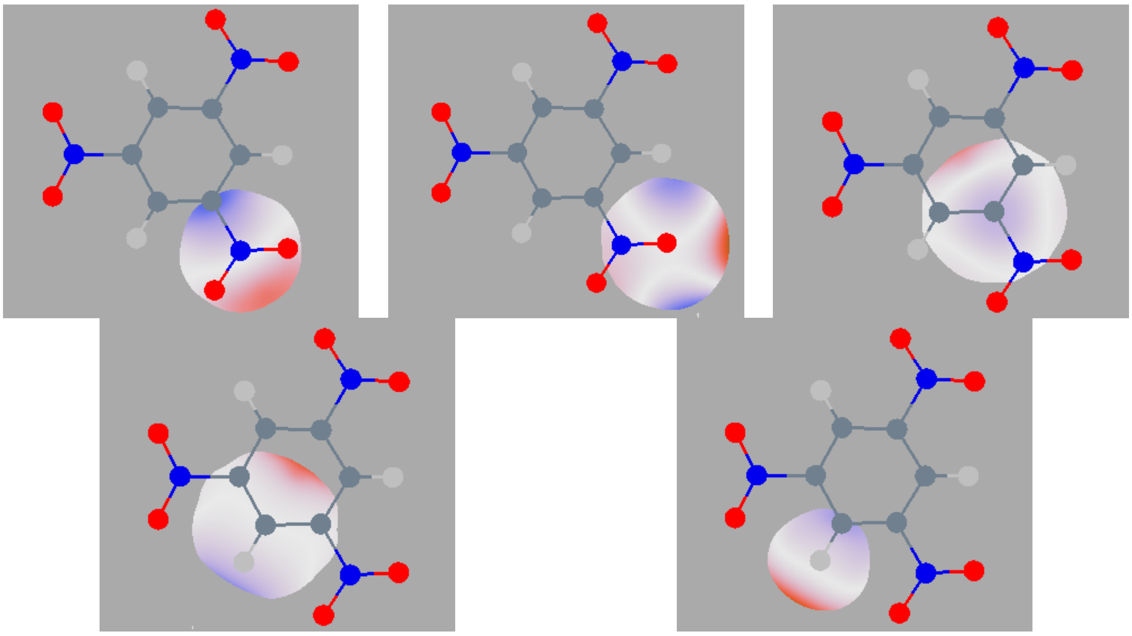


Figure 31: *An electrostatic potential (eV) map of the higher order multipole moments on the atoms of TNB to illustrate the shape of each unique atom in TNB. The map is computed using the ISA distributed multipole analysis (ISA-DMA) on the iso-density surface of* 10^{-3} $electron/bohr^3$ *around the each unique atom in the static optimized isolated molecular structure. On a potential scale of* $+0.4\ eV$ *(red) and* $-0.4\ eV$*(blue).*

5.3.2 The polarization model from atomic polarizabilities

To obtain the atomic polarizabilities within CamCASP one can either restart/continue the calculation from the ISA multipole moments or recalculate the Kohn-Sham molecular orbitals and their energies. For TNB the calculation was continued from the ISA multipole moments using the PBE0 functional alongside the CS00 asymptotic correction and the ALDA+CHF kernel in NWChem (an analogous level of theory to the multipole moments). In CamCASP, we calculate an accurate charge-density susceptibility $\alpha(\mathbf{r},\mathbf{r}')$ (**2.6.1.1**), for the optimized TNB molecule. The charge-density susceptibility is then distributed between atoms to give the atomic polarizabilities $\alpha_{tt'}^{ii'}$, which can be used to obtain the change in multipole moment $\Delta Q_t^{i'}$ on atom i' in response to the net electric field V_t^i on atom i. CamCASP initially calculates the non-localized polarizabilities, which can then be used directly to calculate dispersion energies using the dispersion program within CamCASP. In a non-local model, the electric field at one nucleus can induce a change in the multipole moments of other atoms in the same molecule. This methodology is very accurate and includes charge-flow effects, which are important for low-dimensional delocalized systems like 1-D wires or conjugated polymers,[383] but also a very complicated and elaborate model. Additionally, it is very computationally expensive, obtaining the dispersion term requires a quadruple sum over all sites. Consequently, the non-localized polarizabilities are localized and refined in order to obtain a simpler, cheaper localized model.

In previous studies like that of $C_6Br_2ClFH_2$,[60] the distributed molecular properties were obtained using the Williams-Stone-Misquitta (WSM) scheme.[38, 60, 165, 227, 287] The WSM scheme refines the polarizabilities using a combination of CamCASP and molecular electronic structure program like DALTON[378] or NWChem to do a SAPT(DFT) calculation of the local (point-to-point) responses, which is the change in potential at points in an array around the molecule in response to a point charge at any of the points. In this study we derive our model using the ISA-Pol method, this is analogous to the WSM scheme in many ways. However, with the ISA-Pol method, the main difference is the ISA-A calculation used to compute the multipole moments for the electrostatic model is then fed into the distributed polarizability module in CamCASP, which is then utilized to derive the local polarizabilities of the molecule. The main benefit of this scheme is that not only is it faster and more efficient[38] but also able to refine the final polarization model into a form suitable for our models; whether that be an intricate anisotropic Rank $l = 4$ model or a simpler but still accurate dipole-dipole model using isotropic polarizabilities. As with the pyridine study, this study of TNB uses a Rank $l = 1$ polarization model, which ignores the effects of induced higher moments like induced quadrupoles but is more transferable between simulation codes.

$Atom\ type$	$\left|\alpha_{iso}^{ab}\right|$	$\alpha_{10,10}^{ab}$	$\alpha_{11c,11c}^{ab}$	$\alpha_{11s,11s}^{ab}$
O	8.11	8.30	5.26	5.07
N	1.50	9.42	11.66	1.30
C_N	1.96	13.18	9.76	3.21
C_H	13.23	4.42	22.58	4.72
H	2.67	1.75	0.49	1.07

Table 19: *The symmetry unique atomic polarizabilities for each like-like atom type pair in TNB in atomic units up to Rank $l = 1$ (within a local axis system). The off-diagonal terms are not included, but the modulus of the isotropic polarizabilities ($\left|\alpha_{iso}^{ab}\right|$) has been included for comparison as they are closely related to the C_6 dispersion term. The ORIENT atomic polarizabilities input file for TNB can be found* ***Appendix 5.A*** *in atomic units.*

For O, N & C_N, we see the magnitudes of polarizabilities of in-plane of the bonds ($\alpha_{10,10}^{ab}$) and along the bonds ($\alpha_{11c,11c}^{ab}$) is larger than those out of plane. As another sanity check the modulus of the ISA-derived isotropic polarizabilities should be closely related to their atomic dispersion coefficient counterparts (***Table 20***), which is indeed observed. The $\left|\alpha_{iso}^{ab}\right|$ on oxygen is larger than that of nitrogen and appropriately its $C_6^{\iota\kappa}$ dispersion coefficient is also larger. Equally the isotropic polarizabilities for the homogenous pairs of both N and C_N are similar in magnitude and hence their $C_6^{\iota\kappa}$ dispersion coefficients are close.

5.3.3 The dispersion model: Dispersion Coefficients

The atom-atom dispersion coefficients are computed by integration over the distributed polarizabilities calculated at imaginary frequency.[285] Formerly for pyridine,[38] these were then localized utilizing the Williams-Stone-Misquitta (WSM) scheme[60, 285, 384] but the TNB model employs the ISA scheme to obtain dispersion coefficients. Deriving the dispersion coefficients from the atomic polarizabilities using the iterated stockholder atoms (ISA) approach reduces the coefficients to products of the individual-molecule properties, allowing the dispersion energy coefficients to be calculated just once and reduces dependence on the intramolecular geometry.[164, 385] In **Chapter 4** the charge density was found to depend heavily on the intramolecular geometry,[168] but derived terms like the dispersion coefficients may change with conformation but they are unlikely to change significantly under the ISA methodology.

$\iota-\kappa$	Dispersion coefficients (eV Å^{2n+4})		
	$C_6^{\iota\kappa}$	$C_8^{\iota\kappa}$	$C_{10}^{\iota\kappa}$
$O-O$	15.23	34.10	449.52
$N-N$	3.72	32.18	503.00
C_N-C_N	4.00	26.84	1310.59
C_H-C_H	26.86	50.89	2302.19
$H-H$	1.78	3.64	8.86

Table 20: *The isotropic dispersion coefficients ($C_{2n+4}^{\iota\kappa}$) for like-like atom type pairs in TNB in units of eV Å^{2n+4} up to n = 3. The full list of dispersion coefficients can be found in the complete potential input file in* ***Appendix 5.A****.*

Similar to the polarization model, the functional form of the dispersion models can be changed to suit our force-field. By limiting the atomic polarizabilities to dipole-dipole, the classic isotropic C_6 model could be chosen or higher-order polarizabilities can be used to yield dispersion coefficients up to C_{12}/R^{12}. The non-empirical model for TNB uses an isotropic C_{10}/R^{10} dispersion model analogous to that of pyridine.[167] Atoms are assumed to be spherical, this approximation is necessary as anisotropic dispersion models cannot be used in DMACRYS. Furthermore, the dispersion is kept isotropic for the sake of simplicity as the issues and complexities associated with an anisotropic model currently outweighs its benefits.[223]

$$V_{disp}^{MN} = \sum_{i\in A, k\in B} -f_6(\beta_{AB}R_{ik})\frac{C_6^{\iota\kappa}}{R_{ik}^6} - f_8(\beta_{AB}R_{ik})\frac{C_8^{\iota\kappa}}{R_{ik}^8} - f_{10}(\beta_{AB}R_{ik})\frac{C_{10}^{\iota\kappa}}{R_{ik}^{10}}$$

99

where M and N are two interacting molecules, i and k are the interacting atoms of type ι and κ in molecules M and N, and C_n are the calculated dispersion coefficients.

5.3.4 Damping

This study uses a single parameter (β_{MN}) Tang-Toennies damping function[271] (**2.8**) that tends to 0 as $R \to 0$, and 1 as $R \to \infty$. The Tang-Toennies damping function is given as

$$f_n(\beta_{MN}R_{ik}) = 1 - \exp(-\beta_{MN}R_{ik})\left(\sum_{m=0}^{n}\frac{(\beta_{MN}R_{ik})^m}{m!}\right)$$

100

where β_{MN} is the damping parameter for a molecular pair M and N. The issues with a single damping parameter are discussed in **Chapter 2** and **5.5.4** but as this is the only damping function accepted by DMACRYS, we must use it. The electrostatic expansion is not damped because we include the penetration energy in the short-range portion of the potential (**2.9**), however, the dispersion and polarization energies are. The final dispersion (β_{disp}), and polarization damping parameters (β_{pol}), are obtained by sampling important TNB dimers and calculating their second-order SAPT(DFT) dispersion ($E_{DISP}^{(2)}$) and induction ($E_{IND}^{(2)}$) energies, discussed in further detail in **5.5.4**. A preliminary guess of β_{disp} is initially derived from the ionization potential of the two monomers M and N.[273]

$$\sqrt{2I_M} + \sqrt{2I_N} = \beta_{disp}$$

101

Using TNB's experimental IP of 0.4028 a.u.[380], $\beta_{disp} = 1.8$. This method is used to obtain the dispersion damping parameter for the initial model derived from the first-order SAPT(DFT) interaction energies, which is used for the basin-hopping calculations (**5.5.1**).

5.4 Calculating the first-order short-range energy

5.4.1 Decomposing the short-range energy

Once all damped long-range contributions have been removed from the total interaction energy we're left with the short-range contribution E_{SR}, which can be fit to a sum of exponentials using the density overlap model[274] (**5.4.3**). In **Chapter 2** a detailed explanation is given on the derivation and role of the short-range energy in intermolecular potentials. For TNB, we model the short-range energy within a force-field using the following functional form

$$V_{SR} = \sum_{i\in M, k\in N} G\exp\left[-\alpha_{00}^{\iota\kappa}\left(R_{ik} - \rho^{\iota\kappa}(\Omega_{ik})\right)\right] \tag{102}$$

In the previous study of pyridine[38, 167] higher order induction terms were included as δ_{int}^{HF} but in our study of TNB we do not include this term as it is small and expensive to compute (**5.5.1**). E_{SR} is obtained by only calculating the 1st and 2nd-order SAPT(DFT) interaction energies of a large number of different dimer configurations, and then the short-range exponential in Equation 102 is fit to E_{SR} to obtain the term V_{SR}. As seen in ***Figure 31*** some of the atoms of TNB are very anisotropic, thus, it is important to ensure V_{SR} is anisotropic to better capture atomic shape.

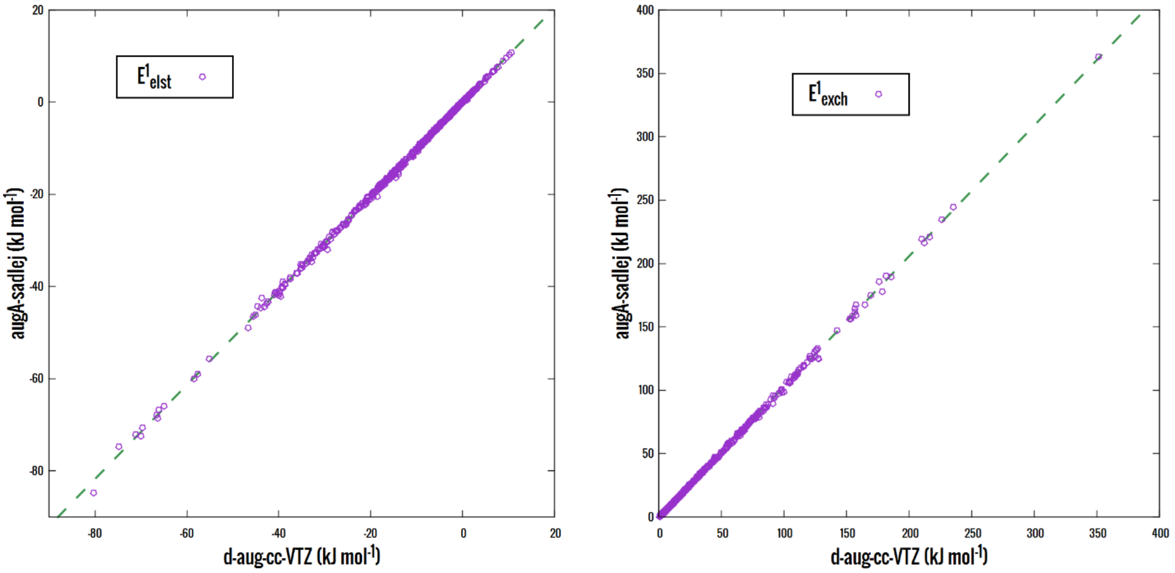
5.4.2 Obtaining the first-order SAPT(DFT) energies

$$E_{int}^{(1)}[SAPT(DFT)] = E_{exch}^{(1)} + E_{elst}^{(1)} \tag{103}$$

The $E_{int}^{(1)}[SAPT(DFT)]$ dimer interaction energies were computed for 2000 dimer pseudo-random configurations of various orientations and varying relative centre of mass distances R. For a non-biased 1st order fit, the dimer configuration space needs to be sampled adequately but also fairly, which can be difficult even for a planar molecule that is held rigid. The space is sampled using the quasi-random Sobol sequence, and Shoemake's algorithm,[38] for sampling angular space (Shoemake uses quaternion space to keep sampling uniform yet random), implemented in the CamCASP program. In the case of two rigid planar molecules a six dimensional space $(R_x, R_y, R_z, N_x, N_y, N_z, \alpha)$ must be sampled. Using the angular axis system, R_α denotes the centre of mass of molecule M, while $N_\alpha$the centre of mass of molecule N and α the angle between the two. To maintain consistency with our distributed multipole moment ($V_{elst}^{(1)}[DM]$) calculations, the first-order SAPT(DFT) energy calculations used the PBE0 functional, the CS00 (Casida-Salahub) asymptotic correction,[382] with an AC-shift of $0.0539\ a.u.$, and the ALDA+CHF linear-response kernel. The main basis-set for computing the molecular wave-function was augA-sadlej, while the auxiliary and atomic-auxiliary basis sets were aug-cc-pVTZ. These are also identical to the basis sets used in the distributed moments calculations. The distributed electrostatic energy ($V_{elst}^{(1)}[DM]$) were then calculated for the 2000 dimer configurations using the Rank 4 ($l = 4$) ISA distributed multipoles moments obtained in **5.3.1** and subtracted from the non-expanded electrostatic energy $E_{elst}^{(1)}$ to give $E_{pen}^{(1)}$ for each dimer configuration (Equation 102). When calculating the $V_{elst}^{(1)}[DM]$, all basis-sets were monomer centred (MC type), however, a dimer centred (DC) type auxiliary basis-set is used for the density fitting step (**2.2.4**). The density fitting is used to fit the atomic densities ($\rho_i(\mathbf{r})$) to the calculated molecular orbitals.

5.4.2.1 Sanity Checks

One sanity check is to ensure all the exchange energies are positive, as $E_{exch}^{(1)}$ is a purely repulsive term. Another is to check the effect of basis set. For the first 1200 points generated, the $E_{elst}^{(1)}$ and $E_{exch}^{(1)}$ contributions were computed using the augA-Sadlej as their main basis, and they were also checked against energies computed using the larger d-aug-cc-VTZ (main) basis and d-aug-ccpVQZ (auxiliary and atomic-auxiliary) basis. This was to ensure that the smaller but faster and efficient augA-sadlej basis was producing reasonable energies (***Figure 32***). The single point at high energy in ***Figure 32*** (b) should not be considered an outlier. Higher energy points were included as more points were randomly generated.



(a) (b)

Figure 32: *(a)1st order electrostatic and (b) exchange SAPT(DFT) interaction energies for 1200 TNB dimers calculated using the augA–Sadlej basis set (the basis used for generating the non-empirical potential) and the larger d-aug-cc-VTZ basis set. Both terms show a line of best fit gradient of 1.0 (2 s.f.) and an average % difference of 5%, we find that the augA-Sadlej and d-aug-cc-VTZ energies are remarkably close.*

5.4.3 The overlap model

5.4.3.1 Simplifying terms using symmetry

In **Chapter 2**, we discuss the value of utilizing symmetry to simplify the number of terms in our force-field, and we see its implementation with pyridine in **Chapter 3**. With the optimized rigid structure of TNB being D_{3h} and the gas-phase dimers being identical in conformation, the number of total atom-types in our potential is greatly reduced from 18 to just 5 (***Figure 30***), consequently, there are 15 atom-atom type combinations. The symmetrisation process and the anisotropic terms for TNB are detailed in **Chapter 2** but we show the available anisotropic terms ***Table 21*** for ease of reference within this chapter.

K_1	K_2	L_1	L_2	J	Function	Term	Atoms
0	0	1	0	1	$\vec{z} \cdot \vec{R}$	10	All
1	0	1	0	1	$\vec{x} \cdot \vec{R}$	$11c$	$Oxygen$
-1	0	1	0	1	$\vec{y} \cdot \vec{R}$	$11s$	$None$
0	0	2	0	2	$(3(\vec{z} \cdot \vec{R})^2 - 1)/2$	20	All
1	0	2	0	2	$\sqrt{3}(\vec{z} \cdot \vec{R})(\vec{x} \cdot \vec{R})$	$21c$	$Oxygen$
-1	0	2	0	2	$\sqrt{3}(\vec{z} \cdot \vec{R})(\vec{y} \cdot \vec{R})$	$21s$	$None$
2	0	2	0	2	$\sqrt{3}\left((\vec{x} \cdot \vec{R})^2 - (\vec{y} \cdot \vec{R})^2\right)/2$	$22c$	All
-2	0	2	0	2	$\sqrt{3}(\vec{x} \cdot \vec{R})(\vec{y} \cdot \vec{R})$	$22s$	$None$

Table 21: *Allowed anisotropic terms in DMACRYS*[59]*, and the atoms in TNB that require them by symmetry. Here the terms are for the atom-type* ι *in an atom-atom pair* $\iota - \kappa$ *with the directionality* $\iota \rightarrow \kappa$*. The anisotropic term* $\rho^{\iota\kappa}(\Omega_{ik})$ *(Equation 102) is defined by an S function that can be defined by the terms* K_1, K_2, L_1, L_2, J*.For example, for an atom-atom pair* $\iota - \kappa$*, the 11c term on type* ι *can be defined as* $J, K_1, L_1 = 1, K_2, L_2 = 0$ *while the 11c term on type* κ *can be defined as* $J, K_2, L_2 = 1, K_1, L_1 = 0$.

5.4.3.2 Weighting and Fitting method

The theory behind the weighting and fitting treatments is detailed in **2.9.5**. There are a number of weighting schemes available within CamCASP (**2.9.5.2**), but for TNB, the Gaussian-log weighting[274] $w(e_p) = \exp\left[-\alpha\left(\ln(e_p/e_0)\right)^2\right]$ is used for the construction of the density overlap models, where the $\alpha = 1/\ln 10$ by default. This weighting function is very good for data that spans multiple orders of magnitude. e_0 is chosen depending on part of the potential energy surface one deems most important. For us a good sampling of the repulsive wall is ideal, therefore a large multiple of the absolute global minimum dimer energy was chosen. Its default value is 20 $kJ\ mol^{-1}$.

For these non-empirical models, the distributed density-overlap fits were performed using the Gaussian-log weighting scheme[386] (**2.9.5.2** Scheme 5), where e_0 was set to $100\, kJ\, mol^{-1}$, $\alpha = 1/\ln 10$, and the energy maximums and minimums E_{max}, E_{min} were set to $500\ kJ\ mol^{-1}$ and $0.0001\ \ kJ\ mol^{-1}$ respectively, in order to properly sample repulsive configurations. Energy values outside this range are given a weighting of zero. An overlap model constraint (λ) of $1x10^{-4}\ kJ\ mol^{-1}$ was used, which by default $\lambda = 10^{-8}$. However, the resultant short-range terms from tighter constraints were found to be very unphysical, and as a result $\lambda = 10^{-4}$ was used for TNB. The unphysical terms may have been a product of the poor sampling (**5.4.4.2**).

5.4.4 Relaxing the fit to obtain Model0-iso and Model0-aniso

Once the total density overlap values S_p^{ik} and fitted K^{ik} parameters have been fit to $E_{SR}^{(1)} SAPT(DFT)$ (**2.9.3**), one is left with a tentative model of both the isotropic and anisotropic short-range terms. For TNB, these initial potentials were dubbed Model0. As combining rules are not used the resultant potentials have a total of 15 atom-atom pair terms. The fitted energies using these initial parameters were then compared to the reference (calculated) energies by ORIENT[108] to determine the quality of the fit. For a good fit, the weighted root mean square deviation (wRMSD) between the reference energies and the fitted energies should be around 10-15%, with residual energies (***Table 22***) of under $10\ kJ\ mol^{-1}$. The wRMSD of both potentials after fitting was $\approx 11\%$.

5.4.4.1 Relaxing the short-range parameters to first-order SAPT(DFT) energies

After the initial fitting, the next step is allowing limited relaxation of the short-range parameters $\alpha_{00}^{\iota\kappa}$ and $\rho_{00}^{\iota\kappa}$ (Equation 102) to obtain our final potential (Model0) parameters. Both $\alpha_{00}^{\iota\kappa}$ and $\rho_{00}^{\iota\kappa}$ are heavily correlated; the value of $\alpha_{00}^{\iota\kappa}$ is dependent on $\rho_{00}^{\iota\kappa}$ (and $\rho^{\iota\kappa}(\Omega_{ik})$ if the potential is anisotropic) and vice-versa. Ideally one would relax only $\rho_{00}^{\iota\kappa}$ as our anisotropic terms are also dependent on $\alpha_{00}^{\iota\kappa}$. $\alpha_{00}^{\iota\kappa}$ directly affects the slope of our exponent (Equation 102), thus changing this would result in the most dramatic change in our potential but could also reduce the physicality of our ρ parameters. In order to obtain the best fits, the short-range parameters in the 1st order models are relaxed simultaneously. Normally, ORIENT calculates $\alpha^{\iota\kappa}$ and $\rho^{\iota\kappa}$ for each atom pair and then for all atoms simultaneously, under the presumption that all terms are correlated. In this study the total correlated constants have not been calculated as certain atoms cannot be sampled well enough; this also has other consequences, which are discussed in further detail in **5.4.4.2**.

Consequently, the anchors on $\alpha_{00}^{\iota\kappa}$ were made tighter (set to 1.0) than those on the isotropic $\rho_{00}^{\iota\kappa}$ (set to 0.1), while the anchors on the anisotropic terms $\rho^{\iota\kappa}(\Omega_{ik})$ are set to 1.0. The anchors on the anisotropic terms are made tight as their initial approximation, with the overlap model, is reasonably good. We find the anisotropic terms

(**Appendix 5.A**) to be very closely related to the distributed multipole moments, as expected as both represent the non-sphericity of the atomic charge distribution. This is because the overlap model uses the atomic charge density to obtain the short-range terms, consequently, even the initial approximation is very close to reality.

The relaxation step within CamCASP occurred in stages as opposed to simultaneously relaxing all parameters due to insufficient data and sampling of the carbon and hydrogen repulsive walls. Therefore, it was wise to start with the most important atoms first (i.e. oxygen) and end with the least important (hydrogen). By employing chemical intuition and also investigating the higher order multipole moments (***Figure 31***) of each atom in TNB, one was able to determine which atoms would be the most important to the short-range potential. This was based off their anisotropy (atom-shape), atomic charge density (electronegativity) and how well they had been sampled by the dimer configuration space. ***Figure 31*** and ***Table 18*** suggest that the most important atoms are N and O, whereas H is the least important to the molecular anisotropy.

5.4.4.2 Analysis of the final first-order isotropic and anisotropic fits

In order to see how well the $E_{SR}^{(1)}$ dataset was sampling intermolecular interactions, what improvements had been made on the isotropic potential and also to whittle down terms that were not important and atoms that should be kept isotropic, a fully anisotropic potential was first created. As mentioned in **5.4.3.1**, certain functions used to define many of the anisotropic terms can be ruled out due to symmetry. In addition, some atoms can be made isotropic if their repulsive wall is only sampled in a limited range of orientations because most of their atomic charge density is within the repulsive wall of other atoms. This is the case with the hydrogens attached to the aromatic carbons and even the carbons themselves due to the anisotropy of the NO_2 groups.

Potential	Weighted RMS $(kj\ mol^{-1})$	O largest aniso-term	N largest aniso-term	C_N largest aniso-term	C_H largest aniso-term	H largest aniso-term
Isotropic	1.157					
O only	0.852	-0.108 (ρ_{20})				
N & O only	0.768	-0.120 (ρ_{20})	-0.078 (ρ_{22c})			
CN, N & O	0.979	-0.096 (ρ_{20})	-0.051 (ρ_{20})	-0.331 (ρ_{22c})		
CN, CH, N & O	0.964	-0.098 (ρ_{20})	-0.046 (ρ_{20})	-0.329 (ρ_{22c})	-0.056 (ρ_{22c})	
Fully anisotropic	1.026	-0.080 (ρ_{20})	-0.057 (ρ_{20})	-0.353 (ρ_{22c})	0.068 (ρ_{20})	-0.262 (ρ_{20})

Table 22: *The weighted RMS residual energies in* $kJ\ mol^{-1}$ *for the various 1st order short-range models and the largest symmetry allowed term in* $\rho^{\iota\kappa}(\Omega_{ik})$ *other than* $\rho_{00}^{\iota\kappa}$*, in Bohr for each atom type*

A methodical way of whittling down which terms are truly needed is by adding anisotropy to each atom (starting with the most important atom, oxygen) on a term by term basis and observing the effect on the short-range terms (***Table 22***). Atoms showing unphysical anisotropic terms should be made isotropic. Typically, the weighted RMSD is a very good indicator of whether anisotropy is improving the quality of a short-range fit and generally if the RMSD does not decrease then adding anisotropy to an atom could be detrimental to the overall potential. As all anisotropic fits in ***Table 22*** were started from the isotropic best fit, it is possible there were multiple minima and the quality of a fit could have been improved if it was not well determined.

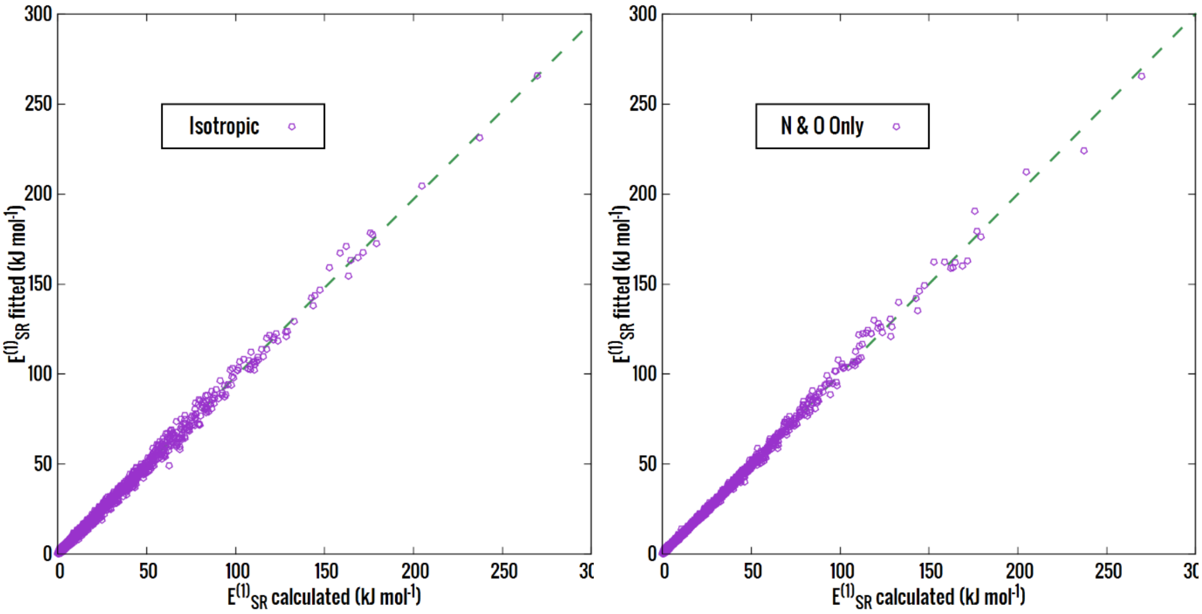
From the outset, making only the oxygen anisotropic improves the fit considerably, significantly reducing the weighted residual error from 1.16 to 0.85. This drastic improvement in quality highlights the importance of including shape in the highly anisotropic oxygen atoms. We also find the ρ_{20} term to be the most anisotropic. ρ_{20} describes the rank 2 term along the bond, and its dominance agrees with the chemistry of TNB and the multipolar moments, where the d_{z^2} moment is the largest (in the local axis). A substantial portion of the molecular charge density can be found on the electronegative NO_2 groups and therefore they are heavily involved in the intermolecular interactions that govern crystal structures and the stability of various polymorphs, especially under conditions that sample higher up the repulsive wall. As a result, accurately capturing the shape of these atoms is crucial if one wants to reliably predict the structure of weakly bound organic energetics at elevated temperatures and pressures. In **Table 22** one can see the magnitude of anisotropy on ρ_{20} decreases as we include anisotropy on more atoms, however, this still remains the most dominant anisotropic term, for both oxygen and nitrogen. Including the symmetry allowed anisotropy terms on the nitrogen further improves the quality of the fit, with very realistic final parameter values. Interestingly, in the fully anisotropic model ρ_{20} is the most dominant anisotropic term for all atomic types except for C_N. The nitrogen and oxygen atoms are sampled very well in the dimer configuration space used for the $E_{SR}^{(1)}$ calculations emphasizing the importance of these atoms in gas-phase interactions.

In **Table 22**, one can see that the anisotropic terms on the central atoms (C and H) are unphysically large. The ρ_{20} term on C_H should certainly not be larger than the oxygen atom, which suggests C_H was poorly sampled. Judging from these results, C_N, C_H and H were also not well sampled/defined within our configuration space. This is likely due to these atoms being blocked by the electron rich NO_2 groups. Upon close inspection their anisotropic terms are very unphysical and are simply numerical values produced by our model. For instance, much of the anisotropy on the hydrogen should be engulfed by the connecting carbon as the charge density on C_H is substantially larger, however, we find the ρ_{20} on the hydrogen to be enormous (almost 4 times larger than the oxygen) in our completely anisotropic potential. Most importantly, this greatly influences the anisotropy of our oxygen atoms. There is a dramatic decrease in the final ρ_{xx}^{O} parameters for oxygen as the model attempts to spread the anisotropy between the less well-defined carbon and hydrogen atoms. Chemical intuition says that oxygen and nitrogen should be the most anisotropic and are the most important in intermolecular interactions in the lattice, therefore, the fits in ***Table 22*** that include shape on the aromatic atoms are not as chemically true. Even though anchors on the anisotropic terms were held tight during the relaxation stage, the preliminary fitting stage the overlap model was unable to prescribe realistic values to the anisotropic $\rho^{\iota\kappa}(\Omega_{ik})$ of the central atoms.

	Anisotropic terms				
Atom-Type (ι)	ρ_{10}^{ι}	ρ_{11c}^{ι}	ρ_{20}^{ι}	ρ_{21c}^{ι}	ρ_{22c}^{ι}
O	0.05	-0.05	-0.12	0.03	0.09
N	0.03	0.00	-0.02	0.00	-0.08

Table 23: *The values (in atomic units) of the anisotropic short-range terms $\rho^{\iota\kappa}(\Omega_{ik})$ used in the final anistropic Model0. Only the nitrogen and oxygen atom types were made anisotropic for reasons discussed in **5.4.4.2**. The complete potential file can be found in **Appendix 5.A**.*

To conclude, despite the ISA atomic iso-density contours for all atoms being anisotropic (***Figure 31***), the final Model0-aniso only includes anisotropy on the oxygen and nitrogen atoms in TNB as the anisotropy of the central atoms is not well determined. In ***Table 23*** the values used are listed. The largest anisotropic term on oxygen ρ_{20}^{ι}, which is 2nd order along the bond, correlates well with the atomic shape seen in ***Figure 31*** and the largest multipole moment Q_{22c} (***Table 18***).



(a)

(b)

Figure 33: *SAPT(DFT) calculated* $E_{SR}^{(1)}$ *energies contrasted against the fitted energies for the (a) Model0-iso and (b) Model0-aniso with anisotropy on the N & O only. The weighted root mean squared (RMS) residuals energies were found to be* $1.16\ kJ\ mol^{-1}$ *and* $0.77\ kj\ mol^{-1}$ *respectively.*

The plots of the final relaxed energies and the SAPT(DFT) calculated energies (***Figure 33***) suggest one can obtain a very good isotropic fit. A short-range potential with a final weighted RMS residual energy of $2\ kJ\ mol^{-1}$ would be considered satisfactory, thus our isotropic model with a weighted RMS of $1.16\ kJ\ mol^{-1}$ is very good. As seen in ***Figure 33*** and ***Table 22***, including atomic anisotropy can substantially reduce the weighted RMSD though the improvement in the quality of the fit is not incremental. It should be stressed that, while anisotropic terms may be transferred between atom pairs (a single $\rho_{10}^{\iota\kappa}$ for all oxygen pairs), $\alpha_{00}^{\iota\kappa}$ and $\rho_{00}^{\iota\kappa}$ depend on both atom types. Accordingly, the anisotropic terms used in the Model0 is identical to those used the final DIFF model.

5.4.4.3 Does sampling higher up the repulsive wall improve short-range parameters?

A possible way to improve on the sampling of the central atoms is to attempt to sample more repulsive structures where the dimers are closer together and include this data in our fitting. By setting a more repulsive R_{min} and R_{max} when generating the pseudo-random configurations, one can force the dimers closer together with the hope that the atomic shape of atoms on the aromatic ring will be better sampled. Hence, the 1st order SAPT(DFT) energies and distributed overlap integrals for a supplementary 2000 “tighter” configurations were calculated using the 2000th configuration from the initial data set as a starting point, meaning no orientations were repeated. However, including all 2000 new structures was found to heavily bias our model towards more repulsive structures, which pulled the potential energy surface minimum up. There was also a dramatic increasing in the percentage error between our fitted and calculated intermolecular perturbation theory energies as the range of energies increased significantly. One could argue that the cut-offs of $E_{SR}^{(1)}$ could have been tighter, but it is the total energy that should be scrutinised. Instead one can use high quality 2nd order SAPT(DFT) dimer energy calculations on the most important dimers on the PES to relax the short-range parameters and increase the reliability and accuracy of the non-empirical models. Which leads us to the question of how does one select the most important configurations to sample on a PES?

5.4.5 Summary of the first-order models and discussion of subsequent steps

The above methodology results in an isotropic and anisotropic analytical model for describing short-range interactions using $E_{SR}^{(1)}$ energies, which can be used to model the TNB crystalline state in simulations. The anisotropic and isotropic Model0s can be combined with the previously derived long-range terms (**5.3**): the distributed multipoles, polarizabilities, and isotropic dispersion coefficients. Using Equation 101 a dispersion damping parameter can be approximated, while an appropriate polarization damping parameter can be assumed and refined after testing the model. This would give a non-empirical force-field that could be used in simulations, and validated with comparisons to experimental data. In this thesis, we focus on developing accurate and reliable potentials for energetic molecules, which can reliably sample the true potential energy surface. Accordingly, one should evaluate the precision of these models against high quality theoretical data, like the 2nd order SAPT(DFT) ($E_{int}^{(2)}[SAPT(DFT)]$) interaction energies. There are four advantages of analysing the potential against higher-order energies:

1. To assess whether certain contributions can be omitted. The highly non-additive many-body terms, which arise from interactions that cause changes in charge distribution (2nd or higher order terms), require very careful consideration when transferring them from gas-phase dimer calculations to the solid-state.
2. The damping parameters can be determined by fitting to 2nd order SAPT(DFT) energies. Approximating these may not be suitable for larger molecules like TNB as it is likely they do not follow the trends found for small near-spherical molecules.[38]
3. Knowing $E_{int}^{(2)}[SAPT(DFT)]$ can allow for the refinement of some of the short-range parameters to absorb errors or approximations.
4. Further tests can be made on the accuracy of the model's PES and its range of validity. This is a key consideration, as the pseudo-random sampling method of dimer configurations used to obtain the $E_{SR}^{(1)}$ dataset may not be adequate enough for highly anisotropic molecules. The repulsive wall around each atom must be sampled as well as it could be in any physical situation (e.g. under pressure or at ambient conditions), however for some atoms, like H in TNB, its repulsive wall is buried beneath the charge density of the larger atoms, or like the carbon atoms in TNB, can only be accessed/sampled at a few orientations. In **5.4.4.3**, R_{min} and R_{max} are changed in order to generate an additional 2000 tighter configurations, but this still fails to improve the quality of the potential. Without considering the total interaction energy, it is hard to confidently conclude that the relevant areas of the intermolecular potential energy surface of TNB has been sampled.

The testing and refining of a non-empirical model is iterative and interlinked. Thus, the subsequent section starts with an investigation of the higher order induction terms δ_{int}^{HF} and $E_{CD}^{(2)}$. Can these terms be neglected from the total intermolecular energies? We investigate a viable method of calculating $E_{int}^{(2)}[SAPT(DFT)]$, albeit at considerable computational expense, and obtaining $V_{int}[Model]$. $E_{int}^{(2)}[SAPT(DFT)]$ was calculated for three distinct datasets:

1. A set of gas-phase dimer configurations obtained using an approach analogous to that used for pyridine.[38] The minimum energy configurations of the TNB dimer were located, and then scanned by varying the centre of mass distance R in order to determine the model's damping parameters.
2. A new set of pseudo-random configurations were generated to sample considerably more configuration space. This is crucial for describing condensed phase properties.
3. The most common two-molecule contacts found in TNB crystal structures generated in a preliminary CSP study. The models were not parameterised to this dataset; it was only used as a test-set the force-fields.

5.5 Testing and improving the models using the total SAPT(DFT) dimer interaction energies

5.5.1 Omitted Terms

5.5.1.1 3rd to infinite order contributions to the induction

For pyridine,[38] the total SAPT(DFT) energy was calculated in two-steps:

1. Calculations of the second-order interaction energy using SAPT(DFT).
2. Calculations of the δ_{int}^{HF} correction energy, which is needed for third to infinite-order contributions to the induction (**2.6.3**).

For the pyridine potential, the total two-body SAPT(DFT) interaction energy (including higher up to infinite order terms) was used and may have had detrimental effects to the resultant force-field (**3.4.5**). The decomposition of the total intermolecular interaction energy and the methods of describing the higher order energy terms is elaborated in **Chapter 2**. The δ_{int}^{HF} correction energy was calculated for a single gas-phase minima of TNB (***Figure 34***) and the higher order induction contributions to the total energy were found to be small. δ_{int}^{HF} was found to be approximately 1 $kJ\ mol^{-1}$ around the minimum, which is less than 4% of the total interaction energy. While this could be considered a substantial contribution to the energy, it is such a costly energy to compute (approximately 16 – 20 hours per point) that its cost currently outweighs the benefits of inclusion, but this is not always the case.[38, 46, 167] The inclusion of the δ_{int}^{HF} correction energy has been found to be detrimental to non-empirical potentials for organic molecules, where higher order induction contributions are small.[167] For the non-empirical model of pyridine (**Chapter 3**), δ_{int}^{HF} coupled with other effects, could have been the cause of a slightly over-binding potential.[167]

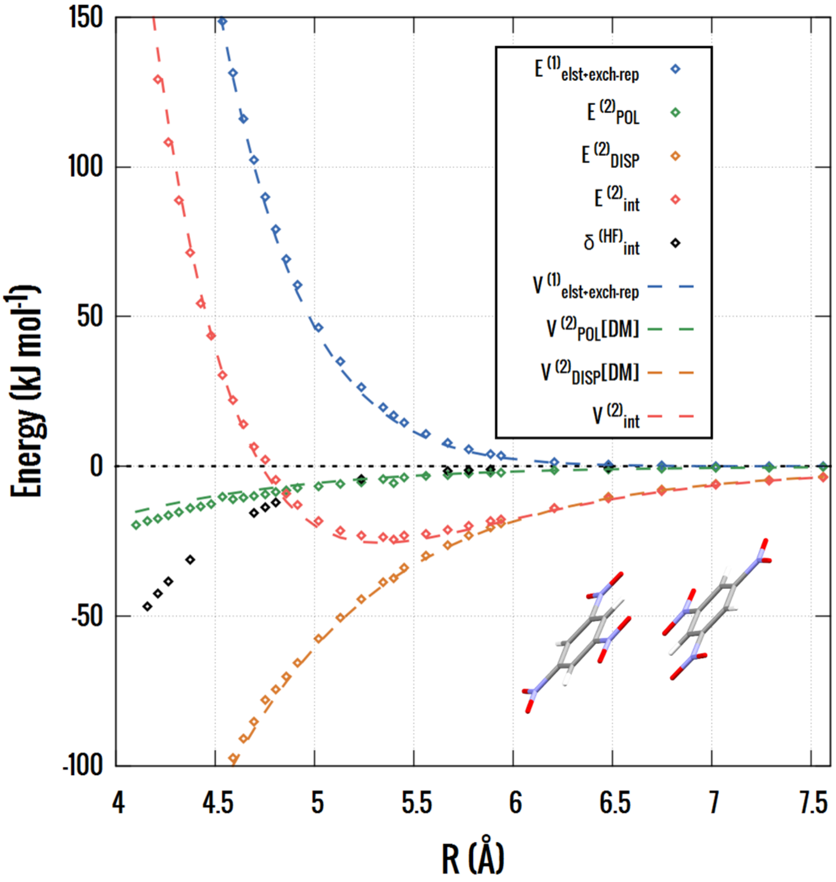


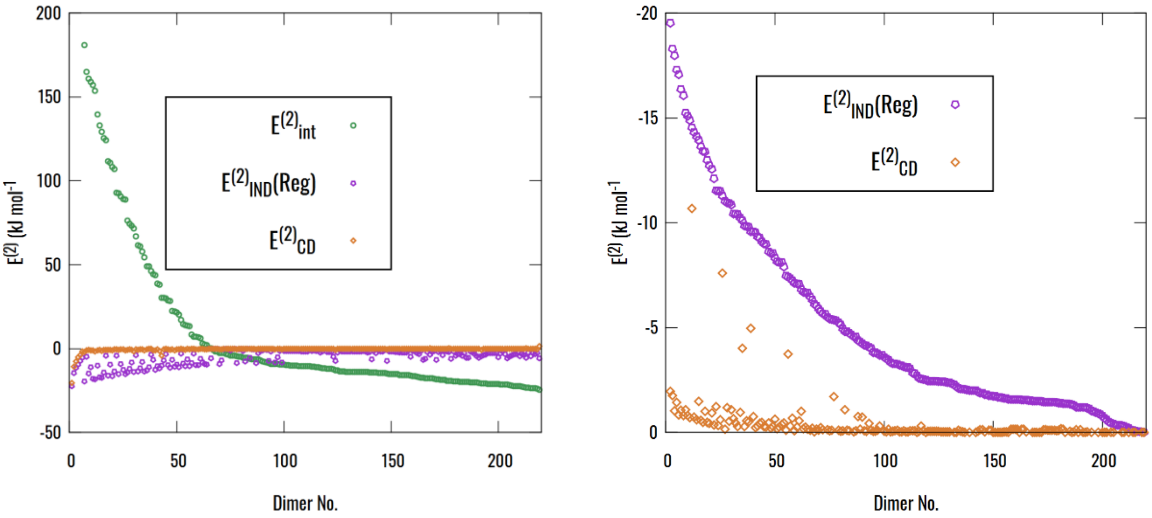
Figure 34: *$E_{int}^{(2)}$ SAPT(DFT) energy breakdown including the δ_{int}^{HF} correction energy (points) and comparison with the Model0-aniso (dashes) for the S3 dimer. $\beta_{disp} = 1.65$ and $\beta_{pol} = 1.00$*

5.5.1.2 Defining the polarization energy and neglecting the second-order charge-delocalization energy

The charge-delocalization term is typically a stabilizing energy that describes the delocalization of charge density between two molecules in close proximity (**2.6.2**). When partitioning the SAPT(DFT) intermolecular interaction energy of a system, the charge-delocalization energy is said to be the difference between the 2nd order induction energy ($E_{IND}^{(2)}$) and the 2nd order polarization energy ($E_{POL}^{(2)}$), which is also the regularized induction energy ($E_{IND}^{(2)}(Reg)$). The charge-delocalization energy's origins and nature are discussed in further detail in **2.6.2**.

$$E_{CD}^{(2)} = E_{IND}^{(2)} - E_{IND}^{(2)}(Reg) \tag{104}$$

If the contribution of $E_{CD}^{(2)}$ to the induction energy is significant enough it can be absorbed into the short-range exponential term by relaxing $\alpha_{00}^{\iota\kappa}$ and $\rho_{00}^{\iota\kappa}$. However, after computing $E_{CD}^{(2)}$ using Equation 104, we find that only the very repulsive configurations ($E_{int}^{(2)} > 150\ kJ\ mol^{-1}$) have notable $E_{CD}^{(2)}$ energies (***Figure 35*** (a)). Generally, the charge-delocalization is tiny and in most cases negligible in comparison to the polarization energy (***Figure 35*** (a)). As a result, $E_{CD}^{(2)}$ is not included in our non-empirical models to reduce the number of approximations/energies the models absorb. In ***Figure 35*** (a) one can see the importance of the polarization energy ($E_{IND}^{(2)}(Reg)$) as its contribution to the 2nd order interaction energy is small but still considerable, especially for the more repulsive structures. ***Figure 35*** (b) further emphasises that the TNB induction energy is dominated by polarization contributions, at least for the majority of configurations seen the gas-phase dimers.



(a) (b)

Figure 35: *(a) The 2nd order SAPT(DFT)intermolecular interaction energy* $E_{int}^{(2)}$ *compared with the components of the 2nd order induction energy, the regularized induction energies* $E_{IND}^{(2)}(Reg)$ *(also known as* $E_{POL}^{(2)}$*) and the charge-delocalization energy* $E_{CD}^{(2)}$ *energies for 200+ gas-phase dimer configuration scan and crystal dimer configurations. The dimers have been ordered by decreasing interaction energies* $E_{int}^{(2)}$*. (b) Showcases the dominance of the polarization energy (*$E_{IND}^{(2)}(Reg)$*) compared to the charge-delocalization energy for the same dimer configurations. In (b) the dimers have been ordered by increasing polarization energies.*

5.5.2 Calculating second-order SAPT(DFT) energies

5.5.2.1 Computational considerations

The remaining 2nd order terms can be included by fitting of our force-field to a grid of numerically dense pair interaction energy calculations (**5.4.5**). The initially proposed methodology for calculating the higher order SAPT(DFT) energies of TNB dimers was analogous to the treatment previously used on the 11-atom pyridine, where higher order calculations including calculation of the 2nd to infinite order induction terms were carried out using solely CamCASP on over 1000 different configurations.[38] However, TNB is a substantially larger molecule with 18 atoms. 2nd order energy calculations that took 2-3 hours per dimer for pyridine in CamCASP can take 36+ hours per TNB dimer. What previously took approximately 2000 hours of computing time would take a whopping 40000 hours. An alternative is to do the complete higher-order calculation on the still-evolving quantum code Psi4.[228] The scaling with molecule size is far better with Psi4 as it does well in partitioning and allocating computer memory, and is more efficient when handling large calculations. As it is relatively new it makes use of the latest treatments to be a substantially faster and efficient code. For example, using the augA-sadlej basis the cost per point is approximately 15 hours, more than half as quick. Conversely, as it is a new code there are still bugs and quirks that need to be worked around, especially for interfacing with other codes. Accordingly, many bespoke python codes, scripts and methodologies had to be developed alongside, in order to analyse and readily employ the Psi4 outputs in our codes. It should be noted that the total SAPT(DFT) interaction energies calculated by Psi4 needed rescaling as the exchange dispersion term $E^{(2)}_{exch-disp}$ in the code is unscaled. Furthermore, Psi4 currently does not include the ALDA+CHF linear-response kernel that is used in the molecular properties calculations. Instead the code uses the ALDA kernel which is more suited for PBE calculations.[284] Consequently, we used a combination of Psi4 to obtain our molecular wave-functions and CamCASP for the SAPT calculations.

5.5.2.2 Methodology

The 2nd order energy calculations were carried out on geometries obtained from previous configuration scans (**5.5.3.1.3**), pseudo-random dimer configurations (**5.5.3.3.1**) and the most popular two molecule contacts found in TNB crystal structures (**5.5.3.2**). A combination of Psi4[228] and CamCASP was used to calculate these interaction energies. The calculations were once again performed using the PBE0 functional, the CS00 asymptotic correction,[382] with an AC-shift of $0.0539\ a.u.$, and the ALDA+CHF linear-response kernel in order to maintain consistency with the 1st order SAPT(DFT) energy calculations. The main basis-set for computing the molecular orbitals was the augA-sadlej basis in the MC+ (monomer centred with mid-bonds) format with a 3s3p2d2f mid-bond set. The auxiliary basis sets used was the aug-cc-pVTZ in the DC+ format (dimer centred with mid-bonds). Psi4 is used to calculate the wave-functions for each dimer configuration, while CamCASP uses these orbital energies to complete the SAPT portion of the calculation. A bespoke Python[31] code was used to extract the dimer geometries of all the configuration scan orientations and convert them into the correctly formatted CLUSTER input files for the interaction energy calculations. Another homemade script was then used to extract the appropriate energy terms and convert them into $kJ\ mol^{-1}$.

5.5.3 Dimer configuration datasets

5.5.3.1 Gas-phase global and local minima and configuration scans

5.5.3.1.1 Basin-Hopping

Basin-hopping is a two-part, iterative global optimization treatment, that employs a global stepping algorithm with local minimizations at each step.[387] It is a simulation technique in the program ORIENT,[108] used to find the global and local minima on a potential energy surface (PES) of a homo/hetero-dimer. The limited-memory BGFS (Broyden–Fletcher–Goldfarb–Shanno) aka L-BFGS optimization algorithm[388, 389] is used by ORIENT to find local and global minima of an intermolecular energy model. Basin-hopping typically shines when used on "bumpy" PESs and was used as a preliminary process for choosing important isolated dimers for $E_{int}^{(2)}[SAPT(DFT)]$ calculations for the further refinement of the short-range terms.

For the TNB calculations, the maximum number of steps used was 100,000, the maximum displacement $\Delta x = 0.10$ of each step and a maximum rotation $\Delta\theta = 30°$. An energy corresponding to kT at a temperature of $500K$ in kelvin was used as the criterion/instructions for allowing or rejecting a possible step. Basin-hopping simulations run either until a specified number of successful steps have been taken or a certain number of minima have been found. A basin-hopping run was initially carried out using the isotropic Model0 and contrasted with the results for the same run using the empirical FIT[371] potential for the dispersion-repulsion contributions alongside an identical electrostatic model as Model0. This was to check the sensibility of our model.

The program CLUSTER which is a part of the suite of programs CamCASP,[107] is then used to analyse the minima found by the basin-hopping simulation. CLUSTER re-orientates the geometries into their principle axis and similar minima are clustered based on their relative energies, relative moments of inertia in principle axes and a similarity cut-off, which is the similarity probability between two minima. These quantities need to be fine-tuned to each unique molecule.

5.5.3.1.2 Basin Hopping: Results

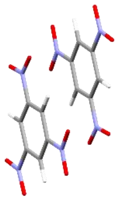
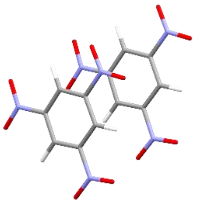
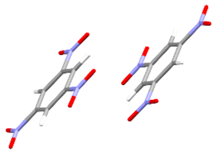
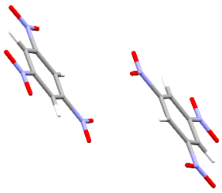
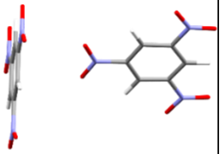
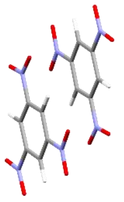
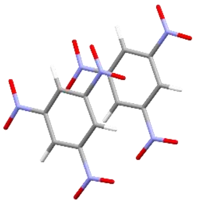
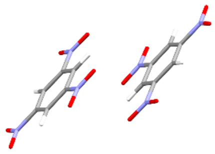
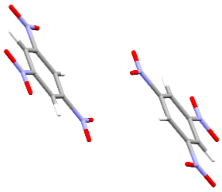
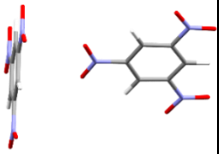
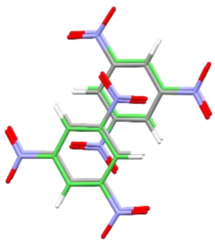
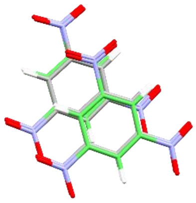
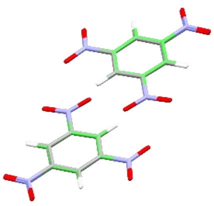
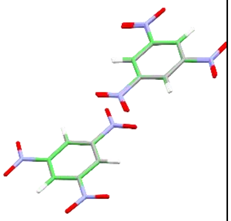
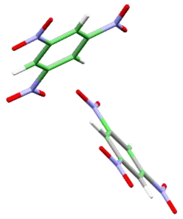
<table>
<tr><th colspan="5">Model0-iso</th></tr>
<tr><th>S1</th><th>S2</th><th>S3</th><th>S4</th><th>T1</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>-26.258 kJ/mol</td><td>-24.128 kJ/mol</td><td>-21.015 kJ/mol</td><td>-20.434 kJ/mol</td><td>-18.566 kJ/mol</td></tr>
<tr><th colspan="5">Empirical FIT</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>-26.275 kJ/mol</td><td>-22.911 kJ/mol</td><td>-22.911 kJ/mol</td><td>-22.712 kJ/mol</td><td>-20.791 kJ/mol</td></tr>
<tr><th colspan="5">Overlays</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>RMSD2 – 0.21</td><td>RMSD2 – 0.19</td><td>RMSD2 – 0.07</td><td>RMSD2 – 0.07</td><td>RMSD2 – 0.04</td></tr>
</table>

Figure 36: *The first row displays the 5 basin-hopping minima with their corresponding label (S for stacked, T for T-shaped) found with the isotropic version of the non-empirical Model0 and their dimer interaction energies computed with the model in* $kJ\ mol^{-1}$*. The following row displays the energies and dimers computed with the empirical FIT model. The final row are structural overlays of the 5 configurations found by Model0-iso (grey) and the empirical FIT model (green). Their* $RMSD_2$ *values are given in Å.*

We find almost identical minima when using the Model0-iso and empirical FIT force-fields (**Figure 36**). In the overlays of the 5 dimer minima obtained using the two models, all root mean square deviations ($RMSD_n$) are below 0.25Å. The energy models favor having the two molecules parallel to one another as opposed to the T-shaped configurations and other configurations found in the crystal (**5.5.3.2**). The lowest energy dimer in both FIT and model0 potential energy surfaces, is the S1 structure, which is strongly stabilized by interactions between NO_2 and the aromatic π-cloud. Things take an interesting turn when we compare the 2nd and 3rd most stable structures in FIT and Model0. Model0 highly favours the staggered mirror stacking of S2 over S3 by over $3\ kJ\ mol^{-1}$, while FIT slightly favours the double NO_2---NO_2 interactions found in S3, the differences between S2 and S3 being only $0.0005\ kJ\ mol^{-1}$. One could suggest that this is due to the non-empirical model's improved description of the dispersion, through the iterated stockholder atoms (ISA) approach. S4 is stabilized by a balance of NO_2---NO_2 and NO_2---aromatic π interactions. The least stable minimum found with both models is the T1 structure, where the oxygen atoms in the NO_2 group interact with the aromatic π-cloud of the neighboring molecule. However, T-shape configurations, and other configurations, are seen in the polymorphs of TNB due to crystal packing forces and the presence of more molecules in the local interaction cluster. This highlights the importance of using the most popular two molecule contacts found in the crystal structure as a test set, especially if we want these non-empirical models to be transferable between phases. Therefore, the assumption that the gas-phase dimer structures appear in the observed crystal structures does not hold for TNB.

5.5.3.1.3 Configuration scan

Obtaining the TNB basin-hopping minima is just an initial analysis of the favoured orientations in the potential energy surface and not necessarily the most informative or complete method of sampling the important dimer configurations needed for the $E_{int}^{(2)}[SAPT(DFT)]$ calculations. More information on the PES can be gleamed by undertaking radial scans of the configuration space on either side of each dimer to obtain additional geometries around the repulsive wall and regions where the long-range interactions are damped. We must restrict the configuration scan along the radial direction as the configuration space for two rigid molecules has 6 degrees of freedom. Hence, only the radial separation vector (R_x, R_y, R_z) varies with scaling the while angle and rotation axis remains fixed.

For each basin hopping dimer, 33 configurations were generated by keeping the relative orientation of each dimer fixed, and fraction scaling the centre-of-mass separation (R) between the pair of molecules. The range of R-scalings was between 0.76 and 1.4. Separations below 1 correspond to the repulsive region of the potential while greater values capture long-range dispersive interactions, therefore, suitable values must be chosen. A dense sampling of the repulsive region is required for a reliable comparison of our short-range potential with second-order SAPT(DFT) interaction energies. ORIENT[108] was used to do energy calculations on each configuration using the Model0-aniso for an initial comparison with $E_{int}^{(2)}[SAPT(DFT)]$ and to determine the importance of 2nd order terms. It should be noted that the distributed polarizabilities and the dispersion model in this Model0-aniso were damped using approximate parameters (**5.3.4**) as the final damping parameters were obtained from these calculations.

5.5.3.2 Crystal Dimers

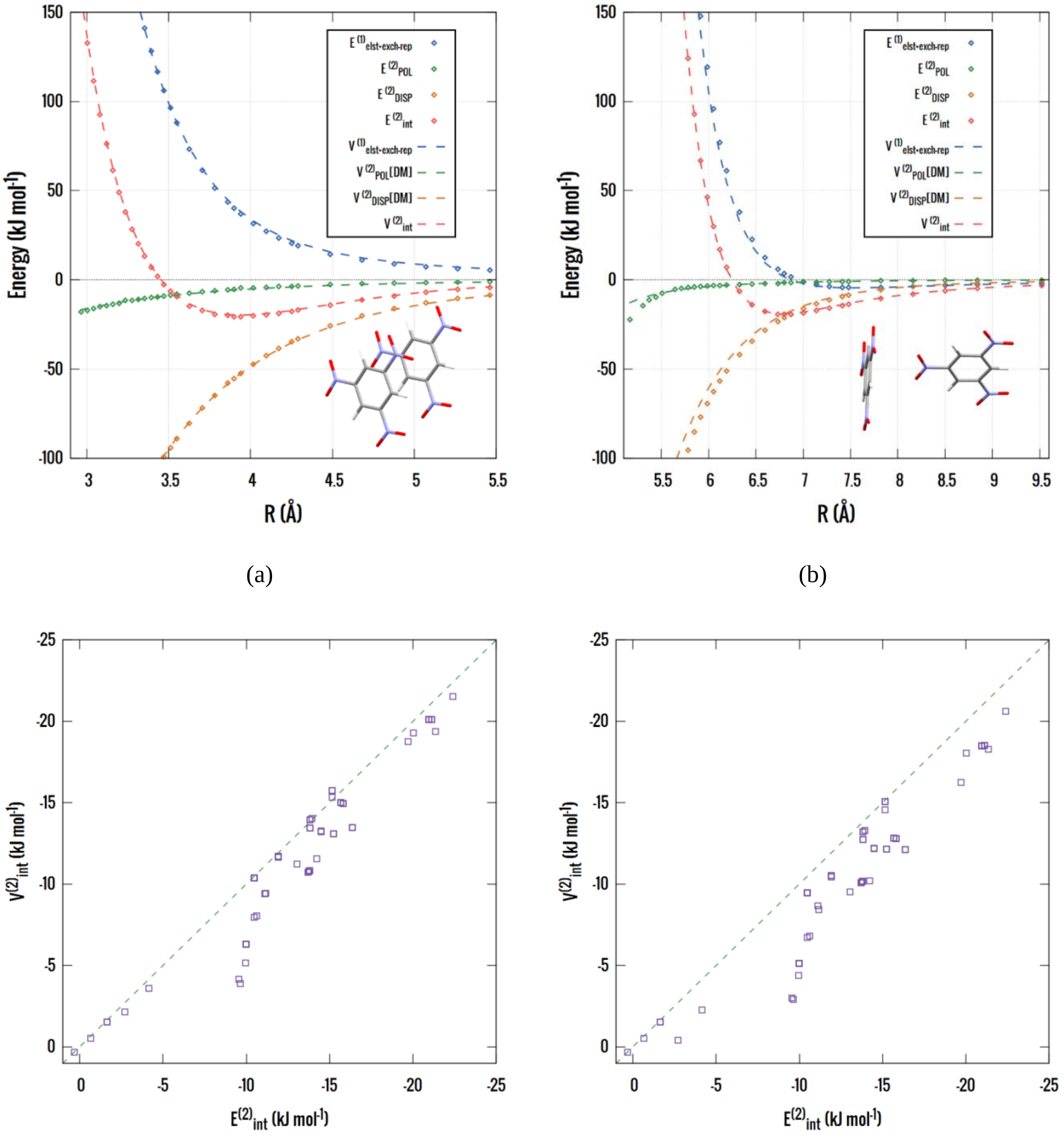
An initial Crystal Structure Prediction (CSP) Study was carried out using CrystalPredictor 1.8[35, 36] to generate a million putative $Z' = 1$ crystal structures of TNB within the 59 most probable space groups (this was also the main generation step for the CSP study carried out in **Chapter 6**). The generated structures were further optimized using the FIT potential and iterated stockholder atoms (ISA) computed point charges and clustered based on their structural similarities. This generated approximately 4 thousand unique structures. These structures were re-optimized using the FIT model again but with the ISA multipole model that included higher order terms up to hexadecapole (Rank $l = 4$) (**5.3.1**). Any optimized structures whose second derivative properties showed that they were not true minima were discarded, and the remaining structures were re-clustered. The 100 most stable structures were analysed in detail for their similarities to each other, the gas-phase dimers, and the 66 most common two molecule contacts were chosen. This was done using the similarity tool in Mercury[291] which determines how many molecules (n) of a maximum coordination cluster (15 for crystals, 2 for crystal/dimer comparisons) can be matched within a 20% distance in intermolecular atom-atom distances and 20° in interatomic intermolecular angles, and reports the optimum RMSD_n (root mean square deviation of n molecules) of the overlay, ignoring hydrogen or deuterium atoms.

Analysis of the two-molecule configurations seen in the FIT+ISA's lowest energy structures suggest that the most common contacts in the thermodynamically plausible crystal structure differ significantly from those found in the gas-phase. These additional data points can act as a test/check for our non-empirical models and ensures that we do not bias our fitting towards an inadequate range of the TNB PES.

5.5.3.3 Avoiding over-fitting

5.5.3.3.1 Pseudo-random dimers

How does one avoid biasing a potential to particular configurations? Force-field development is an iterative process and initially in this TNB study, the pyridine methodology (**Chapter 3**) was used to create a relaxed model. Model0-aniso was relaxed solely to configurations derived from the 5 most stable gas-phase dimers found on the TNB PES. The process of relaxation and the choice of damping parameters is elaborated in further detail in **5.6**, however, in order to explain the need for pseudo-random configurations in this section we will discuss and showcase the relaxed "ModelD".



(a) (b)

(c) (d)

Figure 37: *$E^{(2)}_{int}[SAPT(DFT)]$ energy breakdown and comparison for the S1 (a) and T1 (b) gas-phase dimers, using the relaxed ModelD, which is Model0 relaxed to $E^{(2)}_{int}[SAPT(DFT)]$ calculations on the configuration scan dimers using a dispersion damping parameter of $\beta_{disp} = 1.65$. $E^{(2)}_{int}[SAPT(DFT)]$ and $V^{(2)}_{int}[model]$ were also compared for the most common crystal contacts using ModelD and both damping dispersion parameters, $\beta_{disp} = 1.6$ (c) and $\beta_{disp} = 1.65$ (d).*

Using CamCASP, Model0 was relaxed to the $E^{(2)}_{int}[SAPT(DFT)]$ energies of the 160 structures found in the configuration scans of the 5 gas-phase minima to create a relaxed model dubbed ModelD. Two dispersion damping parameters were used, $\beta_{disp} = 1.6$ and $\beta_{disp} = 1.65$, as both these values approximated the SAPT(DFT) dispersion energy the best (***Figure 39*** & ***Figure 41***). No other minima were used and the configuration space was only limited to the most stable gas-phase dimers. This was due to the great cost of these higher-order calculations and in hopes of being resourceful. For pyridine an astounding 1000+ dimer configurations were used in the fitting to $E_{int}^{(2)}[SAPT(DFT)]$, however, for TNB one must be smarter and try and sample the areas of the PES that are most important. Hence, minima from the configuration scans were used with the belief that the most significant contacts in the gas-phase would be the most dominant in the crystalline state.

A comparison of these relaxed energies $V_{int}^{(2)}[modelD]$ and the $E_{int}^{(2)}[SAPT(DFT)]$ contributions for the S1 and T1 gas-phase dimers (***Figure 37***) suggests an exceptional fit but also highlights one major issue. While the relaxed potential now better described the gas-phase, the model had clearly been over-fit. There was a marked deterioration in its ability to predict crystal contacts, relative to the unrelaxed Model0-aniso (contrasting ***Figure 37*** (c) & (d) with ***Figure 41***). This not only emphasised the importance of retaining the crystal configurations to be used as an independent test set for our fits but also the need for uncorrelated data. Consequently, in the later stages of the force-field development process an additional 282 pseudo-random dimer structures were added: to supplement the $E_{int}^{(2)}[SAPT(DFT)]$ dataset, to ensure a non-bias fit and also increase the space sampled.

5.5.3.3.2 Methodology

The most affected crystal contacts were those with energies between -10 and $-25\ kJ\ mol^{-1}$ (***Figure 37***); contacts close to the PES minimum. In order to address this issue and better model this region of the energy surface, 190 random dimers between -10 and $-22\ kJ\ mol^{-1}$ (the energy of the most stable pseudo-random dimer) were included to the dataset. It is equally important to maintain a thorough description of the repulsive region thus an additional 92 repulsive structures with energies were included whose energies ranged from 3 to $65\ kJ\ mol^{-1}$. Firstly, 1000 additional dimer configurations were generated using the sampling method described in **5.4**. The R_{min} and R_{max} were kept the same as our initial scan (**5.4.2**). However, the scan used the 2000th configuration from the initial data set as a starting point, thus new areas of the configuration space (previously unexplored by the configurations used for the $E_{SR}^{(1)}[SAPT(DFT)]$ calculations) were sampled. Energy calculations were carried out on the new structures using ORIENT and the Model0-aniso created from the $E_{SR}^{(1)}[SAPT(DFT)]$ calculations. The structures were then ranked according to their dimer interaction energies. The first 200 (most stable configurations) and the last 130 (most repulsive) configurations were then chosen for the higher order SAPT(DFT) calculations. However, $E_{int}^{(2)}[SAPT(DFT)]$ calculations using Psi4/CamCASP for 48 of those structures failed and they were not included in the dataset, hence the pseudo-random dataset comprised of a total of 282 structures. The geometries can be found in the **Appendix CD**.

5.5.4 Determining damping parameters

The multipolar dispersion and polarization damping parameters were determined using the configuration scan dataset and checked against the crystal structure dimers. This was done so that the effect of each coefficient could be investigated as a function of R. We found there was no perfect value that covered all ranges of R.

5.5.4.1 *Damping the multipolar polarization energy*

As we already have a theoretical approximation of the dispersion damping parameter ($\beta_{disp} = 1.8$) (**5.3.4**) but not the multipolar polarization damping parameter. β_{pol} must be approximated using the $E_{int}^{(2)}[SAPT(DFT)]$ data. The initial choice is not only approximate but also tentative. In order to not bias the choice of damping parameter to the dimer configurations sampled, the process can be iterative, with the damping model re-assessed and adjusted if need be once the full potential energy surface is obtained. The damping parameter can very easily be changed and for most cases should be reassessed after preliminary CSP studies have been completed using final DIFF and Model0 force-fields. The first approximation of β_{pol} for TNB was used throughout the study though it was one of the weakest points in the potential. β_{pol} was the first to be fit to the $E_{POL}^{(2)}[SAPT(DFT)]$ of the most important dimer orientations. For various values of β_{pol}, the 2nd-order multipolar polarization energy $V_{POL}^{(2)}[DM]$ was estimated by not iterating our model to convergence. The value of β_{pol} was adjusted till *non-iterated* energies best match $E_{POL}^{(2)}[SAPT(DFT)]$. As the polarization energy is a many-body term; it is dependent on terms involving the interaction of multiple sites. Thus, these damping parameters should depend on the types of interacting pairs and possibly their orientation. For example, a N and a O interacting in dimer T1 should, in principle, require a different damping parameter to an N and O interacting in dimer S1 or a C and H in dimer T1. Therefore, by employing a single parameter damping function like our Tang-Toennies, we struggle to obtain a single β_{pol} that correctly describes the 2nd order polarization energy potential surface for all TNB dimer orientations in ***Figure 38***. The more anisotropic configurations, S3, S4 and T1 require stronger damping coefficients (***Figure 38***), which was a trend also seen in the dimers of pyridine.[38] We see that the damping parameters vary dramatically between conformations and across centre of mass distances (R), thus it is difficult to choose a single damping parameter to cover all distances. The most important regions to be damped are around equilibrium distances, highlighted in orange in ***Figure 38***. We can see that this still varies between configuration scans. Moreover, a large spread is seen in the most common two molecule contacts found in TNB crystal structures; further stressing the pitfalls of using a single parameter damping function. By weighting the gas-phase dimers based on their presence in crystal structures and inspecting the effect of β_{pol} on $V_{POL}^{(2)}[DM]$ for the two-molecule contacts present in the crystal, a final polarization damping parameter of 1.0 was chosen.

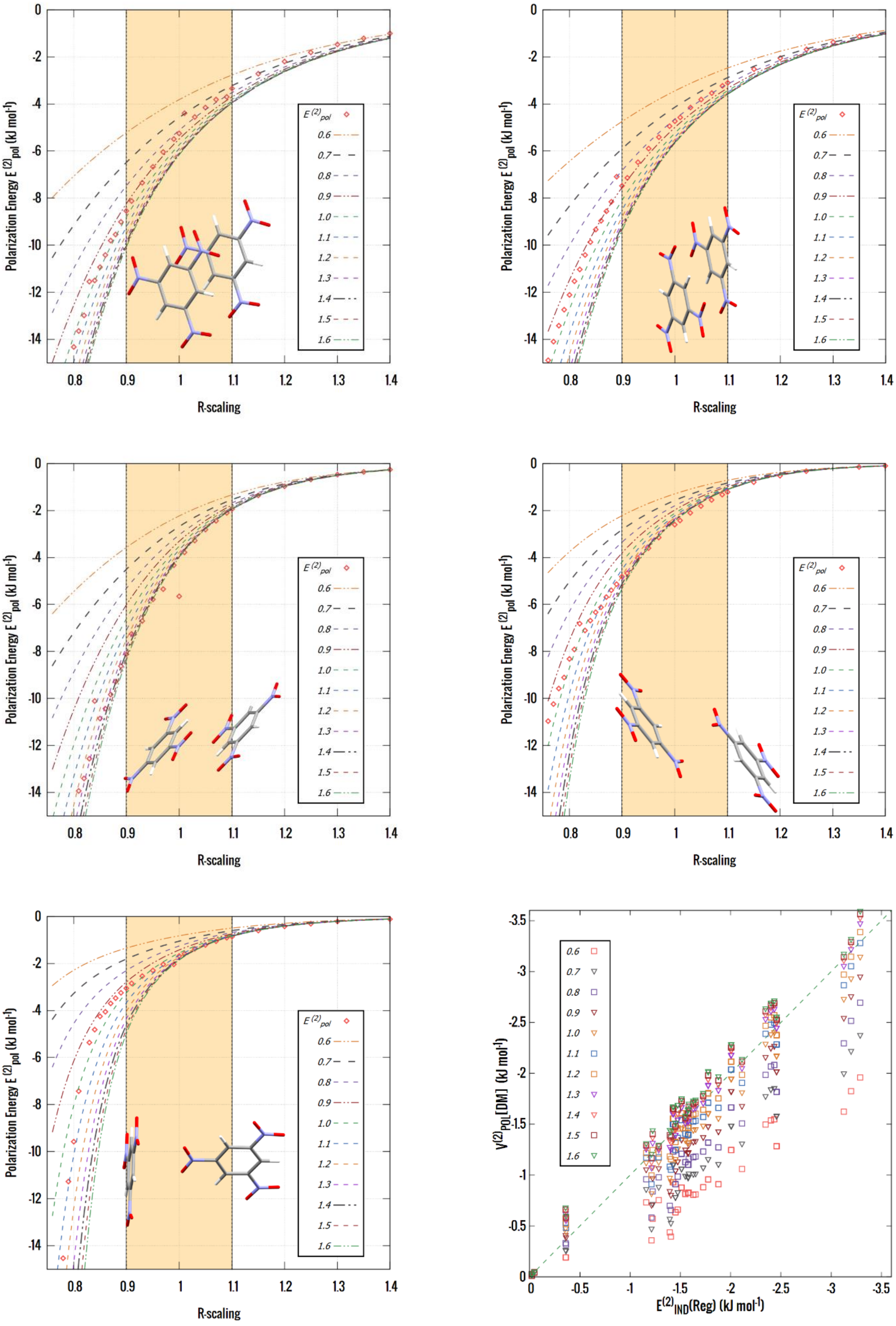


Figure 38: *A comparison of* $E^{(2)}_{POL}[SAPT(DFT)]$ *against the Model0-aniso* $V^{(2)}_{POL}[DM]$ *for the 5 gas-phase dimers and crystal structures (bottom right) using various polarization damping coefficients* (β_{pol}) *and the initial dispersion damping coefficient guess* $\beta_{disp} = 1.8$.

5.5.4.2 Damping the multipolar dispersion energy

The dispersion model derived from the distributed multipole expansion must also be damped to avoid the unphysical behaviour as $R \rightarrow 0$. Moreover, $E_{DISP}^{(2)}[SAPT(DFT)]$ includes exchange effects in short-range that are absent from our long-range multipole expansion. Using a damping function helps combat these divergence issues and absorb some of the short-range effects while the remaining contributions are absorbed in the fitting of our short-range terms. In ***Figure 39*** we compare a various damping parameters and our initial guess ($\beta_{disp} = 1.8$) for both gas-phase and crystal dimer configurations.

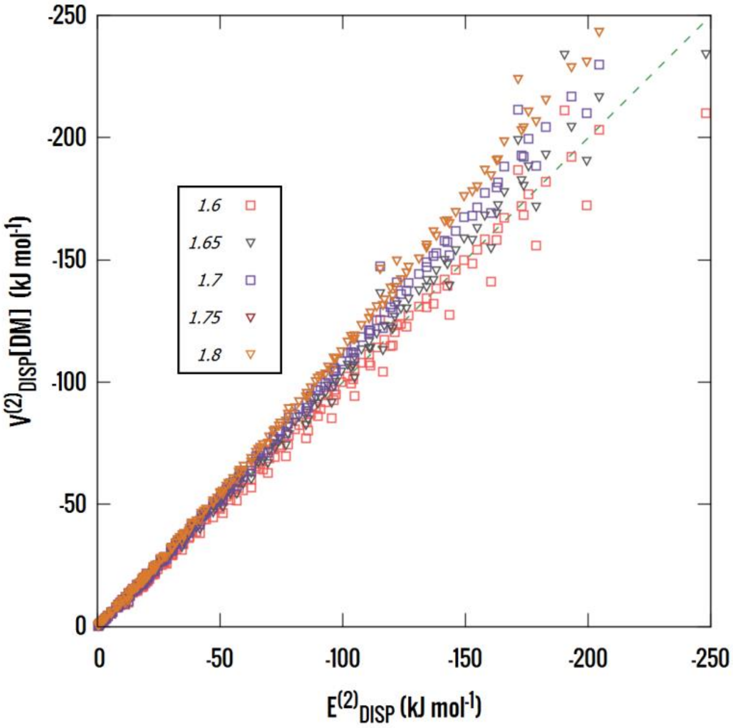
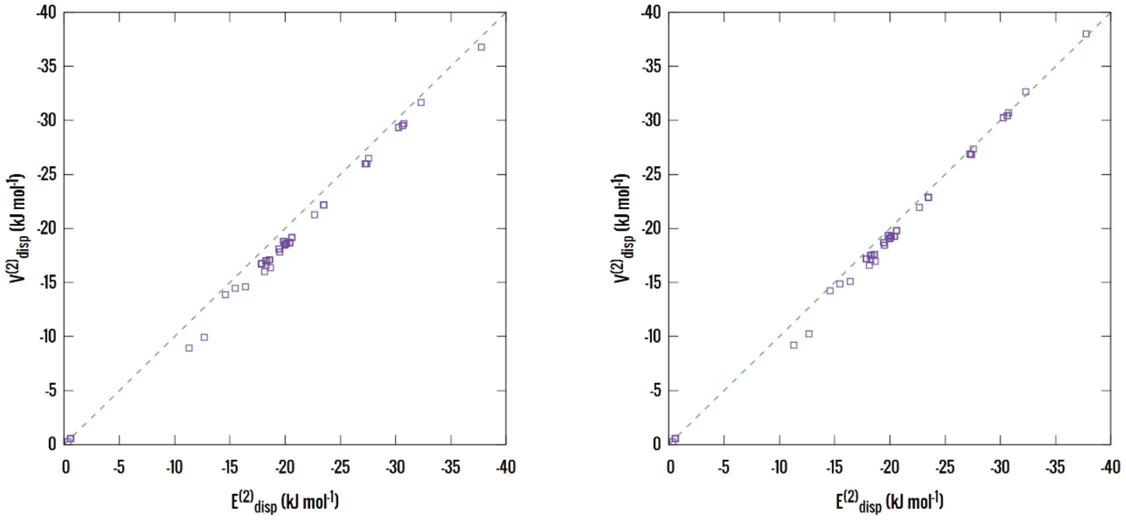


Figure 39: *A comparison of $E_{DISP}^{(2)}[SAPT(DFT)]$ against the Model0-aniso $V_{DISP}^{(2)}[DM]$ across the 200+ dimer configurations sampled in the 5 gas-phase minima and the most stable TNB crystal structures using various damping coefficients β_{DISP}. The initial guess of 1.8 is included for comparison*

Figure 39 suggests that our initial guess of $\beta_{disp} = 1.8$ was sensible. By simply using the ionization energies of monomers M and N in a dimer, an approximate dispersion parameter can be obtained that is suitable for the higher dispersion energies. However, the best correlation was found using $\beta_{disp} = 1.65$. One could argue that $\beta_{disp} = 1.6$ exhibits a better correlation as a substantial portion of points fall on the correlation line and only a few points are slightly under-binding. In addition, the under-binding nature of 1.6 may help combat the overall over-binding nature of our non-empirical potentials discussed in detail in our previous study of pyridine[167] (**Chapter 3**). A quantitative analysis of the datasets (***Table 24***) reveal that unweighted RMSD between the SAPT(DFT) and potential energies was the lowest when $\beta_{disp} = 1.65$ $(RMSD = 4.07\ kJ\ mol^{-1})$ when compared across all energies. Inspecting the dispersion energies around the most important part of the PES, the minima, $(0 > -50\ kJ\ mol^{-1})$ we find the RMSD of all the parameters become very comparable. However, $\beta_{disp} = 1.65$ still has the lowest RMSD.

		Dispersion Damping Parameter (β_{disp})				
		1.6	1.65	1.7	1.75	1.8
RMSD (kJ/mol)	$0 > -250$	4.65	4.07	7.63	12.22	10.44
	$0 > -50$	1.22	0.91	1.06	1.49	1.30

Table 24: *A comparison of the root mean squared deviations (RMSD) in kJ mol*$^{-1}$ *between the calculated* $E^{(2)}_{DISP}[SAPT(DFT)]$ *and the Model0-aniso* $V^{(2)}_{DISP}[DM]$ *across the 200+ dimer configurations sampled in the 5 gas-phase minima and the most stable TNB crystal structures using various damping coefficients* β_{disp}.



(a) (b)

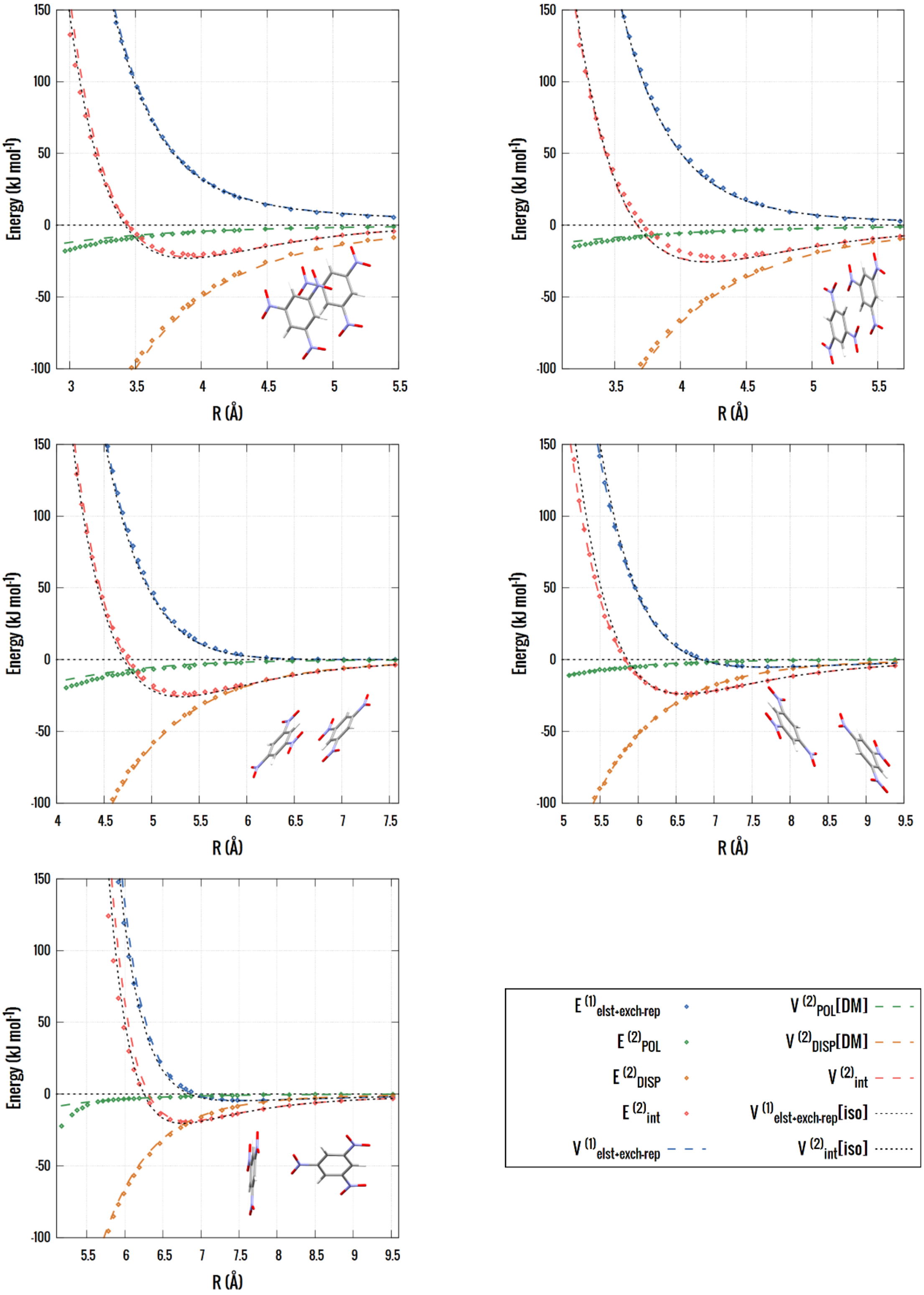
Figure 40: *A comparison of the SAPT(DFT) calculated* $E^{(2)}_{DISP}$ *and the Model0-aniso* $V^{(2)}_{DISP}[DM]$ *for the most common two molecule contacts in TNB crystal structures* $\beta_{disp} = 1.6$ *(a) and* $\beta_{disp} = 1.65$ *(b).*

5.5.5 Testing Model0

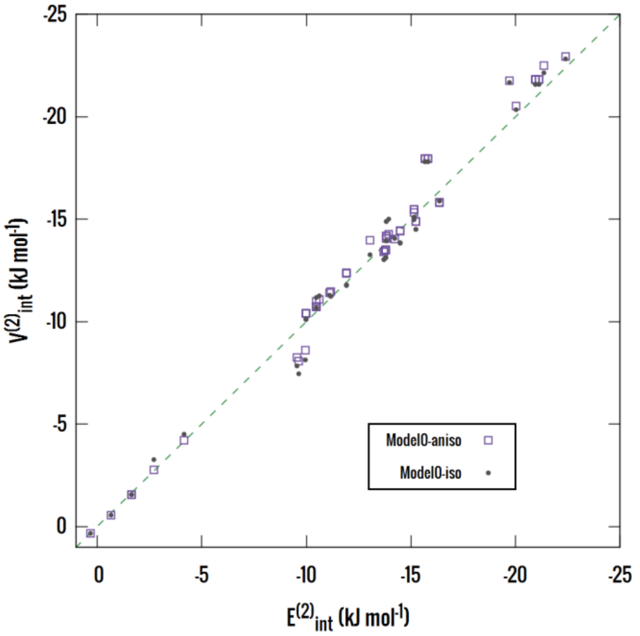
As both $\beta_{disp} = 1.65$ and 1.6 are very comparable (***Figure 40***), a final test and comparison of the two damping dispersion parameters was carried out using the new polarization damping parameter $\beta_{pol} = 1.00$. $E^{(2)}_{int}[SAPT(DFT)]$ and $V^{(2)}_{int}[Model]$ for the Model0-aniso was compared for the different dimer configurations to investigate how well our approximations (Tang-Toennies damping, using a single isotropic damping parameter, ignoring charge-delocalization effects) do. While $E_{elst}^{(1)}[SAPT(DFT)]$ contains both the penetration and multipolar electrostatic energies (**2.9.1**), the $E_{pen}^{(1)}$ has been absorbed into our non-empirical potentials exchange-repulsion term. Accordingly, we must compare the summation of the exchange-repulsion and electrostatic energies (***Figure 41***)

$$V_{exch-rep}^{(1)}[model] \approx E_{exch-rep}^{(1)}[SAPT(DFT)] + E_{pen}^{(1)} \tag{105}$$

Both dispersion parameters perform amazingly for the gas-phase dimers, however, $\beta_{disp} = 1.65$ performs the best in predicting the crystal dimer interaction energies (***Table 24***, ***Figure 40***, ***Figure 41*** & **Appendix 5.B**). A key thing to note is how well both models perform in long-range, even before the potentials have been relaxed. We see that for all gas-phase configurations the long-range portion of potential energy surface is near enough perfect. This is an amazing achievement as it stresses the real accuracy, reliability and power of derivation and the ISA partitioning methods used to obtain the distributed molecular properties.


Energy (kJ mol-1)
150
100
50
0
-50
-100
R (Å)
3
3.5
4
4.5
5
5.5
4
4.5
5
5.5
6
6.5
7
7.5
5
5.5
6
6.5
7
7.5
8
8.5
9
9.5
E (1)elst+exch-rep
E (2)POL
E (2)DISP
E (2)int
V (1)elst+exch-rep
V (2)POL[DM]
V (2)DISP[DM]
V (2)int
V (1)elst+exch-rep[iso]
V (2)int[iso]

(a)



(b)

Figure 41: *The first 5 graphs (a) compare the $E_{int}^{(2)}[SAPT(DFT)]$ energy breakdown (points) with the Model0-aniso (coloured dashes), with $\beta_{disp} = 1.65$, for the 5 gas-phase dimers, while the last graph (b) compares $E_{int}^{(2)}[SAPT(DFT)]$ & $V_{int}^{(2)}[Model]$ for crystal structures using SAPT(DFT) (points) with the dashed line signifying equality. Model0-iso (black dashes for the gas-phase dimers and dots for the crystal dimers) is included for comparison.*

5.6 Further adjustments to the potential

After finalising the damping parameters of the 1st order models (Model0-iso and Model0-aniso) and selecting which higher-order terms to include, the next step is to relax the short-range portion of the potential to the $E_{int}^{(2)}[SAPT(DFT)]$ energies of the selection of dimers discussed in **5.5.3**. In **5.5.5**, we found both dispersion parameters $\beta_{disp} = 1.6$ & 1.65 perform exceptionally well. However, based on the unweighted RMSD between $E_{int}^{(2)}[SAPT(DFT)]$ and $V_{int}^{(2)}[model0]$, we settled with $\beta_{disp} = 1.65$. ***Figure 38*** reminds us that it is not possible perfectly describe the polarization contribution using a single damping parameter as it leads to systematic errors in some orientations. Yet it is the simplicity of our damping models that allow us to hypothetically get away with approximating the damping parameters and not having to calculate 2nd order energies. Nonetheless, using $\beta_{pol} = 1.00$ is a good enough approximation if these and other errors can be absorbed by adjusting the short-range parameters. All considered, Model0 does exceptionally well in modelling the gas-phase dimers and even the local crystal contacts (***Figure 41***). What is truly exciting about this result is that while calculating the higher order energies may be very computationally exhaustive for more complex molecules, a 1st order model is very accessible, furthermore, using informed approximations like Equation 101 can result in solid guesses of the damping parameters without the need for higher order energy calculations. Subsequently, we begin the relaxation process from a very solid starting point in Model0-aniso.

5.6.1 Relaxing to second-order SAPT(DFT) energies

The relaxation process was completed by employing a combination of modules within CAMCASP and ORIENT to relax the $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$ short-range parameters for the Model0-aniso to the $E_{int}^{(2)}[SAPT(DFT)]$ energies of the 160 structures found in the configuration scans of the 5 gas-phase minima and the 282 pseudo-random structures generated. The $E_{int}^{(2)}[SAPT(DFT)]$ and $V_{POL}^{(2)}[DM]$ energies were included in an ORIENT input file that contained the co-ordinates of each configuration using an angular axis system. The $V_{elst}^{(1)}[DM]$ and $V_{DISP}^{(2)}[DM]$ energies were not included in the input file as ORIENT can calculate these terms from their respective multipole moments and dispersion coefficients file. Currently, ORIENT cannot calculate the non-iterated polarization energy $V_{POL}^{(2)}[DM]$ in the initial re-parameterization step of the relaxation process. As a result, these energies must be inputted separately and subtracted from the potential and the energies fitted can be given as

$$e_{fit} = E_{int}^{(2)}[SAPT(DFT)] - V_{elst}^{(1)}[DM] - V_{DISP}^{(2)}[DM] - V_{POL}^{(2)}[DM]$$

106

i.e. the short-range terms are relaxed to accommodate for $E_{exch}^{(1)}$, $E_{pen}^{(1)}$, $E_{CD}^{(2)}$ and the residual errors made by the damped distributed multipole expansions of the polarization and dispersion contributions. The relaxation process within ORIENT uses the below Boltzmann weighting scheme (Equation 107) as the Gaussian-log weighting scheme (**2.9.5.2**) used in CAMCASP is not available to ORIENT and the total energies go through zero, so the Gaussian-log weighting scheme cannot be used.

$$w(e) = \exp\left[\frac{\left(e_m - V_{int}^{(2)}[Model]\right)}{e_0}\right]$$

107

The energy defining the Boltzmann weighting $e_0 = 100$ kJ mol^{-1} is based off the empirical rule that the value of e_0 should be about $4 \times$ the modulus of the global minimum. $e_m = -26$ kJ mol^{-1} as the SAPT(DFT) calculated global minimum for the gas-phase structures is $\sim > -26$ kJ mol^{-1}. Before relaxation the datasets were scanned by CAMCASP and any dimers above $E_{max} = 2000\ kJ\ mol^{-1}$ and below $E_{min} = -30\ kJ\ mol^{-1}$ were discarded as energies so high up the repulsive wall are likely to be erroneous, and any energy considerably below the global minimum is likely to include an error from the multipole expansion. This is a precautionary measure, because for TNB there were no dimer configurations outside this bound under any of the models.

The $\rho_{00}^{\iota\kappa}$ term in Equation 102 was relaxed first; while this term may not improve the fit as well as relaxing the anisotropic terms $\rho^{\iota\kappa}(\Omega_{ik})$ or the $\alpha_{00}^{\iota\kappa}$ term, it is the most appropriate term to relax. Of the short-range terms, the isotropic tern $\rho_{00}^{\iota\kappa}$ can absorb the most changes while maintaining some sort of physicality. The anisotropic terms should remain untouched as they are closely related to the multipole moments in their directionality and magnitude and a massive dataset would be required to justify any deviation from the overlap model. Additionally, as $\alpha_{00}^{\iota\kappa}$ and $\rho^{\iota\kappa}(\Omega_{ik})$ are dependent on one another, changing $\alpha_{00}^{\iota\kappa}$ too much could result in very unphysical parameters. It could be argued that only terms on important atoms like N and O should be relaxed but the relaxation of $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$ for all atomic sites is allowed with strong anchors to prevent any parameters from slipping too far from their initial value while still greatly improving the fit.

For each dataset of dimer configurations the Model0-aniso was employed as a starting point and $\rho_{00}^{\iota\kappa}$ is re-fitted first to create a transitory model, which is then used as a starting point for the relaxation of the $\alpha_{00}^{\iota\kappa}$ terms. Not only does this prevent the values from slipping but also greatly improves on the fit. The anchors for both $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$ were given values of 1.0, which are fairly conservative to prevent over-fitting. For each model the relaxation was done in stages, as used for obtaining the anisotropic terms in the overlap model (**5.4.3**), starting with all pairs involving oxygen, then all pairs containing N (including the $N - O$ pair, which is re-adjusted), then C_N, afterwards C_H and finally H. This is done using the Python fitting codes included in the CAMCASP package.

5.6.2 Results & Discussion

5.6.2.1 Selecting a final non-empirical force-field

Multiple relaxed models were created and compared to see the effect of sampling various areas of the configuration space and observe how changing $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$ affects each models ability to predict the structures in both gaseous and crystal phases. The two step fits were performed on 4 different datasets (***Table 25***) generating 8 tentative 2nd order models, a model relaxed to all 160 gas-phase dimers found in the configuration scan was initially made, dubbed ModelD. This is the model, which initially informed our decision to include pseudo-random dimers in our dataset (**5.5.3.3**). 3 additional models were created using the pseudo-random data set. A model using only the 282 pseudo-random dimers (ModelR) and a model using all 442 configuration scan and pseudo-random dimers (ModelA) was created. A more informed model (ModelS) was also generated using the all 282 pseudo-random configurations and a select set of gas-phase dimers. In order to appropriately sample the gas-phase PES but not over-fit, 9 configurations from each of the 5 gas-phase minima were used at R-scalings of the dimers corresponding to $0.76,\ 0.81,\ 0.85,\ 0.90,\ 1.0,\ 1.1,\ 1.15,\ 1.35,\ 1.4$.

(Tentative) Model	№	$E_{int}^{(2)}[SAPT(DFT)]$ Energy Range (kJ mol^{-1})	Initial weighted RMS using model0	Weighted RMS after relaxing		Non-weighted RMS against crystal SAPT(DFT)	
				ρ_{00}	α_{00}	ρ_{00}	α_{00}
FIT+ISA	~	~	~	~	~	1.537	
Model0	~	~	~	~	~	0.794	
ModelD	160	1000 → −25	7.78	2.97	1.24	3.419	3.540
ModelR	282	66 → −20	1.26	1.07	0.89	0.699	0.573
ModelS	327	1000 → −25	2.91	1.88	1.39	0.705	0.469
ModelA	442	1000 → −25	4.57	2.17	1.51	0.678	0.464

(a)

(Tentative) Model	№	$E_{int}^{(2)}[SAPT(DFT)]$ Energy Range (kJ mol^{-1})	Non-weighted rms against gas-phase dimers (SAPT(DFT) (all energies)		Non-weighted rms against gas-phase dimers (SAPT(DFT) (R-scaling ≥ 0.85)	
			ρ_{00}	α_{00}	ρ_{00}	α_{00}
FIT+ISA	~	~	58.2		6.6	
Model0	~	~	81.0		9.7	
ModelD	160	1000 → −25	17.1	10.3	1.9	1.1
ModelR	282	66 → −20	78.5	57.7	9.4	7.3
ModelS	327	1000 → −25	46.1	25.7	5.1	3.1
ModelA	442	1000 → −25	29.6	20.0	3.3	2.0

(b)

Table 25: *A comparison of the root mean squared deviations (RMSD) in kJ mol^{-1} between the SAPT(DFT) calculated energies and the computed energies using various models. Table (a) compares the RMSD when used for the 66 crystal dimers, before relaxation and after relaxing $\rho_{00}^{\iota\kappa}$ and then $\alpha_{00}^{\iota\kappa}$. Table (b) investigates the changes in RMSD for the 160 gas-phase dimer energies across all R-scalings and from close to equilibrium → long-range*

From ***Table 25*** one can see how dramatically each fit improves once $\alpha_{00}^{\iota\kappa}$ is allowed to relax. This is because changing this term affects the slope of our repulsive term (Equation 102). As discussed in **5.5.3.3**, ModelD is a deceiving case. With a low weighted RMSD after relaxing $\rho_{00}^{\iota\kappa}$ then $\alpha_{00}^{\iota\kappa}$ and the lowest non-weighted rmsd when compared against the gas-phase dimer $E_{int}^{(2)}[SAPT(DFT)]$ energies (***Table 25*** (b)), one would believe ModelD to be the best performing model. However, comparing the RMSD against the crystal dimer $E_{int}^{(2)}[SAPT(DFT)]$ energies in ***Table 25*** (a) (also illustrated in ***Figure 37***) we see the ModelD dataset is performing very poorly when modelling orientations not found in the gas-phase minima. ModelD is highly correlated as it is fitting to only 5 orientations at different distances. Accordingly, the fits improve on the modelling of these orientations but give a significantly worse representation of the dataset that samples contacts from TNB crystal structures.

The other 3 datasets sample a range of orientations while giving fits of comparable quality. It is not appropriate to use ModelR as the final model as the pseudo-random dataset on its own samples too limited a range of important configurations, while ModelA includes too many gas-phase dimer structures thus may have biased our potential. Furthermore, increasing the number of structures to include all points does not significantly improve on the fit as seen by the non-weighted RMSD against the crystal dimers (***Table 25*** (a)). ModelS seems to be the most appropriate dataset as it has a balance of both random and dimer structures while including some high energy points and giving a good fit in the energy ranges sampled by the contacts in the TNB crystal structures. Consequently, we deem ModelS the most complete non-empirical force-field we have created for TNB and from henceforth will refer to ModelS, with both $\rho_{00}^{\iota\kappa}$ then $\alpha_{00}^{\iota\kappa}$ relaxed, as TNB's Distributed Intermolecular Force-Field (DIFF).

5.6.2.2 Comparing the short-range anisotropic terms

The short-range anisotropic terms for the DIFF model remain unchanged and are identical to those found in the Model0-aniso in ***Table 23***. We see some changes in the $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$ terms upon relaxation in ***Table 26***. As the relaxation step allows parameters on all atom pairs to relax, we see the most of the notable changes occurring in pairs involving the carbon atoms. Previously mentioned in **5.4.4.2**, these central atoms were poorly sampled by the configurations in $E_{int}^{(1)}[SAPT(DFT)]$ calculations used for the overlap model stage. Therefore, it is possible that the atoms on the aromatic ring with considerable electron density like the carbon atoms were not well-described by our overlap model, resulting in the changes seen in ***Table 26*** upon relaxation. The hydrogen atoms on the other hand have a smaller local charge density and are less anisotropic. This is emphasized in their higher order multipole moment maps in ***Figure 31*** and their atomic polarizabilities in ***Table 19***. Accordingly, a poor sampling of $H \cdots X$ contacts would not be as detrimental to their overlap model description.

	$\rho_{00}^{\iota\kappa}(a.u.)$			$\alpha_{00}^{\iota\kappa}(a.u.)$		
$\iota \cdots \kappa$	Model0-aniso	DIFF	Δ	Model0	DIFF	Δ
$O \cdots O$	5.669	5.681	0.012	2.000	1.994	−0.007
$O \cdots N$	5.446	5.399	−0.047	2.135	1.969	−0.166
$O \cdots C_N$	5.880	5.664	−0.217	1.950	1.882	−0.069
$O \cdots C_H$	5.954	6.085	0.131	1.830	1.818	−0.011
$O \cdots H$	4.776	4.684	−0.092	1.935	1.836	−0.099
$N \cdots N$	4.905	4.939	0.034	2.534	2.531	−0.003
$N \cdots C_N$	5.602	5.703	0.101	1.983	1.902	−0.081
$N \cdots C_H$	5.406	5.577	0.172	1.721	1.701	−0.020
$N \cdots H$	3.995	4.005	0.010	1.773	1.753	−0.019
$C_N \cdots C_N$	6.252	6.253	0.001	2.008	1.997	−0.011
$C_N \cdots C_H$	6.147	6.059	−0.088	1.710	1.505	−0.205
$C_N \cdots H$	4.764	4.827	0.063	1.844	1.789	−0.055
$C_H \cdots C_H$	6.138	6.217	0.079	1.607	1.650	0.043
$C_H \cdots H$	5.139	5.332	0.193	1.963	1.945	−0.019
$H \cdots H$	3.573	3.576	0.003	1.972	1.969	−0.003

Table 26: *A comparison of $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$ values for all atom-type pairs in both the Model0-aniso and the final relaxed DIFF model in atomic units. Like-like pairs have been highlighted for clarity.*

The differences in the $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$ terms between DIFF and Model0-aniso stresses the importance of having mixed datasets. We see C_N and C_H vary considerably once the configuration scan dataset is included. The carbon atoms are better sampled in the gas-phase dataset (**5.5.3**), as we see these favour stacked configurations (***Figure 36***). This preference is not something that would be represented by pseudo-random sampling, and further highlights the true importance of exploring the configuration space. Choosing where in configuration space to sample is crucial to generating a physically representative model. Additionally, weighting or selecting configurations carefully using an informed chemical intuition is imperative to preventing potential bias.

One observes the following relationship between atom-atom contacts and the isotropic repulsion parameter $\rho_{00}^{\iota\kappa}$:

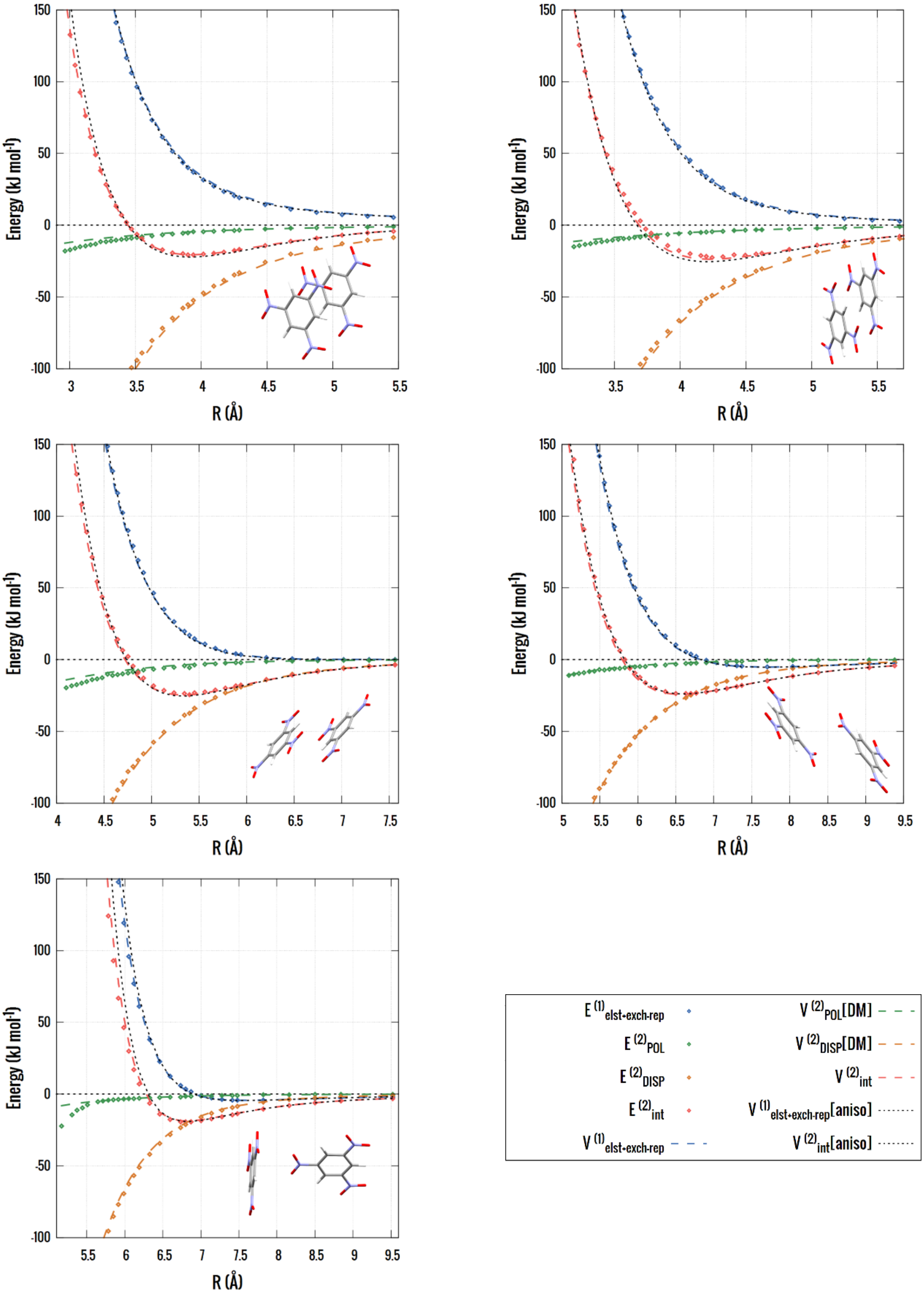
$$H \cdots H \ll X \cdots H < X \cdots X$$

Contacts involving the larger, more electron rich atoms C, N & O have larger $\rho_{00}^{\iota\kappa}$s, which suggests the potential is indeed capturing the physics behind the interactions. This also mirrors the results obtained from the molecular properties derived directly from the molecular charge distribution (***Table 19***, ***Table 20*** & ***Figure 31***). The $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$ terms for TNB are on the same order of magnitude as those found in pyridine (**Appendix 3.A**), however, the $\rho_{00}^{\iota\kappa}$ terms for the $H \cdots H$ and $X \cdots H$ contacts are notably smaller. This is likely due to the presence of the electronegative nitro groups pulling local charge away from the hydrogen atoms and aromatic ring in general, and consequently shielding these atoms. Even with the differences in functional groups between pyridine and TNB, we find the $\alpha_{00}^{\iota\kappa}$ terms to be very comparable. For example $\alpha_{00}^{\iota\kappa}$ is between $1.960 - 1.965$ for $H \cdots H \ (DIFF[pyridine])$ compared to 1.969 for $H \cdots H(DIFF[TNB])$. This comparison acts also as an additional sanity-check to ensure that our final force-field is sensible.

5.6.2.3 Modelling the gas-phase with the DIFF model

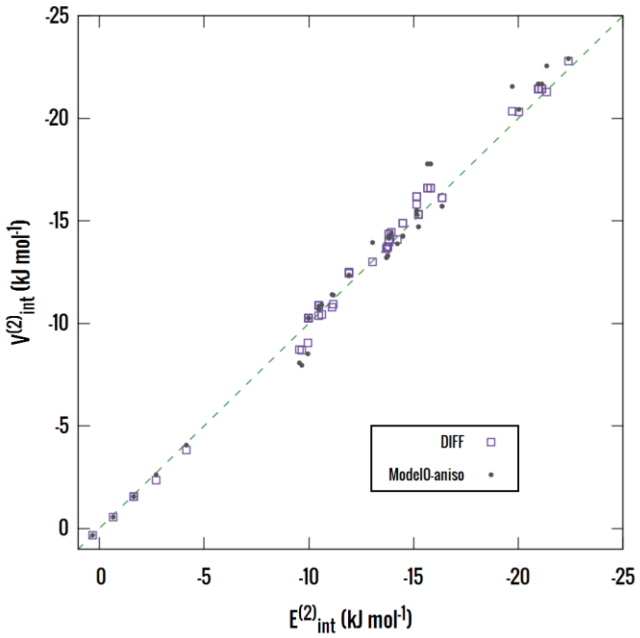
Whilst ***Figure 41*** showed the initial approximate model (Model0) was already good, ***Figure 42*** shows that the DIFF model better predicts the equilibrium energies of the gas-phase dimers and the energies of the most common TNB crystal contacts. In ***Figure 42*** we see that the potential does an amazing job in predicting energies above equilibrium bond-distances, for each confirmation the long-range portion of the PES is almost perfect. This really emphasizes on the accuracy and trueness of our long-range model. Deriving the electrostatic, polarization and dispersion models from a materials molecular properties is not computationally impractical for a large number of energetic and pharmaceutical molecules, and hence should be at least considered in contemporary non-empirical force-field development going forward.

For these non-empirical models, our main concerns lie in accurate modelling of the damped dispersion and polarization at short-range. In ***Figure 42*** can clearly see the chief sources of error in our potential are in the short-range/damped portion of the PES, in particular, the polarization and dispersion contributions. As stated earlier, this is likely due to the over-simplification of our Tang-Toennies damping function and the obvious shortcomings of using a single damping parameter to describe an anisotropic molecule like TNB. Most readily accessible codes do not accept orientation dependent or atom-atom specific damping parameters. However, we see that this should certainly be considered in the development of future codes and force-fields, especially those intended to be used for highly accurate and reliable CSP on anisotropic organic molecules.



Energy (kJ mol-1)
R (Å)
150
100
50
0
-50
-100
3
3.5
4
4.5
5
5.5
4
4.5
5
5.5
6
6.5
7
7.5
5
5.5
6
6.5
7
7.5
8
8.5
9
9.5
$E^{(1)}_{elst+exch-rep}$
$E^{(2)}_{POL}$
$E^{(2)}_{DISP}$
$E^{(2)}_{int}$
$V^{(1)}_{elst+exch-rep}$
$V^{(2)}_{POL}$[DM]
$V^{(2)}_{DISP}$[DM]
$V^{(2)}_{int}$
$V^{(1)}_{elst+exch-rep}$[aniso]
$V^{(2)}_{int}$[aniso]

(a)



(b)

Figure 42: *$E^{(2)}_{int}[SAPT(DFT)]$ energy breakdown and comparison with the DIFF model for the 5 gas-phase dimers using SAPT(DFT) (points) and our non-empirical potential (dashes) (a). The last figure (b) plots $E^{(2)}_{int}[SAPT(DFT)]$ against $V^{(2)}_{int}[DIFF]$ for the crystal dimers with the dashed line indicating when they are equal. Model0-aniso (black dashes and dots) is included for comparison.*

TNB has definitely pushed the current DIFF development methodology, which in turn has resulted in many changes and improvements to the workflow but the question then arises, did we need to calculate the 2nd order energies? In a sense yes, there is a notable improvement in predictive ability (***Figure 41*** vs ***Figure 42*** & ***Table 25***). While the initial approximation of β_{disp} using Equation 101 was close to our final parameter (1.8 vs 1.65), we see a significant reduction in the RMSD when 1.65 is used over 1.8 (***Table 24***). Furthermore, there is currently no physically meaningful way of approximating β_{pol} except for basing a guess on previous work i.e. using $\beta_{pol} = 1.25$, the polarization damping parameter used in the pyridine study (**Chapter 3**). Yet, this treatment is unreliable and unlikely to extend to novel or larger molecules. Additionally, we find our largest percentage errors in $V^{(2)}_{POL}[DM]\ vs\ E^{(2)}_{POL}[SAPT(DFT)]$ and $V^{(2)}_{DISP}[DM]\ vs\ E^{(2)}_{DISP}[SAPT(DFT)]$ (***Figure 38*** & ***Figure 39***), and relaxing potential parameters to 2nd order energies is vital to reducing these errors, particularly for organic molecular crystals that are dominated by polarization and dispersive interactions. One can certainly obtain a good potential using only $E^{(1)}_{int}[SAPT(DFT)]$ and an informed choice of partitioning treatments. In **Chapter 6**, we will test whether or not higher order energy data is truly required in order to cultivate a force-field fit for the modern demands of CSP.

Cautious and careful planning is also a necessity when developing non-empirical models for larger organic molecules as ***Figure 37*** (c) & (d) emphasises the dangers of assuming one can correctly model the condensed phase using only configurations favoured in the gas-phase. A smart sampling of the configuration space is required, especially for larger molecules like TNB, which are currently quite computationally exhaustive. Furthermore, we must remind ourselves that in making this DIFF some approximations have been made in the SAPT(DFT) calculations that has saved a gargantuan amount of computing time, but these approximations may not be applicable to all molecules. We ignore the effects of 3rd to infinite order energies as we found them to be not crucial to describing TNB's intermolecular interactions (**5.5.1**) and even detrimental in some cases (**Chapter 3**). We have also held TNB in its rigid planar conformation, the true implications of which we shall delve into in the following chapter.

5.7 Conclusions & Further work

We find the most important contributions are the 1st order terms, namely, the exchange-repulsion and electrostatic energies. Modelling these concisely is crucial to producing an accurate non-empirical force-field, and consequently, our 1st order models do well (***Figure 41*** & ***Table 25***). Using only the $E_{int}^{(1)}[SAPT(DFT)]$ dimer interaction energies to derive the non-empirical force-field's short-range parameters does exceptionally well compared to the standard empirical FIT model in predicting the crystal and gas-phase dimers, particularly for Model0-aniso (RMSDs in ***Table 25***). 1st order non-empirical models are certainly accessible for many pharmaceutical and energetic molecules, however, we find the improvements in relaxing to 2nd order dimer energies in order to obtain the final non-empirical model (DIFF) have come at a large cost. We have found a notable improvement in the fit requires a large $E_{int}^{(2)}[SAPT(DFT)]$ dataset. Hence, creating a DIFF for molecules larger than TNB may be very challenging.

Future work should look to drastically increase the high-order energy dataset to sample more hindered atomic sites like those in the centre of the aromatic ring, as a better sampling of the carbon and hydrogen atoms may have led to different $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$ values. Moreover, a large enough dataset would enable a physically meaningful and reliable relaxation of the anisotropic terms. If the anisotropic short-range terms could be relaxed, then it is possible only relaxing terms on the oxygen and nitrogen would still lead to an outstanding fit, as they are the largest contributors to the anisotropy of TNB. Codes that allow for atom-atom damping parameters in force-fields do exist (e.g. OpenMM[390]), and informed by the work in this chapter, future non-empirical force-fields created to be used in similar codes should include both single Tang-Toennies damping parameters and atom-atom damping parameter. Not only would this increase the transferability of the potential but also further reduce errors and the need for re-fitting.

To conclude, the resultant TNB DIFF performs exceptionally well in the gas-phase and is a truly a state-of-the-art bespoke force-field that pushes the current capabilities of modern potential development methods. Many costly calculations were required in order to obtain reliable damping parameters and accurate isotropic short-range parameters ($\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$) but a substantially larger dataset would be needed if one wanted to improve on the anisotropic terms. We will see in **Chapter 6** as we look to Crystal Structure Prediction as the final test, whether the 2nd order SAPT(DFT) terms are truly worth it in comparison to Model0-iso or Model0-aniso.

Appendix 5.A – Input Files for the DIFF

Table of Contents

Further details on the formatting of the input files can be found in the ORIENT manual.

5. Local Axes

Local/atomic axis in ORIENT input format using the TNB's atomic labels.

6. Iterated Stockholder Atoms – Distributed Multipole Moments (ISA-DMA)

The ISA-DMA multipolar moments are given ORIENT input file format. These detail the atom labels, atom types and multipole moment ranks used for TNB's electrostatic model. The atomic positions can also be found in this input file, adjacent to the corresponding atom label.

7. Iterated Stockholder Atoms – Distributed Polarizabilities

The ISA-DMA derived $Rank = 1$ distributed polarizabilities in ORIENT input file format. The values for only 5 atom types is included as the distributed polarizabilities is calculated in the local axes in ORIENT.

8. Model0-iso ORIENT input file

Model0-iso is given in ORIENT input file format; detailing the atom types. For each atom-atom type pair, the isotropic short-range parameters $\rho_{00}^{\iota\kappa}$ (as rho) and $\alpha_{00}^{\iota\kappa}$ (as alpha) is given. The dispersion coefficients ($C_6 \rightarrow C_{12}$) are also given adjacent to $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$. Please see the ORIENT manual for further details on its formatting.

9. Model0-aniso ORIENT input file

Model0-aniso is given in ORIENT input file format; detailing the atom types. For each atom-atom type pair, the isotropic short-range parameters $\rho_{00}^{\iota\kappa}$ (as rho) and $\alpha_{00}^{\iota\kappa}$ (as alpha) and the anisotropic short-range parameters $\rho^{\iota\kappa}(\Omega_{ik})$ (in its shorthand notation) is given. The dispersion coefficients ($C_6 \rightarrow C_{12}$) are also given adjacent to $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$.

10. Distributed Intermolecular Force-Field (DIFF) ORIENT input file

DIFF is given in ORIENT input file format. Its formatting is identical to that of Model0-aniso and it has identical dispersion coefficients and anisotropic short-range terms to Model0-aniso as only $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$ were relaxed. Accordingly, these values have been highlighted for each atom-atom type pair to emphasise the changing terms. As the anisotropic terms are identical they are not included below but the fully functioning input file can be found in the **Appendix CD**.

Local Axes ORIENT Input File

```
Axes
  C1  z from C1 to N1  x from C1 to C2
  N1  z from C1 to N1  x from N1 to O2
  O1  z from N1 to O1  x from N1 to O2
  O2  z from N1 to O2  x from N1 to O1
  C2  z from C2 to H1  x from C2 to C3
  H1  z from C2 to H1  x from C2 to C3
  C3  z from C3 to N2  x from C3 to C4
  N2  z from C3 to N2  x from N2 to O4
  O3  z from N2 to O3  x from N2 to O4
  O4  z from N2 to O4  x from N2 to O3
  C4  z from C4 to H2  x from C4 to C5
  H2  z from C4 to H2  x from C4 to C5
  C5  z from C5 to N3  x from C5 to C4
  N3  z from C5 to N3  x from N3 to O6
  O5  z from N3 to O5  x from N3 to O6
  O6  z from N3 to O6  x from N3 to O5
  C6  z from C6 to H3  x from C6 to C5
  H3  z from C6 to H3  x from C6 to C5
End
```

Iterated Stockholder Atoms - Distributed Multipole Moments ORIENT Input File

```
! Multipole moments for TNB

! Based on DF-type : ISA-GRID


  O1     4.94664312    -4.54086037    -0.00008315     Type    O      Rank   4
     -0.434562
      0.000003      0.003549     -0.012323
     -0.006455      0.000024     -0.000029      0.324657      0.048761
     -0.000044     -0.073265     -0.193155      0.000080     -0.000082
                    0.516849     -0.077248
      0.060450     -0.000168      0.000056      0.238484     -0.250316
                   -0.000077      0.000134     -0.043458     -0.430454

  O2     1.41192200    -6.56512162    -0.00055747     Type    O      Rank   4
     -0.434846
      0.000003      0.009062     -0.009690
     -0.006538     -0.000012      0.000049     -0.115232      0.306717
     -0.000019      0.200679     -0.033305      0.000129     -0.000121
                    0.520471      0.066946
      0.057293     -0.000010     -0.000102     -0.336318      0.086505
                    0.000009      0.000142      0.386318     -0.190862

  O3    -6.40684762    -2.01269802     0.00239806     Type    O      Rank   4
     -0.433911
      0.000009     -0.012463      0.004092
     -0.005566     -0.000336     -0.000174     -0.203970      0.256848
     -0.000206     -0.126735      0.159557     -0.000139      0.000977
                    0.518804     -0.078443
      0.052667      0.000288      0.000992      0.099210      0.329198
```

```
                   -0.000340       0.000693      -0.354599       0.249889

O4    -6.39181863     2.06052698    -0.00229980    Type  O      Rank   4
    -0.433952
    -0.000012      -0.012583      -0.003836
    -0.005595       0.000328      -0.000185      -0.207929      -0.253721
     0.000173      -0.127936      -0.158512       0.000153       0.000994
                    0.520238       0.066922
     0.053037      -0.000253       0.000984       0.094481      -0.329998
                    0.000439       0.000851      -0.360719      -0.240125

O5     4.98029536     4.50331719     0.00059148    Type  O      Rank   4
    -0.434894
     0.000003       0.003824       0.012259
    -0.006627       0.000046      -0.000013       0.323795      -0.053839
     0.000042      -0.071820       0.193577       0.000213      -0.000042
                    0.518287       0.065246
     0.059032      -0.000085      -0.000085       0.242448       0.249095
                    0.000097      -0.000188      -0.031113       0.433109

O6     1.46131566     6.55482262     0.00013606    Type  O      Rank   4
    -0.435336
    -0.000002       0.008799       0.009967
    -0.006382      -0.000025      -0.000040      -0.120439      -0.304755
     0.000050       0.200612       0.031586       0.000058      -0.000061
                    0.518621      -0.078602
     0.055574      -0.000154       0.000014      -0.340679      -0.079859
                    0.000003      -0.000067       0.392108       0.178398

N1     2.66435302    -4.65393213    -0.00028157    Type  N      Rank   4
     0.810235
    -0.000003       0.036294      -0.064245
    -0.027704       0.000003       0.000005       0.005180       0.008870
    -0.000010       0.030906      -0.054396      -0.000006       0.000006
                   -0.109342       0.000613
    -0.178156       0.000030       0.000044      -0.033693      -0.058006
                   -0.000023      -0.000006       0.087690      -0.156927

N2    -5.36320109     0.01996496     0.00006236    Type  N      Rank   4
     0.808852
    -0.000007      -0.073557       0.000293
    -0.027467       0.000006      -0.000067      -0.009781       0.000034
     0.000004      -0.062567       0.000286      -0.000017      -0.000148
                   -0.109239       0.001140
    -0.178510      -0.000020      -0.000480       0.069878      -0.000659
```

```
                    0.000022      -0.000063      -0.181879       0.002818

N3     2.69894067     4.63389725     0.00024188     Type    N      Rank   4
     0.811408
     0.000011       0.036774       0.063929
    -0.027890       0.000006      -0.000001       0.004999      -0.009021
     0.000020       0.031231       0.054177      -0.000013      -0.000002
                   -0.109602       0.001892
    -0.178765       0.000077       0.000035      -0.033006       0.058276
                    0.000046       0.000004       0.092698       0.153338

C1     1.27935016    -2.23526239    -0.00019086     Type   C1      Rank   4
    -0.032712
     0.000003      -0.009048       0.015821
    -0.173242      -0.000013       0.000017      -0.031757      -0.054933
    -0.000002       0.011433      -0.023187       0.000041       0.000015
                    0.337873      -0.002051
    -0.630863      -0.000748       0.000131       0.081407       0.128559
                    0.000014      -0.000118      -0.079689       0.147361

C2    -1.33083197    -2.28497352     0.00008693     Type   C2      Rank   4
    -0.080825
     0.000009       0.033357       0.057064
    -0.024883      -0.000024      -0.000015      -0.001163       0.001908
     0.000110      -0.054499      -0.093388       0.000000       0.000060
                   -0.347075       0.004005
     0.213270       0.000790      -0.000319       0.018934      -0.036966
                    0.000343       0.000125      -0.293120      -0.494830

C3    -2.57602533     0.00944863     0.00003968     Type   C1      Rank   4
    -0.031782
    -0.000001       0.018184      -0.000075
    -0.172466      -0.000008      -0.000128       0.063883      -0.000462
     0.000002      -0.025066      -0.000522      -0.000026       0.000322
                    0.338356      -0.003927
    -0.642363       0.000051       0.000548      -0.149518       0.001289
                   -0.000019      -0.000482       0.165273      -0.002632

C4    -1.31390380     2.29464702    -0.00000756     Type   C2      Rank   4
    -0.080683
    -0.000015       0.033157      -0.057285
    -0.025054       0.000028      -0.000024      -0.001522      -0.001845
    -0.000070      -0.053201       0.094101      -0.000004       0.000081
                   -0.347294       0.003863
     0.213983      -0.000525      -0.000355       0.020380       0.037265
```

```
                   -0.000333       0.000050      -0.278366       0.501804

C5     1.29573976     2.22593281    -0.00002646     Type   C1      Rank   4
    -0.033349
    -0.000002      -0.009378      -0.015902
    -0.173187       0.000025       0.000003      -0.031154       0.055159
     0.000012       0.011003       0.022545      -0.000006       0.000052
                    0.338387      -0.005586
    -0.630171       0.000278      -0.000002       0.078778      -0.131190
                    0.000072      -0.000009      -0.084007      -0.144135

C6     2.64371154    -0.00967729    -0.00015496     Type   C2      Rank   4
    -0.079249
     0.000012      -0.066154       0.000291
    -0.026220       0.000015       0.000008       0.004004      -0.000027
     0.000008       0.107322      -0.000365       0.000005       0.000047
                   -0.345966       0.003869
     0.209459      -0.000005       0.000227      -0.039056      -0.000666
                   -0.000104      -0.000124       0.575276      -0.008045

H1    -2.35807379    -4.04925572     0.00036850     Type    H      Rank   4
     0.178815
     0.000013      -0.022983      -0.039412
    -0.027313      -0.000005      -0.000011      -0.011337       0.019754
    -0.000004       0.017158       0.029637      -0.000004       0.000021
                    0.022529      -0.000263
     0.030438       0.000050      -0.000021      -0.019960       0.035237
                    0.000007      -0.000043       0.008800       0.014730

H2    -2.32811785     4.06637663    -0.00017952     Type    H      Rank   4
     0.178780
    -0.000012      -0.022690       0.039582
    -0.027351       0.000003      -0.000007      -0.011651      -0.019553
     0.000000       0.016977      -0.029747       0.000003       0.000018
                    0.022492      -0.000224
     0.030430      -0.000053       0.000006      -0.020466      -0.034894
                   -0.000017      -0.000039       0.008545      -0.015022

H3     4.68522599    -0.01721540    -0.00030803     Type    H      Rank   4
     0.177917
    -0.000002       0.045885      -0.000148
    -0.027629      -0.000005       0.000002       0.023042      -0.000166
     0.000002      -0.033832       0.000142      -0.000005       0.000000
                    0.022682      -0.000250
     0.030773      -0.000020      -0.000016       0.041565      -0.000296
```

```
                        0.000003       -0.000005       -0.018183        0.000195
```

Distributed Polarizabilities ORIENT Input File

```
ALPHA   TNB  SITE-NAMES  O   O   RANK 1 TO 1 INDEX   0 FREQSQ       0.0000000
      8.299266620033      0.657534134342      0.001168032541
      0.657534134342      5.261042064169      0.002429494229
      0.001168032541      0.002429494229      5.067264425631
ALPHA   TNB  SITE-NAMES  N   N   RANK 1 TO 1 INDEX   0 FREQSQ       0.0000000
      9.423894900420     -0.015361607683      0.016857855523
     -0.015361607683     11.660479549818      0.003124857303
      0.016857855523      0.003124857303      1.295136469016
ALPHA   TNB  SITE-NAMES  C1  C1  RANK 1 TO 1 INDEX   0 FREQSQ       0.0000000
     13.177116503186     -0.019233973562     -0.015010509176
     -0.019233973562      9.762773188295      0.008227011694
     -0.015010509176      0.008227011694      3.212384411971
ALPHA   TNB  SITE-NAMES  C2  C2  RANK 1 TO 1 INDEX   0 FREQSQ       0.0000000
      4.415786598166      0.025494065077      0.005934670086
      0.025494065077     22.582340489149      0.005675237363
      0.005934670086      0.005675237363      4.722126021275
ALPHA   TNB  SITE-NAMES  H   H   RANK 1 TO 1 INDEX   0 FREQSQ       0.0000000
      1.746282716865     -0.008618178385     -0.013895137756
     -0.008618178385      0.493858247139      0.001103132934
     -0.013895137756      0.001103132934      1.067402233819
ENDFILE
```

Model0-iso ORIENT Input File

```
      ! Pair-Potential: Atom-Atom Potential
      ! Sites:
      !      O N C1 C2 H
      O O rho alpha      C6       C8       C10        C12
         00   00    0     0.569986E+01    0.198934E+01    0.254966E+02
0.203798E+03    0.959376E+04    0.128300E+06
      END
      O N rho alpha      C6       C8       C10        C12
         00   00    0     0.541574E+01    0.210381E+01    0.116931E+02
0.212641E+03    0.123982E+05    0.198203E+06
      END
      O C1 rho alpha      C6       C8       C10        C12
         00   00    0     0.585029E+01    0.199973E+01    0.119190E+02
0.172248E+03    0.320058E+05    0.398233E+06
      END
      O C2 rho alpha      C6       C8       C10        C12
         00   00    0     0.598292E+01    0.183016E+01    0.329903E+02
0.246549E+03    0.245650E+05    0.289621E+06
      END
      O H rho alpha      C6       C8       C10        C12
         00   00    0     0.474065E+01    0.198269E+01    0.869292E+01
0.655161E+02    0.182635E+04    0.194930E+05
      END
      N N rho alpha      C6       C8       C10        C12
         00   00    0     0.498873E+01    0.258187E+01    0.624078E+01
0.192346E+03    0.107350E+05    0.290945E+06
      END
      N C1 rho alpha      C6       C8       C10        C12
         00   00    0     0.558045E+01    0.196324E+01    0.641208E+01
0.175404E+03    0.190577E+05    0.574780E+06
      END
      N C2 rho alpha      C6       C8       C10        C12
         00   00    0     0.577418E+01    0.178715E+01    0.137852E+02
0.233419E+03    0.212275E+05    0.436156E+06
      END
      N H rho alpha      C6       C8       C10        C12
         00   00    0     0.443204E+01    0.180278E+01    0.391061E+01
0.646106E+02    0.329293E+04    0.361222E+05
      END
```

```
      C1 C1 rho alpha       C6      C8      C10       C12
         00  00     0     0.620899E+01    0.201153E+01    0.669735E+01
0.160388E+03    0.279707E+05    0.663064E+06
      END
      C1 C2 rho alpha       C6      C8      C10       C12
         00  00     0     0.605659E+01    0.174603E+01    0.138268E+02
0.181587E+03    0.471683E+05    0.649078E+06
      END
      C1 H rho alpha       C6      C8      C10       C12
         00  00     0     0.462393E+01    0.186344E+01    0.400968E+01
0.510491E+02    0.100250E+05    0.109718E+06
      END
      C2 C2 rho alpha       C6      C8      C10       C12
         00  00     0     0.617586E+01    0.164264E+01    0.449526E+02
0.304137E+03    0.491337E+05    0.559020E+06
      END
      C2 H rho alpha       C6      C8      C10       C12
         00  00     0     0.508006E+01    0.201513E+01    0.113262E+02
0.859296E+02    0.668283E+04    0.715384E+05
      END
      H H rho alpha       C6      C8      C10       C12
         00  00     0     0.363215E+01    0.198405E+01    0.297800E+01
0.217686E+02    0.189172E+03    0.138605E+04
      END
```

Model0-aniso ORIENT Input File

```
      ! Pair-Potential: Atom-Atom Potential
      ! Sites:
      !      O N C1 C2 H
      O O rho alpha       C6      C8      C10       C12
         00  00     0     0.566884E+01    0.200027E+01    0.254966E+02
0.203798E+03    0.959376E+04    0.128300E+06
         00  10     1      0.500728E-01
         00  11c    1     -0.483256E-01
         10  00     1      0.500728E-01
         11c 00     1     -0.483256E-01
         00  20     2     -0.120337E+00
         00  21c    2      0.268936E-01
         00  22c    2      0.912824E-01
         20  00     2     -0.120337E+00
         21c 00     2      0.268936E-01
         22c 00     2      0.912824E-01
      END
      O N rho alpha       C6      C8      C10       C12
         00  00     0     0.544592E+01    0.213467E+01    0.116931E+02
0.212641E+03    0.123982E+05    0.198203E+06
         00  10     1      0.257796E-01
         10  00     1      0.500728E-01
         11c 00     1     -0.483256E-01
         00  20     2     -0.175252E-01
         00  22c    2     -0.784986E-01
         20  00     2     -0.120337E+00
         21c 00     2      0.268936E-01
         22c 00     2      0.912824E-01
      END
      O C1 rho alpha       C6      C8      C10       C12
         00  00     0     0.588042E+01    0.195040E+01    0.119190E+02
0.172248E+03    0.320058E+05    0.398233E+06
         10  00     1      0.500728E-01
         11c 00     1     -0.483256E-01
         20  00     2     -0.120337E+00
         21c 00     2      0.268936E-01
         22c 00     2      0.912824E-01
      END
      O C2 rho alpha       C6      C8      C10       C12
         00  00     0     0.595393E+01    0.182963E+01    0.329903E+02
0.246549E+03    0.245650E+05    0.289621E+06
         10  00     1      0.500728E-01
         11c 00     1     -0.483256E-01
         20  00     2     -0.120337E+00
```

```
         21c 00    2      0.268936E-01
         22c 00    2      0.912824E-01
      END
      O H rho alpha       C6      C8      C10       C12
         00  00    0     0.477593E+01     0.193521E+01     0.869292E+01
0.655161E+02    0.182635E+04    0.194930E+05
         10  00    1      0.500728E-01
         11c 00    1     -0.483256E-01
         20  00    2     -0.120337E+00
         21c 00    2      0.268936E-01
         22c 00    2      0.912824E-01
      END
      N N rho alpha       C6      C8      C10       C12
         00  00    0     0.490457E+01     0.253391E+01     0.624078E+01
0.192346E+03    0.107350E+05    0.290945E+06
         00  10    1      0.257796E-01
         10  00    1      0.257796E-01
         00  20    2     -0.175252E-01
         00  22c   2     -0.784986E-01
         20  00    2     -0.175252E-01
         22c 00    2     -0.784986E-01
      END
      N C1 rho alpha       C6      C8      C10       C12
         00  00    0     0.560217E+01     0.198305E+01     0.641208E+01
0.175404E+03    0.190577E+05    0.574780E+06
         10  00    1      0.257796E-01
         20  00    2     -0.175252E-01
         22c 00    2     -0.784986E-01
      END
      N C2 rho alpha       C6      C8      C10       C12
         00  00    0     0.540571E+01     0.172074E+01     0.137852E+02
0.233419E+03    0.212275E+05    0.436156E+06
         10  00    1      0.257796E-01
         20  00    2     -0.175252E-01
         22c 00    2     -0.784986E-01
      END
      N H rho alpha       C6      C8      C10       C12
         00  00    0     0.399525E+01     0.177254E+01     0.391061E+01
0.646106E+02    0.329293E+04    0.361222E+05
         10  00    1      0.257796E-01
         20  00    2     -0.175252E-01
         22c 00    2     -0.784986E-01
      END
      C1 C1 rho alpha       C6      C8      C10       C12
         00  00    0     0.625156E+01     0.200797E+01     0.669735E+01
0.160388E+03    0.279707E+05    0.663064E+06
      END
      C1 C2 rho alpha       C6      C8      C10       C12
         00  00    0     0.614701E+01     0.171046E+01     0.138268E+02
0.181587E+03    0.471683E+05    0.649078E+06
      END
      C1 H rho alpha       C6      C8      C10       C12
         00  00    0     0.476364E+01     0.184384E+01     0.400968E+01
0.510491E+02    0.100250E+05    0.109718E+06
      END
      C2 C2 rho alpha       C6      C8      C10       C12
         00  00    0     0.613784E+01     0.160748E+01     0.449526E+02
0.304137E+03    0.491337E+05    0.559020E+06
      END
      C2 H rho alpha       C6      C8      C10       C12
         00  00    0     0.513878E+01     0.196339E+01     0.113262E+02
0.859296E+02    0.668283E+04    0.715384E+05
      END
      H H rho alpha       C6      C8      C10       C12
         00  00    0     0.357276E+01     0.197229E+01     0.297800E+01
0.217686E+02    0.189172E+03    0.138605E+04
      END
```

DIFF ORIENT Input File

```
      ! Pair-Potential: Atom-Atom Potential
```

```
      ! Sites:
      !      O N C1 C2 H
      O O rho alpha       C6      C8      C10       C12
         00  00    0     0.568097E+01    0.199363E+01    0.254965E+02
0.203797E+03    0.959360E+04    0.128299E+06
~
      END
      O N rho alpha       C6      C8      C10       C12
         00  00    0     0.539913E+01    0.196850E+01    0.116930E+02
0.212641E+03    0.123980E+05    0.198202E+06
~
      END
      O C1 rho alpha       C6      C8      C10       C12
         00  00    0     0.566392E+01    0.188182E+01    0.119189E+02
0.172249E+03    0.320057E+05    0.398248E+06
~
      END
      O C2 rho alpha       C6      C8      C10       C12
         00  00    0     0.608513E+01    0.181833E+01    0.329903E+02
0.246547E+03    0.245649E+05    0.289621E+06
~
      END
      O H rho alpha       C6      C8      C10       C12
         00  00    0     0.468381E+01    0.183627E+01    0.869282E+01
0.655152E+02    0.182636E+04    0.194931E+05
~
      END
      N N rho alpha       C6      C8      C10       C12
         00  00    0     0.493905E+01    0.253092E+01    0.624072E+01
0.192344E+03    0.107351E+05    0.290947E+06
~
      END
      N C1 rho alpha       C6      C8      C10       C12
         00  00    0     0.570289E+01    0.190190E+01    0.641211E+01
0.175403E+03    0.190577E+05    0.574786E+06
~
      END
      N C2 rho alpha       C6      C8      C10       C12
         00  00    0     0.557727E+01    0.170088E+01    0.137852E+02
0.233418E+03    0.212276E+05    0.436146E+06
~
      END
      N H rho alpha       C6      C8      C10       C12
         00  00    0     0.400508E+01    0.175312E+01    0.391049E+01
0.646125E+02    0.329294E+04    0.361223E+05
~
      END
      C1 C1 rho alpha       C6      C8      C10       C12
         00  00    0     0.625305E+01    0.199735E+01    0.669739E+01
0.160389E+03    0.279707E+05    0.663074E+06
      END
      C1 C2 rho alpha       C6      C8      C10       C12
         00  00    0     0.605877E+01    0.150547E+01    0.138267E+02
0.181588E+03    0.471682E+05    0.649096E+06
      END
      C1 H rho alpha       C6      C8      C10       C12
         00  00    0     0.482690E+01    0.178865E+01    0.400952E+01
0.510493E+02    0.100251E+05    0.109716E+06
      END
      C2 C2 rho alpha       C6      C8      C10       C12
         00  00    0     0.621681E+01    0.165041E+01    0.449515E+02
0.304136E+03    0.491335E+05    0.559018E+06
      END
      C2 H rho alpha       C6      C8      C10       C12
         00  00    0     0.533222E+01    0.194477E+01    0.113262E+02
0.859303E+02    0.668292E+04    0.715369E+05
      END
      H H rho alpha       C6      C8      C10       C12
         00  00    0     0.357561E+01    0.196938E+01    0.297810E+01
0.217684E+02    0.189172E+03    0.138606E+04
      END
```

Appendix 5.B – Supplementary analysis

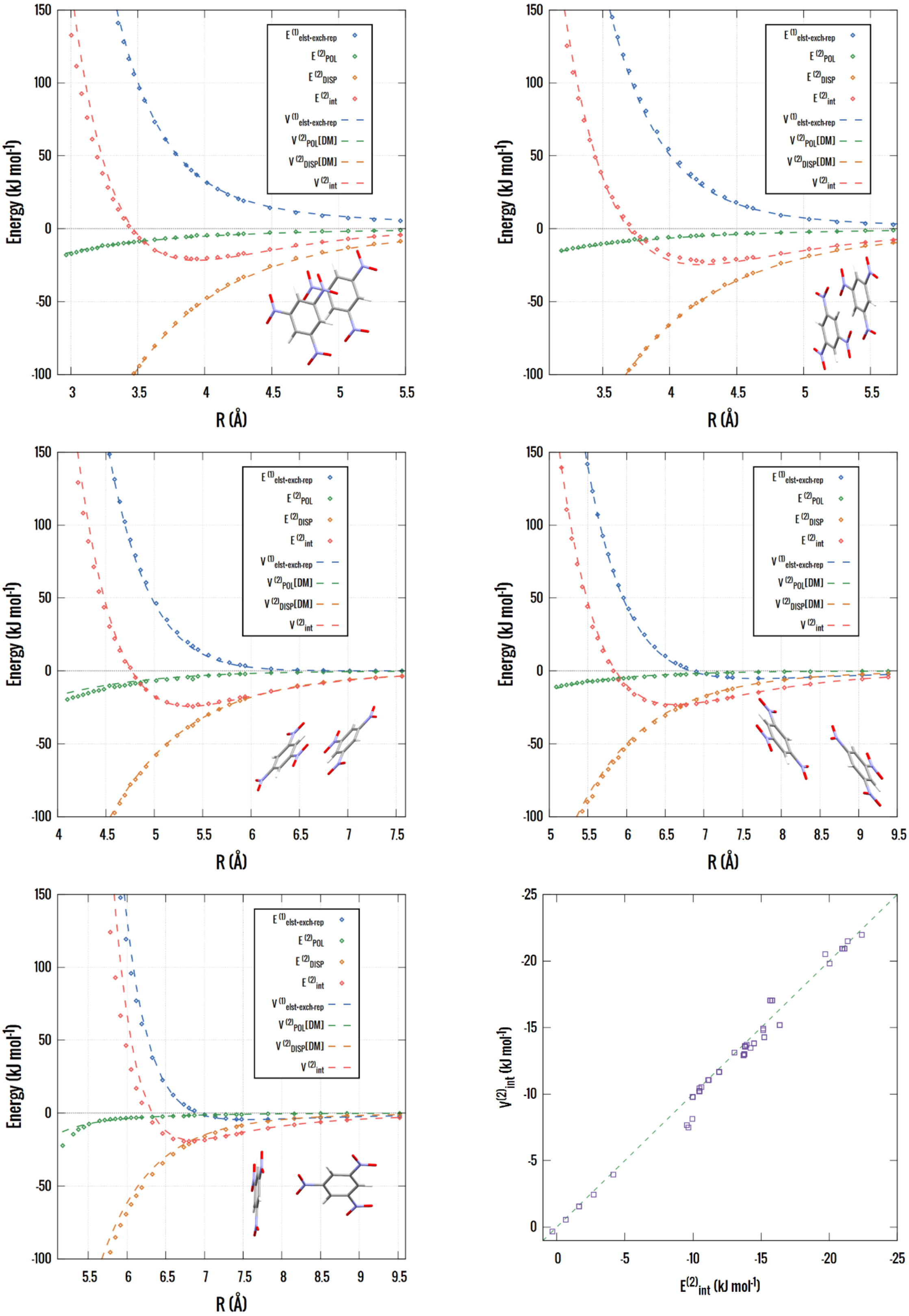


Figure 43: *The first 5 graphs compare the* $E_{int}^{(2)}[SAPT(DFT)]$ *energy breakdown (points) with the Model0-aniso (dashes) for the 5 gas-phase dimers, while the last graph compares* $E_{int}^{(2)}[SAPT(DFT)]$ *&* $V_{int}^{(2)}[Model]$ *for crystal structures using SAPT(DFT) (points) with the dashed line signifying equality.* $\beta_{disp} = 1.6$.

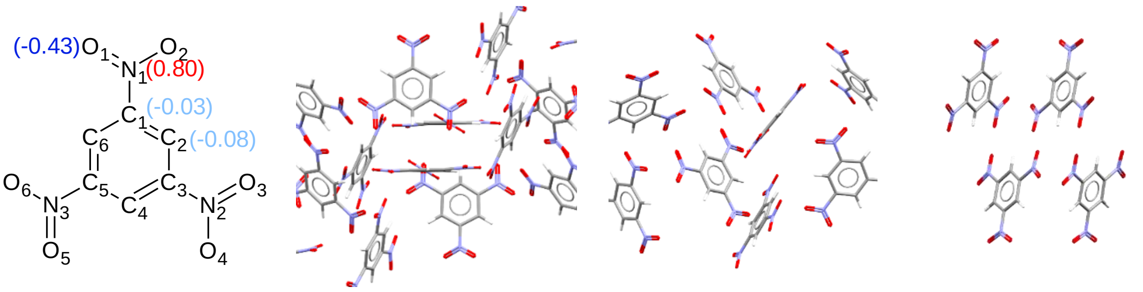
6 From the gas-phase to the crystalline state: Trinitrobenzene

The methodology for developing a rigid non-empirical intermolecular atom-atom force-field for trinitrobenzene (TNB) has been established and detailed in **Chapter 5**. The workflow for TNB builds upon that of pyridine (**Chapter 3**), with a few adaptions to existing codes. The non-empirical distributed intermolecular force-field (DIFF) has been tested on the most significant gas-phase dimers and two molecule contacts in the TNB crystal structures, and shown to reproduce the SAPT(DFT) dimer energies well (**5.6.2.3**). This chapter uses the intermolecular force-fields to model the crystals of TNB, and utilizes these models in a Crystal Structure Prediction (CSP) study. Since, it is not possible to minimize the crystal structures with the explicit modelling of the polarization term using the distributed polarizabilities (**2.6.1.1**), potentials where TNB's small polarization contribution (***Figure 42***) has been absorbed into the electrostatic term are considered and minimized.

In this chapter, the complete non-empirical model that has been relaxed to 2nd order SAPT(DFT) interaction energies ($E_{int}^{(2)}[SAPT(DFT)]$) of 327 TNB dimers will be referred to as the Distributed Intermolecular Force-Field (DIFF). The similarities and differences between the DIFF for trinitrobenzene (TNB) and that for pyridine[38, 167] were mentioned in **Chapter 5**, and this chapter examines how differences in the molecular conformation of the observed and optimized structure, symmetry, and the relative importance of the different contributions to the intermolecular forces affects the crystal lattice. The DIFF uses distributed multipole moments, polarizabilities and dispersion coefficients derived from the molecular charge distribution using the Iterated Stockholder Atoms (ISA) approach,[233] and an anisotropic atom-atom repulsion model, which was fitted to over 2000 $E_{int}^{(1)}[SAPT(DFT)]$ calculations. The DIFF reproduces the pair potential energy surface of TNB as discussed in **Chapter 5**, but in the crystal structure, modelling with the DIFF includes the non-pairwise-additive polarization term. The reliability of non-empirical atom-atom intermolecular force-fields constructed from 1st order SAPT(DFT) dimer interaction energies are also investigated, as these 1st order model potentials are cheaper to develop and could be used for larger systems. These 1st order Model0s, are the preliminary unrelaxed non-empirical force-fields (**5.4.5**). The effects of anisotropy on the short-range contribution is also studied by comparing Model0-iso with Model0-aniso.

6.1 Trinitrobenzene (TNB)

Trinitrobenzene (TNB) is an aromatic explosive that is considered the basic building block for contemporary energetic materials today.[148] TNB is considered a high explosive, that is extremely volatile in its dehydrated powder form.[148] Accordingly, there are many safety precautions associated with experimental studies of the material, limiting the amount of readily available observed data. Even though it was first discovered in 1883,[391] a determination of its crystal structure was not made until 1972,[368] almost a century after. The nitrated-benzene derivative has a melting point of $\sim 400K$,[148, 392] so is solid at room temperature and pressure. Consequently, one would expect the force-fields empirically fit to ambient data to model TNB's intermolecular interactions better than they did pyridine's (**Chapter 3**), as pyridine is a liquid at R.T.P. In the field of crystal structure prediction (CSP), TNB would be considered a medium sized molecule, with 18 atoms, 108 electrons and 3 flexible nitro-groups (NO_2). It is certainly a more challenging molecule for CSP than pyridine and the conformational dependence of its molecular properties and intermolecular interactions were examined in **Chapter 4**.



TNB | Form I | Form II | **Form III**

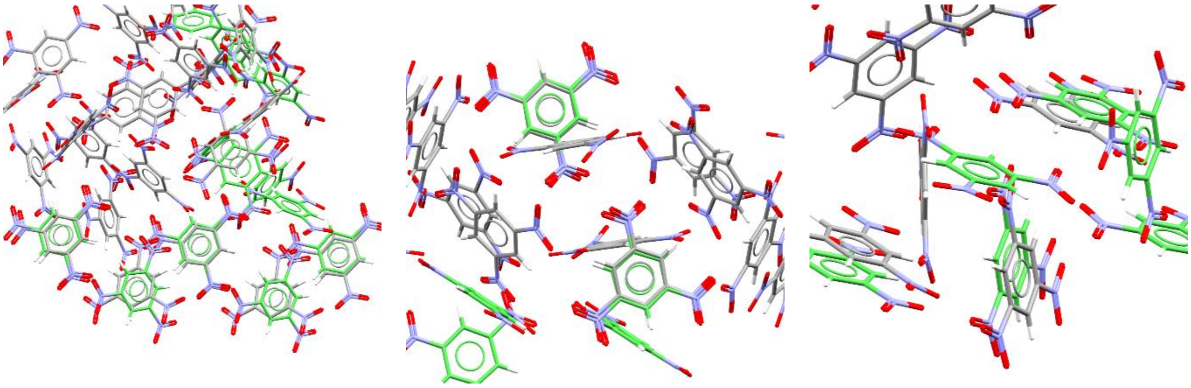
REFCODE	TNBENZ11[263]	TNBENZ12[263]	TNBENZ13[263]
Conditions	Ambient Pressure, $183K$	Ambient Pressure, $120K$	Ambient Pressure, $183K$
Density ($g\ cm^{-3}$)	1.729	1.688	1.717
NO_2 torsion angles (°)	181, 176, 212, 176, 191, 171	178, 176, 184, 185, 181, 193	177, 172, 180

Figure 44: *The polymorphic crystal structures of the energetic molecule trinitrobenzene (TNB).*[263] *The form, Cambridge Structural Database (CSD) reference code, determination conditions, density and NO_2 torsion angles (°) are given below each crystal structure. These were obtained from the experimental Cambridge Structural Database*[359-361] *entries, and visual representations of the structures constructed using CCDC Mercury 3.6.*[291, 362] *The Iterated Stockholder Atoms (ISA) atomic charges computed for the optimized isolated molecular structure at the PBE0/aug-cc-pVTZ level are given in parenthesis in red or blue for positively or negatively charge nuclei.*

From sampling the experimental crystal structures in the Cambridge Structural Database (CSD) in **Chapter 4**, it was found that at ambient temperatures and pressures, the observed conformation of a nitro-group adjacent to two hydrogen atoms does not change significantly from the planar conformation of the optimized isolated molecular structure. Consequently, TNB has been held rigid throughout this investigation. This is an approximation (***Figure 44***) but a necessary simplification for deriving the DIFF (**Chapter 5**). The effect of idealising the molecular conformation is examined in **6.3.2.2**, otherwise the isolated molecular is used in this chapter. The TNBENZ13[263] experimental structure was optimized using the Gaussian09 program[235] and the PBE0/d-aug-cc-pVTZ[203-205, 214] level of theory. The overlays of the three known polymorphs of TNB in ***Figure 45***, show that even though the molecular conformation of each crystal is similar the configurations and close contacts within the crystalline structures differ significantly, with only $2-5$ out of 30 molecules shared between crystal structures. As the polymorphs are so different, the balance of intermolecular interactions in each polymorph varies significantly.

Pyridine is a polar molecule but TNB is not due to the symmetry of its optimized structure. Polarization was found to be a relatively small contribution to the intermolecular energy for TNB's gas-phase dimers (**Chapter 5**), however, it is a non-additive contribution and could prove important to certain configurations in the crystal. It is important to test if the model currently used to estimate polarization effects in CSP (the polarizable continuum model) has any significant effect on the structures that correspond to lattice energy minima (**6.2.2**).

As TNB is an energetic material, its behaviour at elevated temperatures and pressures are of great interest in industry.[137, 350, 355, 393, 394] A genuine description of the repulsive interactions in the short-range is required to accurately measure up the repulsive wall of its potential energy surface (PES). The very weak hydrogen bonds formed by $C-H \ldots O$ and $C-H \ldots N$ contacts, combined with $NO_2 \ldots NO_2$ contacts has been found to greatly influence the crystal structure configurations of TNB.[263, 395] Consequently, it is found that the most sampled two molecule contacts in the crystal differs significantly from those seen in the gas phase (***Figure 44*** & **5.5.3.2**).



Reference	Comparison	Reference	Comparison	Reference	Comparison
Form I TNBENZ11	Form II TNBENZ12	Form I TNBENZ11	Form III TNBENZ13	Form II TNBENZ12	Form III TNBENZ13
N/30 ($RMSD_5$ Å)		N/30 ($RMSD_2$ Å)		N/30 ($RMSD_3$ Å)	
5/30 (1.48)		2/30 (0.35)		3/30 (1.00)	

Figure 45: *The optimum overlays out of 30 molecules of the 3 experimentally known trinitrobenzene polymorphs. The overlays show very few common contacts between the 3 comparisons. The reference molecule is in grey and the comparison molecule in green.*

The polymorphism of TNB was first reported after being induced by the additive trisindane,[263] which resulted in the serendipitous crystallization of forms II ($Z' = 2$) and III ($Z' = 1$) in attempted co-crystallization experiments. Additive induced polymorphism highlights the value of CSP for predicting polymorphs that only appear in unusual crystallization conditions.[263, 396-399] The crystal structures of both $Z' = 2$ forms of TNB show V-shape troughs (***Figure 44***) heavily influenced by polar contacts. The $Z' = 1$, centrosymmetric form III exhibits a more "normal" packing, influenced more by crystal packing forces with no real distinguishable underlying structural motifs.[263, 400] The authors relied on crude lattice energy calculations, using the Dreiding force-field[401] with empirical atomic charges, to suggest that form III was the most stable form, even though form I was the observed densest. Form III was calculated to be almost $25\ kJ\ mol^{-1}$ more stable than form I and $\sim 1.4\ kJ\ mol^{-1}$ more stable than form II.[263] These calculations were very tentative as they were done using a very approximate model.[263] An energy difference of $25\ kJ\ mol^{-1}$ is way beyond the believed energy difference of typical polymorphs but not unheard of.[7] The energy difference is also surprising as form III was discovered over 100 years after the determination of form I.[391] What is certain, is that form I is the kinetically favourable polymorph as it readily crystallises, while forms II and III require additives, and have problems of reproducing single crystal growth. Form II could not be reproduced in another lab,[263] which is a reminder that the appearance, disappearance, isolation and reproduction of polymorphs is still mysterious and can be very erratic. If TNB is monotropic, the melting points of $121.8°\text{C}\ (I)$, $119.2°\text{C}\ (II)$ & $122.7°\text{C}\ (III)$ suggest that form III is the most stable form. However, the polymorphs may be enantiotropically related, and the use of the DIFF should give a much better estimate of the relative stability order of the static lattice. Since there is no recently published experimental polymorph screening for TNB, it is the perfect candidate to explore whether a DMACRYS[59] $Z' = 1$ CSP study using the DIFF PES can give a realistic prediction of further polymorphs.

6.2 Methodology

6.2.1 Models

The CSP study of TNB focuses on one empirical and three non-empirical force-fields:

- The empirical FIT model,[40] uses an exp-6 Buckingham potential with an undamped C_6 isotropic dispersion model. Its short-range parameters were originally fitted to aza-hydrocarbon crystal structures at ambient conditions (**1.2.2.1**).[72] The empirical model is combined with two electrostatic models in this study. GDMA, which is often used with FIT in CSP studies, and ISA-DMA to have the same electrostatic model as the non-empirical potentials (**2.5**).
- Model0-iso, is a non-empirical intermolecular force-field derived from 2000 1st order SAPT(DFT) TNB dimer interaction energies ($E_{int}^{(1)}[SAPT(DFT)]$), which utilizes an isotropic atom-atom short-range model.
- Model0-aniso, is analogous to Model0-iso, however, Model0-aniso employs an anisotropic atom-atom short-range model.
- The distributed intermolecular force-field (DIFF), is the final non-empirical atom-atom intermolecular force-field generated from the relaxation of the $\alpha_{00}^{\iota\kappa}$ and $\rho_{00}^{\iota\kappa}$ short-range parameters to the 2nd order SAPT(DFT) dimer interaction energies $E_{int}^{(2)}[SAPT(DFT)]$ of 327 pseudo-random and important TNB gas-phase dimers (**5.6.2.1**).

The non-empirical potentials all have the same long-range electrostatic, polarization and dispersion models, as specified in **Chapter 5**. Approximations to the DIFF and Model0 are used for lattice energy minimizations because of the impossibility of minimizing with an explicit polarization model in DMACRYS. However, all other quoted lattice energies, employing the non-empirical models, include the fixed-geometry explicit polarization except for DIFF(no-pol), which is the DIFF model with the multipolar polarization energy $V_{pol}^{(2-\infty)}[DM]$ not included. The influence and importance of the polarization contribution, particularly for relaxing structures, is examined by using the polarizable continuum model (PCM) to calculate the electrostatic model (6.2.2). The functional forms of the empirical and non-empirical models are not given in this chapter as they are identical to those used in Equations 86 and 87 in **Chapter 3** and detailed in **Chapters 1, 2** & **5.** The history and details of the FIT model can be found in **Chapter 1**, while the theory behind non-empirical force-fields and the development of the bespoke potentials for TNB can be found in **Chapters 2** & **5**. The ORIENT[108] input files for the non-empirical models can be found in **Appendix 5.A**, and the DMACRYS potential input files used can be found in the **Appendix CD**. The non-empirical force-fields include the Iterated Stockholder Atoms (ISA)[231, 233] derived distributed multipole moments and polarizabilities. The quality and the methodology for deriving the electrostatic and polarizability models can be found in **Chapter 5**.[377, 379, 382]

The CamCASP computed ISA-Pol atomic polarizabilities must undergo rotation and truncation in order to be used in DMACRYS for lattice-energy minimizations, as detailed in **2.6.1.2**. Further details on the electrostatic and polarizability models like their values and relationships can be found in **Chapter 5** & **Appendix 5A.** The FIT potential is coupled with the same ISA electrostatic model as the DIFF, and a Gaussian-Distributed Multipole Analysis (GDMA)[230] derived electrostatic model (**Chapter 2**), as it is often used in CSP studies.[19] The GDMA electrostatic model was obtained from the molecular charge density at the PBE0/6-31G(d, p) level of theory. A different level of theory is used here to the ISA model as this is the standard level of theory used in CSP studies[19] and used to compare our models to current methodologies. The FIT model absorbs induction effects into its empirically fitted short-range parameters, so no explicit polarization model is used.

6.2.2 Polarizable continuum model (PCM)

6.2.2.1 Influence on the electrostatic model

Computing and including the contribution of polarization to the lattice energy is an advanced treatment. For a number of molecular modelling and dynamics codes the polarization term cannot be explicitly included, particularly for modelling organic crystal structures as the induced multipole moments need to be iterated to be self-consistent and all evaluations of the forces, torques and second derivatives have to be evaluated numerically (**2.6.1.3**). Thus a simplified yet accurate functional form that implicitly models the polarization is required. To address this issue and also investigate the importance of the polarization contribution to the relative energies and structures in Crystal Structure Prediction (CSP), non-empirical intermolecular force-fields with implicitly absorbed approximate polarization effects were created. In CSP, the polarizable continuum model (PCM) is usually used to estimate polarization effects and applied as a post-minimization to determine how sensitive the relative lattice energies are to the difference between the charge distribution calculated in vacuum and in a polarizable continuum. As the polarization term in the original model is derived from the charge distribution it is most appropriate to have this term is absorbed into the electrostatic model. The long-range effects of polarization can be absorbed into the electrostatic model by recalculating TNB’s multipole moments in a dielectric polarizable continuum solvation model (PCM) (**2.6.1.4**).

With reference to the solid dielectric constant of nitrobenzene ($3.44D\ at\ 0°C$)[261] and previous work in determining a suitable "general" dielectric constant (ε) for organic solids,[260, 261] TNB's multipole moments were recalculated under a PCM with dielectric constants (ε) of 3, 7, and 11. The effects of each dielectric constant on the isolated molecule can be seen in ***Figure 46***, where $\varepsilon_r = 1$ corresponds to vacuum, and is the multipole moments under no external electric field.

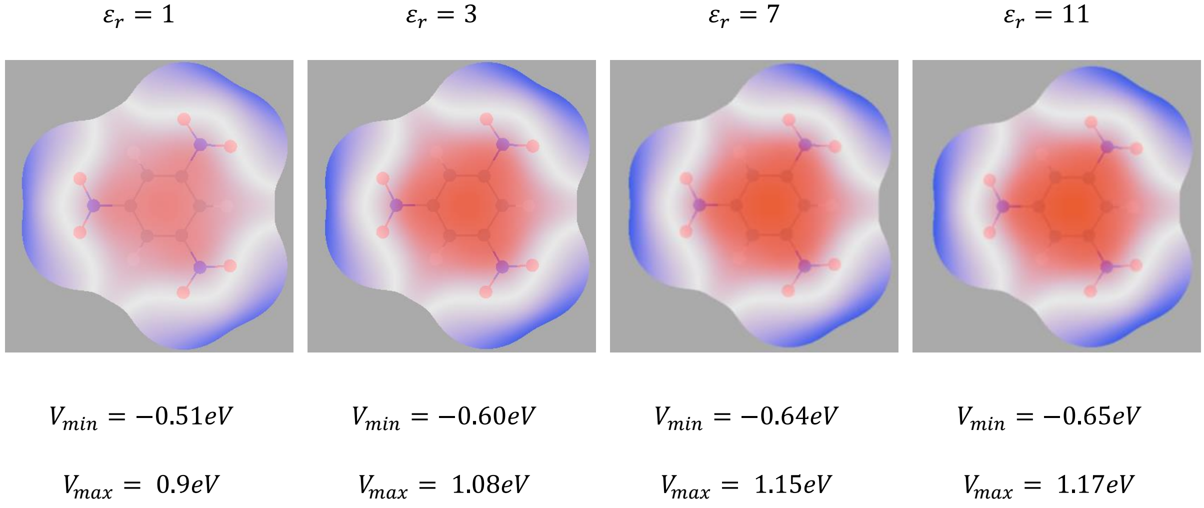


Figure 46: *The electrostatic potential (eV) computed using the ISA distributed multipole analysis (ISA-DMA) on the iso-density surface of* 10^{-3} $electron/bohr^3$ *around the static optimized isolated molecular structures of TNB. The changes in charge density around the TNB molecule, and potential maxima and minima under the various dielectric constants (ε) is observed on a potential scale of +1.5 eV (red) to −1.0 (blue) eV.*

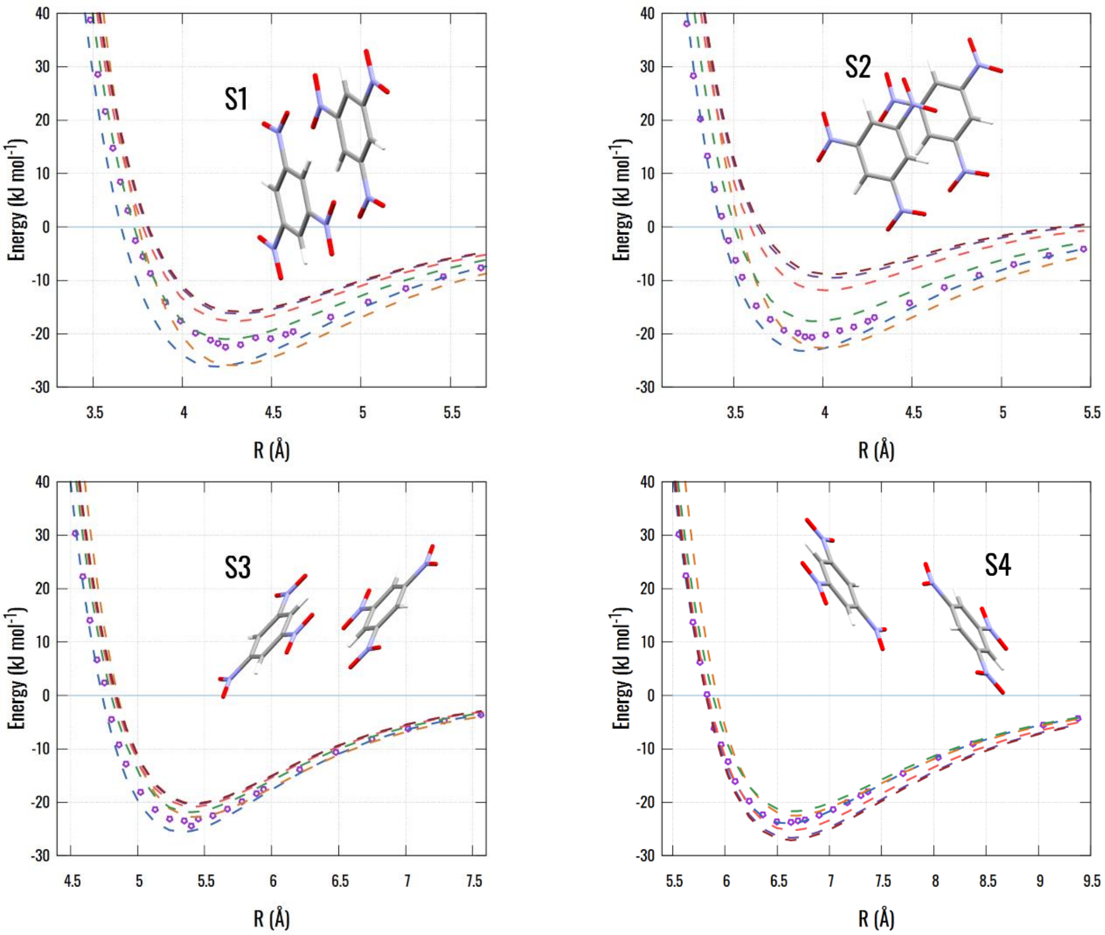
An alternative is to derive the dielectric constant of the molecule using the Clausius–Mossotti relation (Equation 108).[262] This relation derives from the work of the two physicists, though the formula was initially used in the investigation of indices of refraction. With regards to dielectric constants, this theory relates the atomic or molecular polarizability and the molecular density in SI units to the dielectric constant (relative permittivity) of a material, ε_r. The following approximation of the relation assumes the polarizability across the material to be homogeneous and not orientation dependent as it is known to be with water.[165]

$$\frac{\varepsilon_r - 1}{\varepsilon_r + 2} = \frac{N\alpha}{3\varepsilon_0} \tag{108}$$

where, N is the number of molecules per cubic metre, α is the molecular polarizability and ε_0 is permittivity in vacuum. From the charge density calculations in **Chapter 5** TNB's molecular and distributed ISA-Pol polarizabilities were obtained. The total molecular dipole-dipole static polarizability was found to be 121.2 in the atomic units of polarizability ($e^2 a_0^2/E_h$). a.u., computed with CamCASP[107] using the PBE0/AC functional[203-205] and ALDA+CHF hybrid kernel. The experimental density of TNB is approximately 1.7 $g\ cm^{-3}$,[263] and its molecular mass 213 $g\ mol^{-1}$. Therefore, from Equation 108 one obtains a relative permittivity ε_r for TNB of 2.72. This is very close to, and could be rounded up to, the ε_r of 3 that was once proposed as the most suitable "general" dielectric constant for organic solids.[260]

6.2.2.2 *Influence on the gas-phase*

The influence of using a PCM to represent the 2nd order polarization term ($E_{POL}^{(2)}$) in the gas-phase dimer interaction energy with the non-empirical force-field, Model0-aniso, was examined in ***Figure 47***. This is to help assess how using a simpler functional form than the DIFF affects intermolecular interactions. This is an important enquiry, especially if the end goal is to scale this workflow up to larger more complex and computationally challenging systems. It is vital to know what approximations can be made to still produce an accurate non-empirical intermolecular force-field.



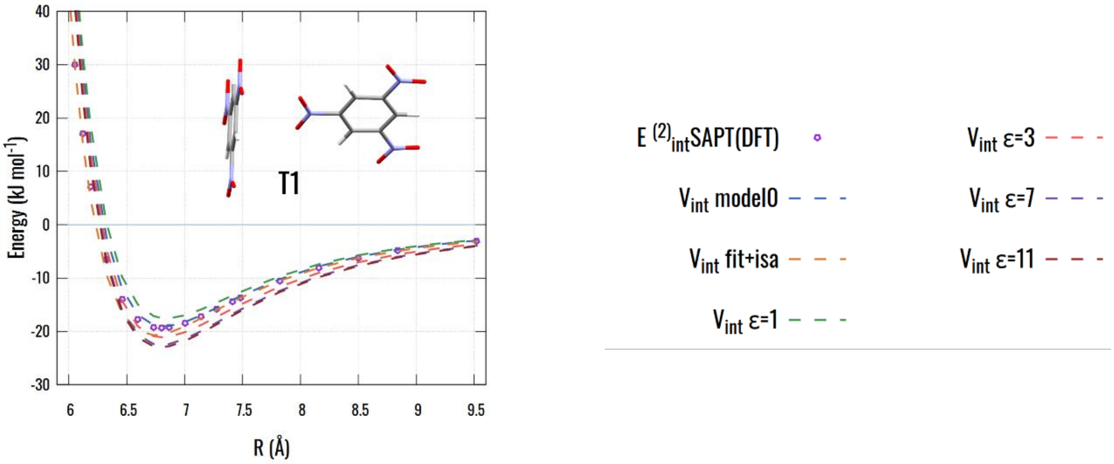


Figure 47: *The intermolecular potential energy* (V_{int}) *with various models compared against the 2nd order SAPT(DFT) interaction energy* ($E^{(2)}_{int}[SAPT(DFT)]$) *for the 5 gas-phase minima of TNB. The force-fields used: the empirical FIT+ISA electrostatic model* ($Rank\ l = 4$)*, the 1st order model0-aniso+ISA electrostatic model* ($Rank\ l = 4$) *and an explicit polarization model (**Chapter 5**), and model0-aniso with the ISA electrostatic model with the four different dielectric constants* ε_r *instead of an explicit polarization model.*

Increasing the dielectric enhances the electrostatic potential gradient of the molecule as one would expect (***Figure 46***). By placing our TNB molecule in an isotropic dielectric, the oxygen surface is found to become more negative and the ring centre more positive. The PCM favours configurations where the positive and negative nodes are close (***Figure 46***). This is the case with the T1 and S4 TNB dimers (**Chapter 5 *Figure 36***), where the PCM is increasingly over-stabilising, with increasing ε. However, configurations where the two molecules are stacked results in increasingly repulsive interactions with increasing dielectric constant. The polarizability is highly directional (the resultant electric field felt is also very directional) but a PCM extremely simplifies the model, failing to capture the anisotropic TNB molecular environment and the anisotropy of other surrounding TNB molecules. Hence, the COSMO PCM fails to mimic polarization effects in the gas-phase. Conversely, the non-additivity of polarization means this error may be much less pronounced in the condensed phase.

The differences between the well-depths of the different dimer configurations (***Figure 47***) with different polarization models, suggests that the polarization would need to be explicitly modelled for studying the gas-phase dimers of TNB. Unsurprisingly, no 2nd virial coefficient data is readily available for TNB, due to the difficulties associated with energetic experiments. So no comparison of the force-fields with gas-phase experimental data could be made for TNB unlike pyridine in **Chapter 3**.

6.2.3 Lattice energy minimization calculations

The lattice energies calculated in this CSP study are the lattice summation over the atom-atom intermolecular interactions. TNB is held rigid, and so the change in intramolecular energy between different crystal structures $\Delta E_{intra}^{molecule} = 0$. Lattice energy minimizations were conducted on the TNB crystal structure in two conformations: its experimentally observed conformations (TNBENZ1X, with standardised $C - H$ bond lengths), these are referred to as *xminexpt* calculations, and its planar rigid conformation, referred to as *xminopt* calculations, in order to see how well the rigid planar structure reproduces the known experimental structures (***Table 27***). In *xminopt* calculations, the optimized isolated molecular structure is pasted into the experimentally observed crystal configurations and minimized. The lattice energy for the crystal structures was minimized using DMACRYS2.2.1.1,[40] with lattice summations being carried out to 15 Å followed by a 2 Å splined correction, with the charge-charge, charge-dipole and dipole-dipole electrostatic contributions being evaluated by Ewald summation. The treatment and workflow used here was identical to that used in **Chapter 3**. During the lattice energy minimization, the forces due to the polarization are not explicitly calculated by DMACRYS as the polarization term is a non-additive contribution and the induced moments have to be solved iteratively to consistency[40] by taking their numerical derivatives (**2.6.1.3**). For the polar molecule pyridine, the lattice energy was minimized with respect to the polarization forces in a very expensive calculation (**Chapter 3** ***Table 6***) but showed to not contribute significantly to changes in cell geometry, thus the forces due to explicit polarization have been omitted for TNB. Instead, this CSP study attempts to model the effects of the polarization forces using a cheaper method, the polarizable continuum model (**6.2.2**). Accordingly, the lattice energy minimizations were also carried out using the DIFF and various electrostatic models under different dielectric constants ($\varepsilon_r = 3, 7$ & 11).

6.2.4 Crystal Structure Prediction (CSP)

The TNB CSP study employed CrystalPredictor 2.0.1[35, 36] to use the optimized isolated molecular structure of TNB as a starting point to search within the 59 most probable space groups[403] and generated 670112 putative $Z' = 1$ crystal structures. The lattice energies of these structures were calculated using the FIT potential and the ISA point charges from the electrostatic model created in **Chapter 5**. The structures were then clustered (**1.1.2.2**) leaving 13956 unique structures. These unique structures were used as the starting point of our study. This is a novel approach as previously in **Chapter 3** and in the Blind Tests,[22] the more elaborate force-fields were used as a post-processing treatment after minimisation with a good intermediary force-field that included distributed multipoles. As a result, the small changes in the force-fields rarely resulted in a significant change in the final lowest energy hypothetical structures, as found for pyridine (**3.3.3.1**). The differences between a point-charge and distributed multipole model, and between the FIT and the non-empirical potentials are likely to be large for TNB, and hence minimisations starting directly from the CrystalPredictor generated structures are likely to sample more of the potential energy surface. Following the investigation in **6.3.1**, all minimizations were performed with the force-fields without any polarization ($\varepsilon_r = 1$), unless otherwise specified. For the non-empirical potentials, the lattice energy was then revaluated in a single-point calculation to include the polarization energy, iterated using the distributed polarizabilities, permanent and induced multipole moments.

After the ~ 14000 structures had been minimized with each model, they were clustered once more and the most stable structures were analysed for their similarities with each other, the experimental forms and the gas-phase minima from **Chapter 5**. The crystal structure similarity tool in Mercury 3.6[291] was used to determine the number of molecules in a maximum coordination cluster (30 for crystals, 2 for dimer comparisons), which matched within a 30% distance in intermolecular atom-atom distances and 30° in interatomic intermolecular angles. The root mean squared deviation ($RMSD_n$) of overlaid molecules was calculated ignoring hydrogen atoms. The number of molecules overlaid was increased to 30 for TNB because of the changes in common CSP practise and the flexibility of TNB, thus the size of the cluster required is larger and the RMSD overlay constraints for finding a match are looser than for pyridine (**Chapter 3**). One should recall that the resultant lattice energy minima from re-minimizing the 14000 structures with the FIT+ISA model was used in the development of the non-empirical force-fields in **Chapter 5** to obtain the most popular two molecule contacts sampled in the lowest energy structures (***Figure 51*** (c)), which were then used as a test dataset for the potentials. The force-field will have performed "well" if the $Z' = 1$ form III is found in the CSP search and the three polymorphs of TNB are amongst the most stable generated structures with a reasonable energy difference considering their melting points (**6.1**).[263]

6.3 Results

6.3.1 The influence of implicit polarisation on cell geometry

Does including implicit polarization by using molecular charge densities calculated empirically in a polarizable continuum improve on traditional empirical methods, which simply absorb polarization effects alongside many other errors? Absorbing the contribution gives us a loose indication of the importance of polarization on the minimized crystal structure. As previously mentioned, one cannot minimize the structures with the explicit polarization model, thus a direct comparison of the PCM and the explicit model cannot be made. Form III of TNB was examined with both isotropic and anisotropic versions of the 1st order model (Model0) and the DIFF to investigate whether the crystal lattices undergo any significant structural changes and variations in the lattice energies when optimized with the non-empirical models compared to the empirical force-fields (***Table 27***).

Form III	Model0-Aniso					FIT+ISA	FIT+GDMA
	Explicit	1	3	7	11		
E_{latt}	−109.6	−98.3	−105.2	−107.8	−108.6	−100.6	−92.2
$V_{elst}^{(1)}[DM]$	−37.4	−37.4	−49.1	−47.9	−48.7	−30.6	−21.6
$V_{POL}^{(2-\infty)}[DM]$	−11.3	~	~	~	~	~	~
Density	1.707	1.707	1.719	1.724	1.725	1.705	1.692
$RMSD_{30}/\text{Å}$	0.774	0.774	0.782	0.785	0.787	0.412	0.306

Form III	Model0-Iso					FIT+ISA	FIT+GDMA
	Explicit	1	3	7	11		
E_{latt}	−108.6	−96.4	−103.3	−105.9	−106.7	−100.6	−92.2
$V_{elst}^{(1)}[DM]$	−37.9	−37.9	−44.2	−47.0	−47.8	−30.6	−21.6
$V_{POL}^{(2-\infty)}[DM]$	−12.0	~	~	~	~	~	~
Density	1.697	1.697	1.709	1.714	1.715	1.7045	1.692
$RMSD_{30}/\text{Å}$	0.633	0.633	0.658	0.666	0.668	0.412	0.306

Form III	DIFF					FIT+ISA	FIT+GDMA
	Explicit	1	3	7	11		
E_{latt}	−108.8	96.6	−103.5	−106.1	−106.9	−100.6	−92.2
$V_{elst}^{(1)}[DM]$	−36.8	−36.8	−44.6	−47.4	−48.3	−30.6	−21.6
$V_{POL}^{(2-\infty)}[DM]$	−12.2	~	~	~	~	~	~
Density	1.715	1.715	1.728	1.732	1.734	1.7045	1.692
$RMSD_{30}/\text{Å}$	0.489	0.489	0.526	0.534	0.538	0.412	0.306

Table 27: *Electrostatic models that include the effects of polarization using the PCM are compared for various ε (*$\varepsilon_r = 1{,}3{,}7{,}11$*), the explicit polarization model and the empirical FIT potential using the* $\varepsilon_r = 1$ *(in vacuum) electrostatic model. The lattice energy (*E_{latt}*), the multipolar electrostatic (*$V_{elst}^{(1)}[DM]$*) and damped multipolar polarisation (*$V_{POL}^{(2-\infty)}[DM]$*) are given in kJ mol*$^{-1}$*. The structural changes are compared in the lattice density (given in g cm*$^{-1}$*) and the 30 molecule structural overlays of the minimized structures and the experimental form III (TNBENZ13*^{263}*, experimental density = 1.717 g cm*$^{-1}$*). The RMSD*$_{30}$ *values are given in* Å.

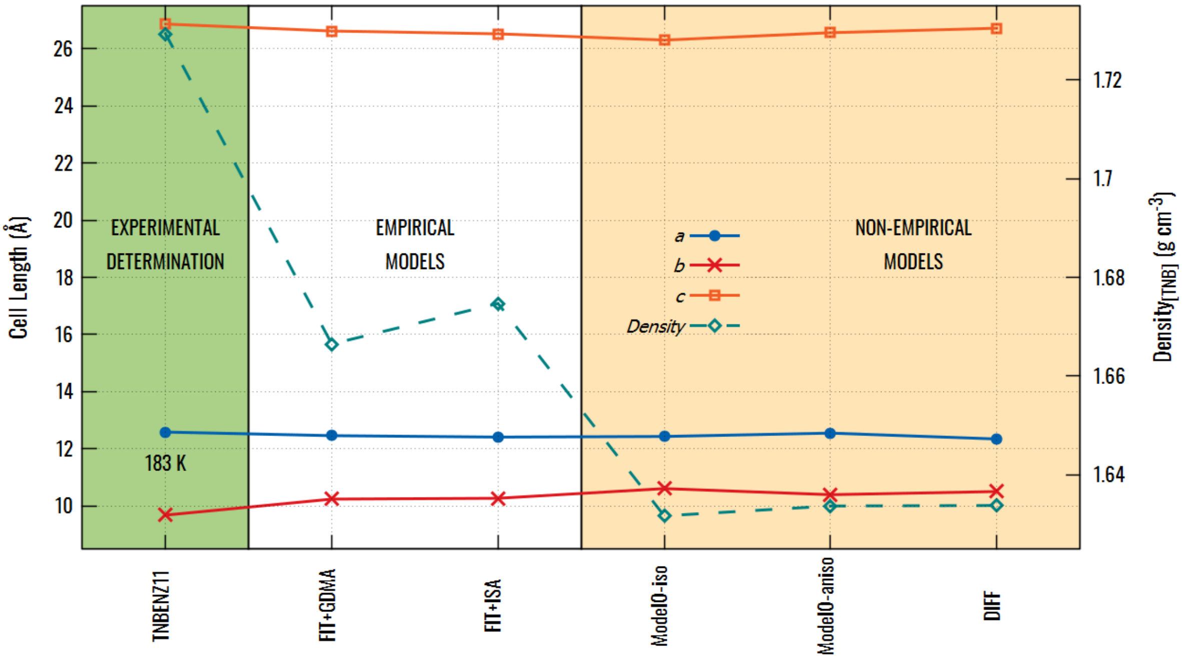
The previous study of the polar pyridine[167] (**Chapter 3**) found that the forces due to polarization had a negligible effect on the structure. One would be right to believe the polarization forces may be significant in a larger, multifaceted molecule like TNB. Including polarization using the PCM increases the predicted density (***Table 27***). The RMSD overlays of the *xminopt* structures with the experimental structure of form III, suggest that including polarization in the electrostatic model has some effect on the resultant structure and results in a poorer overlay with increasing ε_r, though not by a significant margin. The PCM does not seem to stabilise the lattice energy as much as an explicit polarization model (***Table 27***). The most significant changes in lattice energy and structure occur when ε_r is increased from $1 \rightarrow 3$. The gradient on the molecule's electrostatic potential surface (ESP) rises with increasing ε_r (***Figure 46***) and in turn further stabilises the electrostatic model. Increasing the dielectric constant beyond $\varepsilon_r = 3$ has less of an effect on the lattice energy. One can see that increasing the dielectric constant $\varepsilon_r = 11$ (which is well beyond the assumed dielectric constant of TNB, $\varepsilon_r = 2.72$) does not match energy stability gained from having an explicit polarizability model. This is likely due to the electrostatic potential gradient across the molecule, which destabilizes stacked configurations within the crystal lattice, similar to the gas-phase. Therefore, using a PCM in CSP may also result in very different relative energies and rankings that favour specific geometries, where nodes on the opposing electrostatic potential surface are close together. Accordingly, the PCM electrostatic models are used alongside the DIFF model for CSP in ***Figure 52***.

6.3.2 Reproduction of known crystal structures

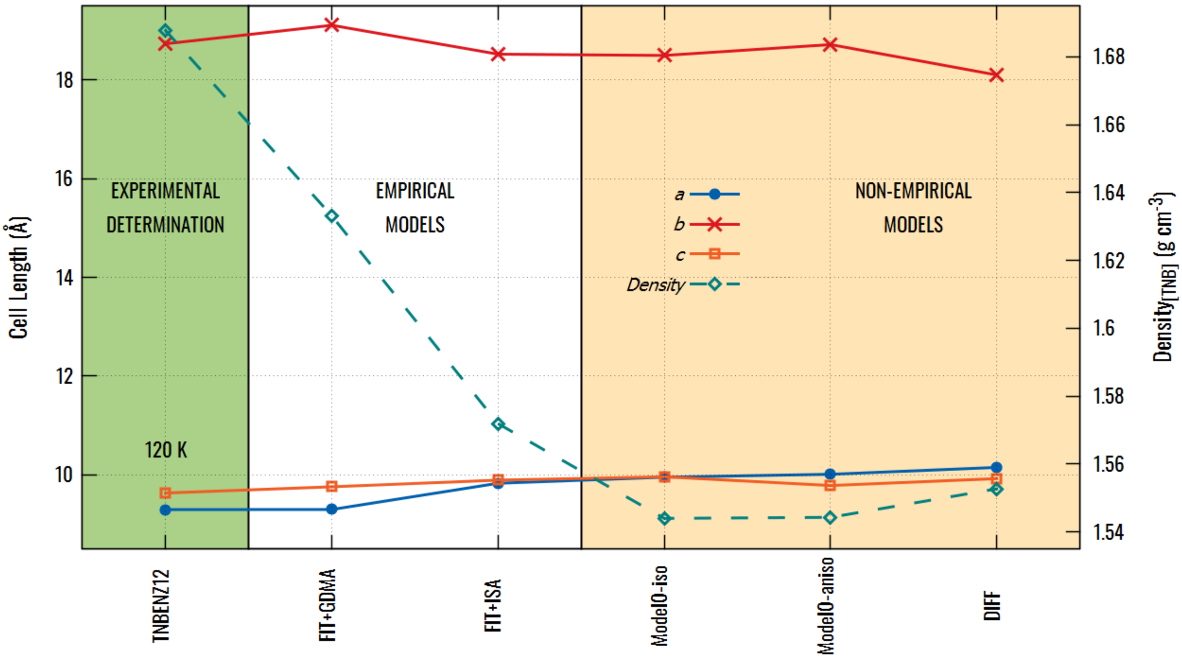
6.3.2.1 Rigid conformation

Gratifyingly, the DIFF model was found to result in a considerable improvement in the overlay $RMSD_{30}$ of the minimized form III structure and the experimental structure compared to the 1st order models (***Table 27***). In ***Figure 48***, the lattice density of form III seems to vary drastically with the force-field and even electrostatic model (FIT+GDMA vs FIT+ISA). For form III, the non-empirical models predict a denser crystal structure than the empirical models that have absorbed the effects of temperature and zero-point vibration, as was the case with pyridine (**3.3.2**). The behaviour of each force-field varies between the polymorphs (***Figure 48***), consistent with each form sampling very different intermolecular interactions. Interestingly, the experimental density for all forms is larger than all the *xminopt* predicted densities (***Table 28***). The experimental crystal structures were determined at ambient pressure and low temperatures,[263] which should have decreased the density, thus this difference is probably because the geometries were modelled using a rigid planar molecular structure. The flexibility of the nitro-groups is likely to allow for denser packing. One should note that the torsion angles of the nitro-groups in the experimental structures of forms I and II of TNB show a larger distribution than those in form III (***Figure 44***).

Consequently, the rigid model approximation used to model these may not have extended as well as it has in form III. Accordingly, minimizations with the experimental conformation were carried out.



Form I (a)



Form II (b)

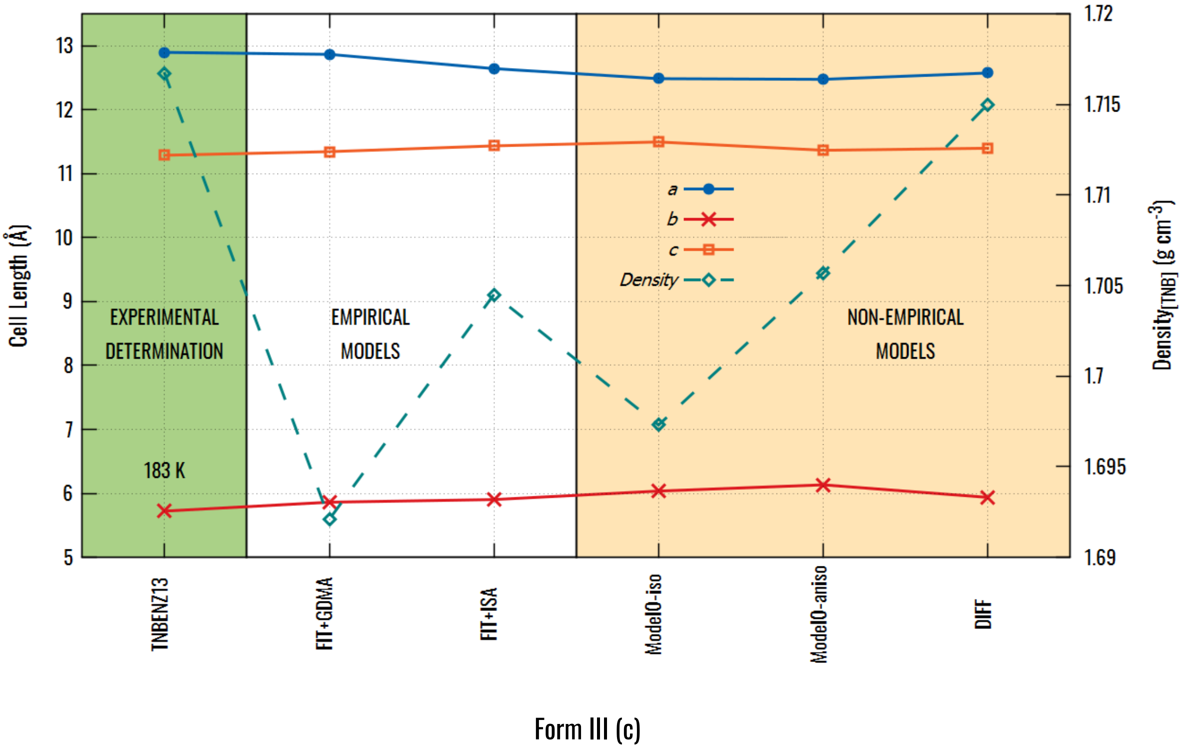


Figure 48: *The experimental crystal structures of forms I-III (a)-(c) of TNB compared with both non-empirical and empirical force-fields. All determinations were at ambient pressure, however, the temperature of each determination varies and is indicated on the graph (form I – 183 K, form II – 120 K, form III – 183 K).*

6.3.2.2 Experimental conformation

Lattice energy minimizations were conducted with the three polymorphs of trinitrobenzene (TNB) in their experimental conformation (*xminexpt*). This exercise was a brief investigation on whether the model had been hindered by approximating TNB as a rigid planar molecule. The ISA-DMA electrostatic and polarization models were recalculated in the experimental conformation to test the effects of the rigid molecule approximation on the cell geometry and lattice energy. If the experimental conformation was to be used in CSP one should recalculate also the DIFF short-range parameters and dispersion coefficients (which are isotropic) in the experimental conformation. These differences should be small as the only neglected changes in the short-range parameters and dispersion coefficients are due to changes in the charge density not associated with geometric change (**Chapter 4**). Furthermore, the experimental conformations are associated with a small conformation energy (**Chapter 4** ***Figure 27***). Hence the original DIFF with these new electrostatic and polarization models should be a very close approximation and substantially better than analytically rotating the multipole moments (**Chapter 4**).

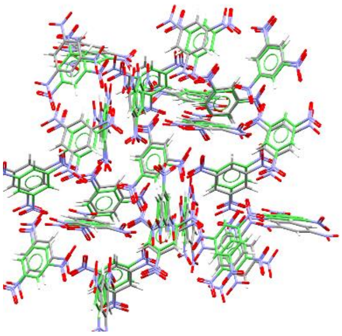
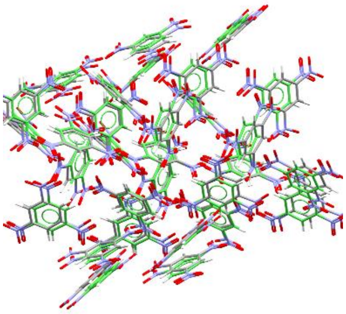
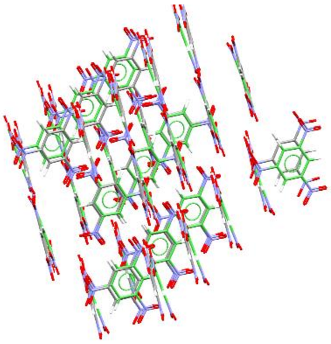
	I (TNBENZ11)			II (TNBENZ12)			III (TNBENZ13)		
	expt	xminexpt	xminopt	expt	xminexpt	xminopt	expt	xminexpt	xminopt
$a/\text{Å}$	12.5870	12.4317	12.3456	9.2970	10.1110	10.1513	12.8960	12.8835	12.5738
$b/\text{Å}$	9.6840	9.8983	10.5082	18.7300	17.7759	18.0981	5.7230	5.7253	5.9370
$c/\text{Å}$	26.8600	27.0772	26.7136	9.6330	9.9927	9.9241	11.2870	11.1944	11.3971
Density $(g\ cm^{-1})$	1.7293	1.6993	1.6338	1.6877	1.5762	1.5527	1.7167	1.7198	1.7150
E_{latt} $(kJ\ mol^{-1})$	~	-105.124	-103.375	~	-123.566	-111.221	~	-123.558	-108.847
$RMSD_{30}/\text{Å}$	~	0.15	0.64 (26/30)	~	0.61	0.57	~	0.25	0.49
Overlay of xminopt with experiment									

Table 28: *The expt (experimental structure), xminexpt (DMACRYS*[59] *minimized crystal structure, with all molecules held rigid in their experimentally observed conformations), and the xminopt (DMACRYS minimized crystal structure, with all molecules held rigid in their gas-phase optimized conformation) crystal structures and intermolecular lattice energies of the three known polymorphs of TNB. The intermolecular lattice energy (E_{latt}) estimates above were calculated using the DIFF force-field. The 30 molecule overlays ($RMSD_{30}$) have been calculated against each respective experimental determination and visualisations of the overlays of the xminopt structures with their corresponding experimental forms shown.*

There is a substantial stabilisation of the lattice energy when the experimental conformation is used (***Table 28***). It should be noted that this lattice energy does not include the intramolecular distortion energy from the optimized to the experimental structure. The largest stabilisation is seen for form III ($14.7\ kJ\ mol^{-1}$) despite this structure containing the most coplanar NO_2 groups. The *xminexpt* structures also have denser computed structures. The *xminexpt* density for form III better matches the approximations and relationships found for the non-empirical models for pyridine (**Chapter 3**). One would expect the computed density to be higher than the experimental density at ambient conditions or at least very close as temperature, pressure, zero-point energy and many-body effects have not been included. The visualisation of the 30 molecule overlays and the reduction in RMSD show that the reproduction of the cell geometry of forms I and III also improves (***Table 28***). The largest improvement is seen for form I which has a greater range of NO_2 torsion angles than form III (***Figure 44***). Interestingly, form II also has a large range of NO_2 torsion angles yet the $RMSD_{30}$ values for the *xminexpt* and *xminopt* structures are very similar. The overall energy rankings of the polymorphs do not change. Both *xminopt* and *xminexpt* calculations predict form II to be the most stable polymorph, however, the energy difference between forms II and III becomes almost negligible. The energy difference goes from $\sim 2.5\ kJ\ mol^{-1}$ to $\sim 0.008\ kJ\ mol^{-1}$, highlighting the stabilisation of the form III once conformational flexibility is included.

6.3.3 Relative stabilities of observed and hypothetical Structures

6.3.3.1 DIFF

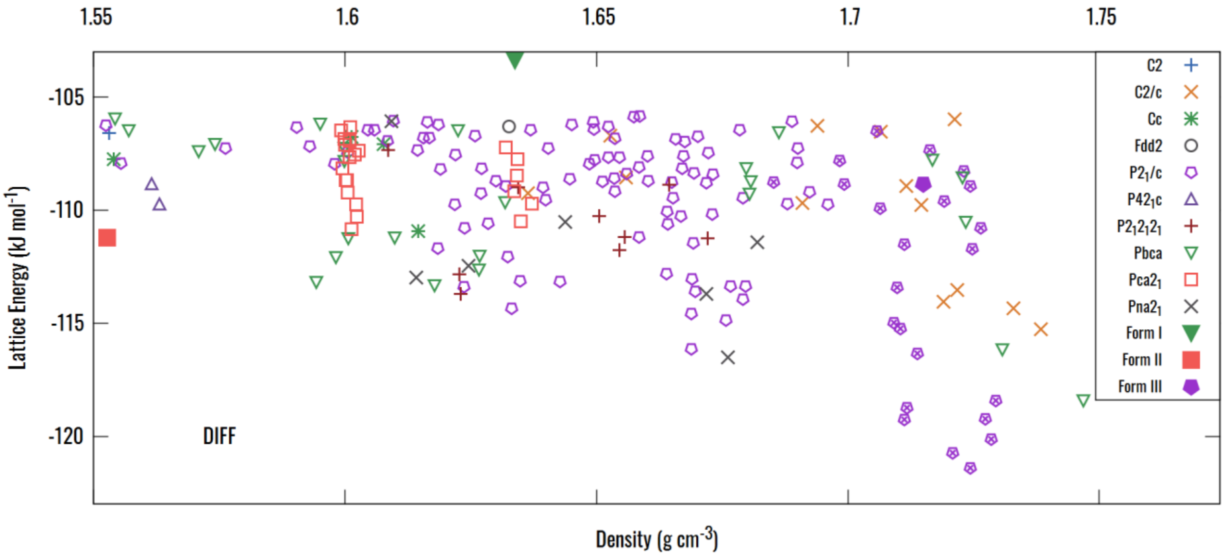


Figure 49: *Analysis of the 200 most stable CSP generated crystal structures of trinitrobenzene ranked by lattice energy against its density, calculated with DIFF. Each point represents a mechanically stable crystal structure, classified by its space group. The lattice energy minima obtained by minimizing the experimental structures with a fixed rigid, planar molecular structure (xminopt) and the DIFF are shown by the opaque symbols corresponding to their space group. Generated structures that had a 30/30 molecule overlay with form III are indicated by the pentagon and cross. This $Z' = 1$ search is not capable of generating forms I and II. The space group, lattice parameters, and lattice energy for the 30 most stable structures can be found in the* ***Appendix 6.A*** *and all structures and a complete table on the* ***Appendix CD****, linked to the structure identifier $_{TNB}\#$.*

The stability of the observed structures was compared to the hypothetical structures generated by the rigid crystal structure prediction study. The CSP study was a $Z' = 1$ search, which was not capable of finding the $Z' > 1$ structures of TNB such as forms I and II, whose *xminopt* structures have been included in the CSP lattice energy landscapes in ***Figure 49***, ***Figure 51*** & ***Figure 52*** for comparison. One can see how the landscape changes significantly with force-field. Across the different force-fields, a varying set of structures are found to be within 5 kJ mol^{-1} of the most stable, which is a likely energy difference between apparently stable polymorphs.[7] The DIFF finds only 9 structures to be in this energy range of the global minimum. It is important to note that the DIFF's lowest energy structure ($_{\text{TNB}}$9955) has a 30 molecule overlay with DIFF *xminopt* form III with a $RMSD_{30} = 0.6$ (***Figure 50***), which is not much larger than the 0.5Å differences between *xminopt* and the experimental structure (***Table 28***). Considering quantum effects and the effects of conformational flexibility, temperature and pressure have not been included, and the experimental uncertainty that is associated with the observed structures, the DIFF has successfully predicted form III as the global minimum.

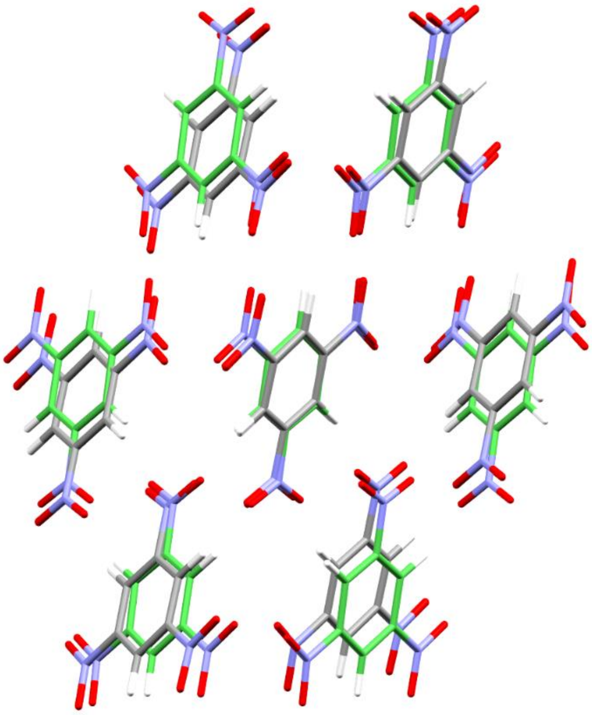
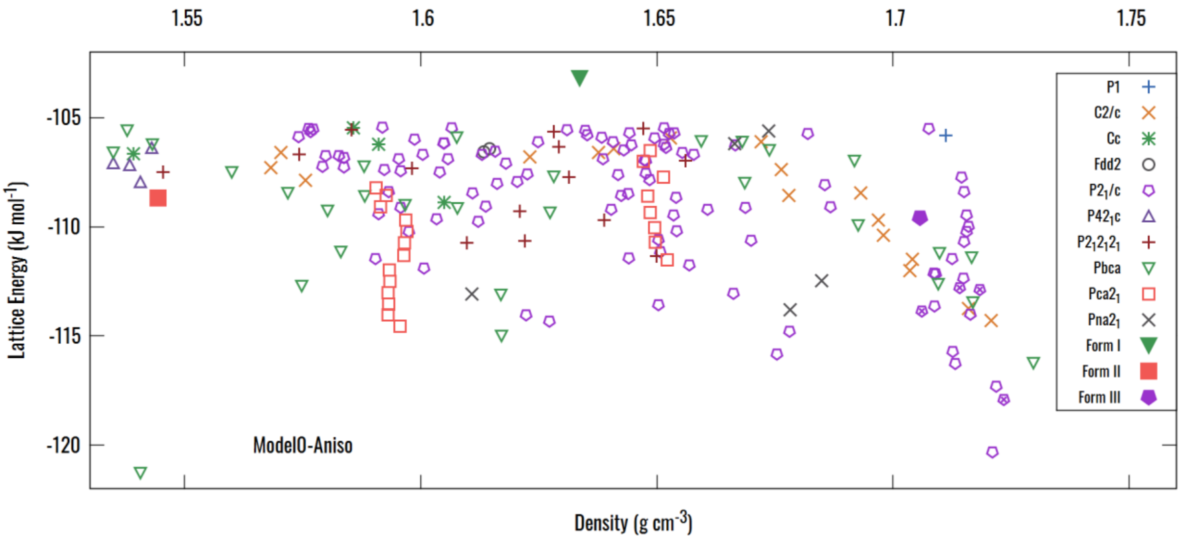


Figure 50: *The 30 molecule overlay of the DIFF global minimum ($_{TNB}$9955) and the xminopt minimized structure of trinitrobenzene, form III with an $RMSD_{30} = 0.604$. This is typical of the other CSP generated putative structures that have a 30/30 molecule overlay with form III, which if found in each potential energy landscape, are indicated by the pentagon and crosses.*

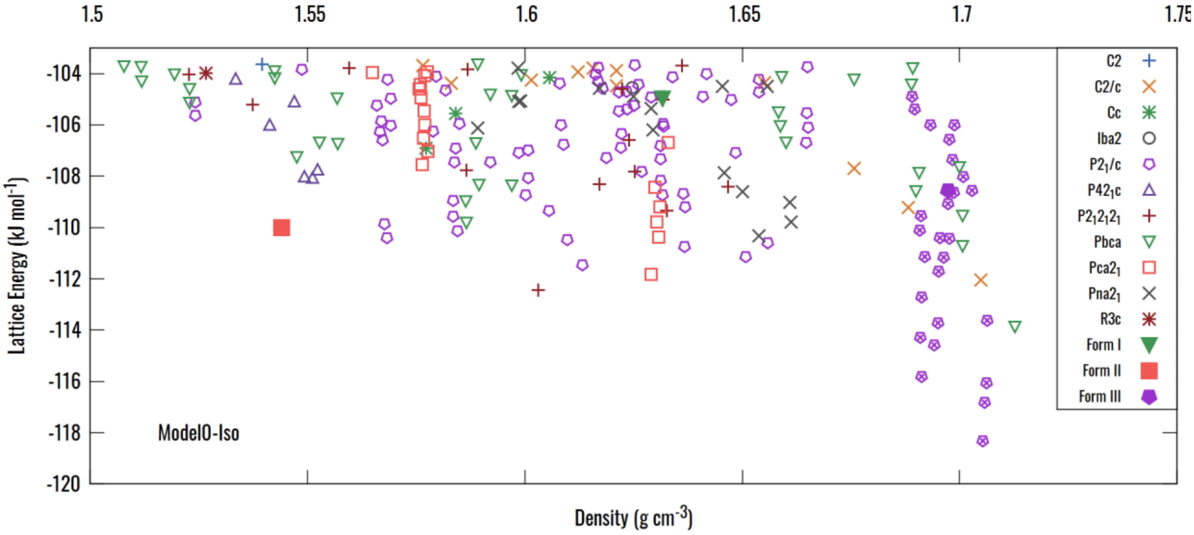
Of the top 10 structures predicted by the DIFF within a 5 $kJ\ mol^{-1}$ energy range of the global minimum, a majority of them are in the same space group as form III ($P2_1/c$) and have a 30/30 molecule overlay with form III. The remaining two are in Pbca, the more stable of which is loosely related to form I with 8/30 overlaying molecules ($_{\text{TNB}}$55), and $Pna2_1$, which could be an undiscovered form as it's $\sim 5\ kJ\ mol^{-1}$ from the global minimum.

6.3.3.2 *Model0 and FIT*

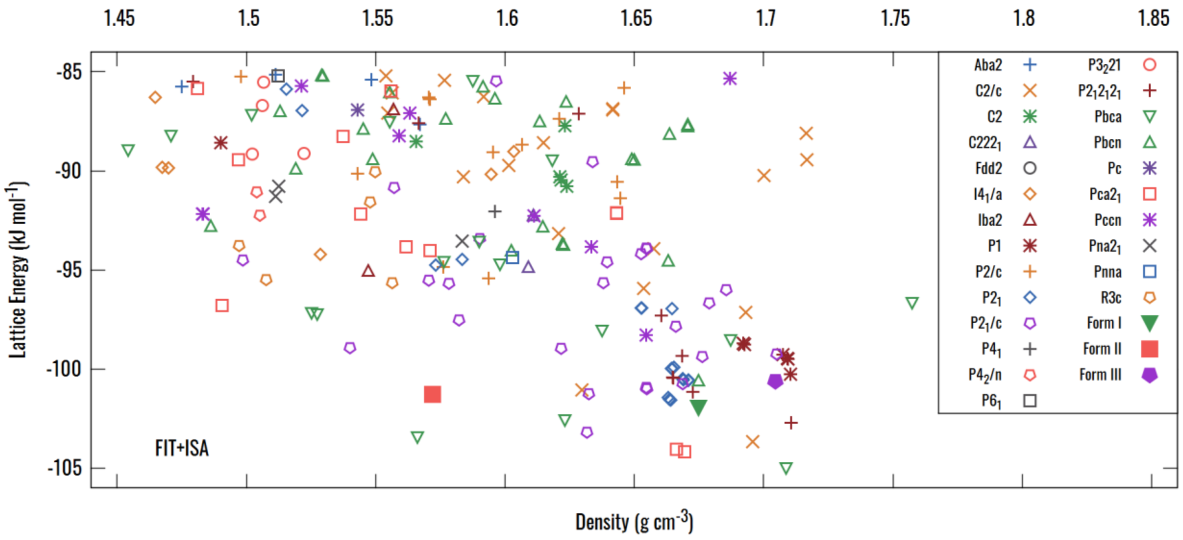
All the non-empirical models found form I to be considerably less stable than the other two forms, which given that it is the most commonly crystallized structure,[263] suggests that it is a kinetic form. Moreover, the order of stability between forms II and III switches with model, with only small relative energy differences ($\sim 1\ kJ\ mol^{-1}$). Only model FIT+GDMA and Model0-aniso found form III to be the most stable of the three polymorphs, the DIFF predicted form II to be the most stable.



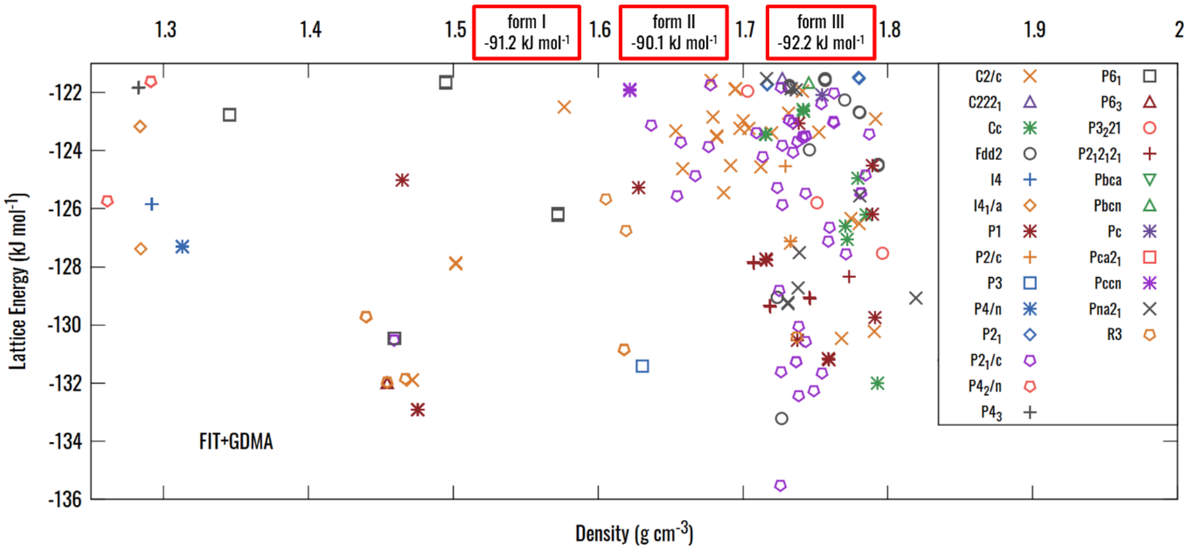
Model0-Aniso



Model0-Iso



FIT+ISA



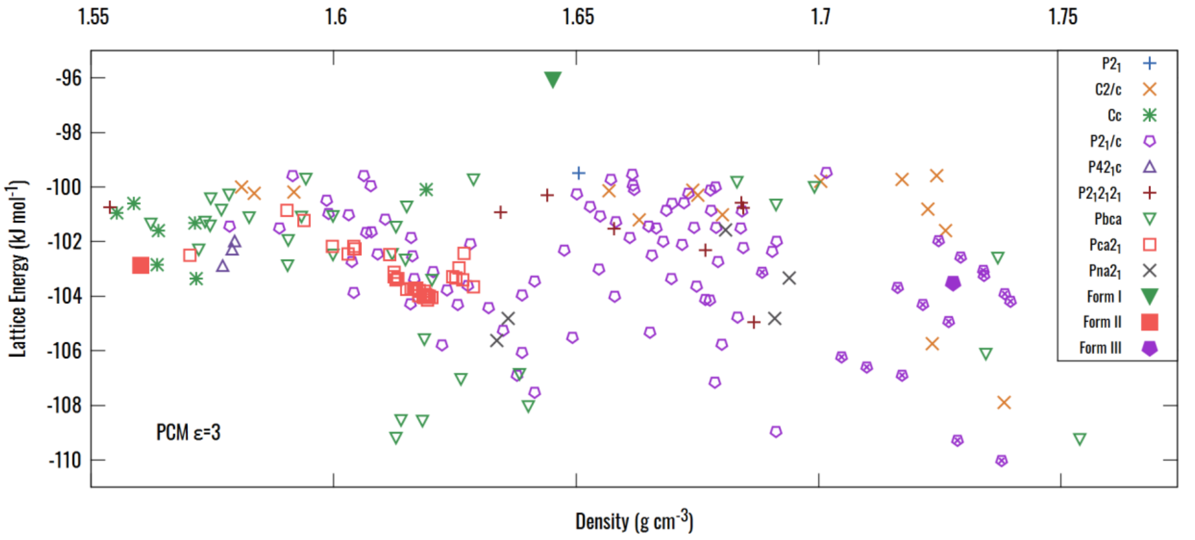
FIT+GDMA

Figure 51: *Analysis of the 200 most stable CSP generated crystal structures of TNB ranked by lattice energy against its density, calculated with the empirical FIT+GDMA (d) & FIT+ISA (c) potentials and the 1st order approximations to the DIFF, Model0-aniso (a) & Model0-iso (b). Each point represents a mechanically stable crystal structure, classified by its space group. The xminopt structures of the three TNB forms are shown by the opaque symbols corresponding to their space group. Generated structures that had a 30/30 molecule overlay with form III are indicated by the pentagon and cross. The space group, lattice parameters, and lattice energy for the most stable structures can be found in the* ***Appendix CD****, linked to the structure identifier* $_{TNB}$*#.*

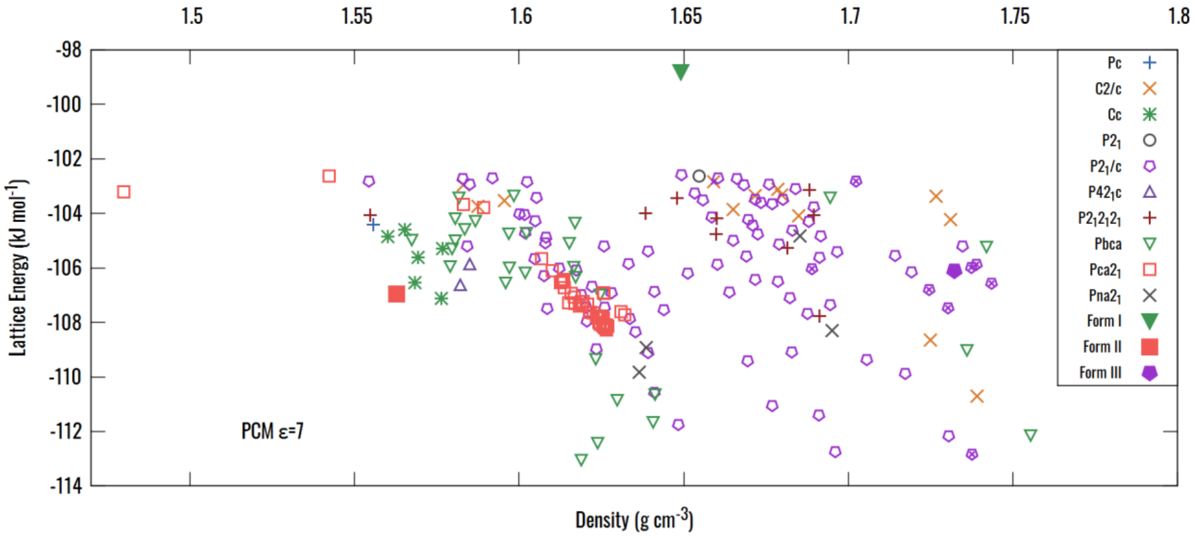
The sensitivity of the potential energy surface to the other models is shown in ***Figure 51***. The global minimum in Model0-iso also overlays with form III, with an $RMSD_{30} = 0.608$. The 1st order model with an anisotropic short-range description (Model0-aniso) found the rigid experimental structure of form III to be significantly higher than the global minimum. Model0-aniso failed to find any structures with a 30/30 overlay with form III within 5 kJ mol^{-1} of its global minimum compared to the DIFF (***Figure 50***). The closest matching structure with a 30/30 overlay was 3rd in Model0-aniso's energy ranking. The actual experimental structure was placed 51st in its energy ranking. Hence, the differences made to the CSP results from the relaxing the short-range parameters to the 2nd order SAPT(DFT) energies, and the inclusion of short-range anisotropy is substantial. If a flexible search was allowed the energy of form III would drop considerably, as found in ***Table 28*** for the DIFF, and the *xminexpt* structure may have been the global minimum for many of these models.

Figure 51 (c) & (d) show the sheer diversity of structures predicted by the empirical models, producing structures in a considerable number of space groups that, judging from the experimental structures and theoretically justified DIFF, may not be favourable for TNB. The empirical FIT+ISA model failed to find any putative structures corresponding to form III in the search but it did to predict it to be 19th in the energy rankings (***Figure 51*** (c)). The FIT+GDMA model also failed to find any hypothetical structures corresponding to form III in its top 200 structures. The best overlay was 5/30 molecules with a structure 83rd in the energy rankings. The FIT+GDMA model also predicted the *xminopt* structure of form III to be an astonishing 1600th in the energy rankings. It did not even make the top 200 structures shown in the above CSP energy landscapes (***Figure 51*** (d)). The drastic variations between these two potential energy landscapes, which use the same FIT *exp*-6 potential, are due to the differences in the intermolecular interactions described by the electrostatic model of each potential. The force-fields in ***Figure 49*** & ***Figure 51***(a)-(c) use the ISA electrostatic model (**6.2.1**), which uses a better quality wave-function and charge partitioning scheme than the GDMA electrostatic model (**Chapter 2**). This clearly makes a substantial difference alongside the repulsion-dispersion potential, as the non-empirical potentials give different low energy structures to the empirical FIT model. The nuances that define the short contacts between nearest neighbours in the TNB crystal structure are complex and intricate, a real challenge to describe accurately, and the classical CSP recipe of FIT+GDMA is most certainly unrealistic for TNB.

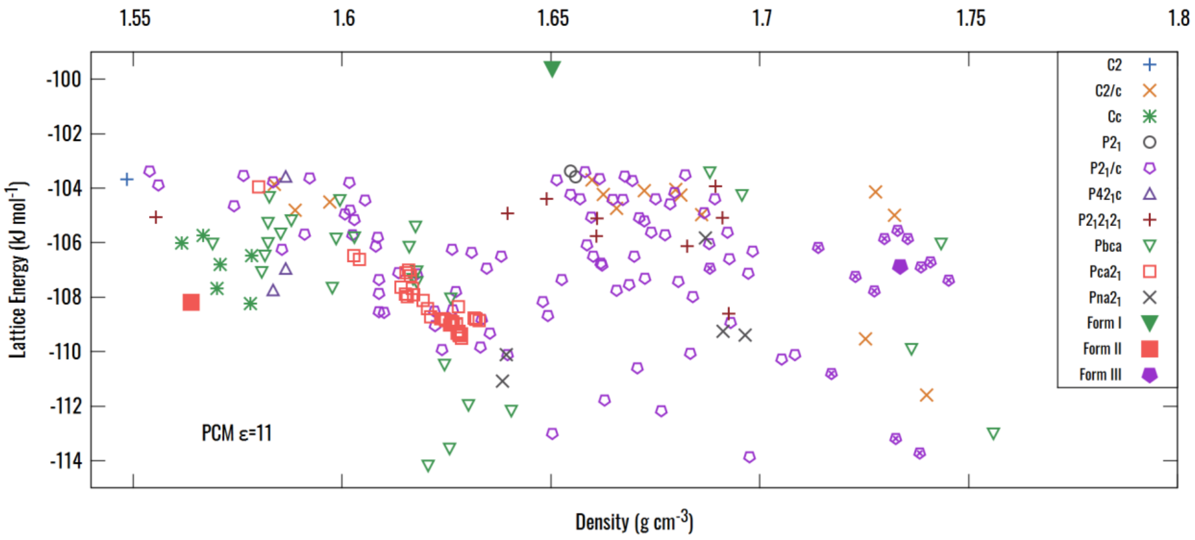
6.3.3.3 *Implicit polarization models*



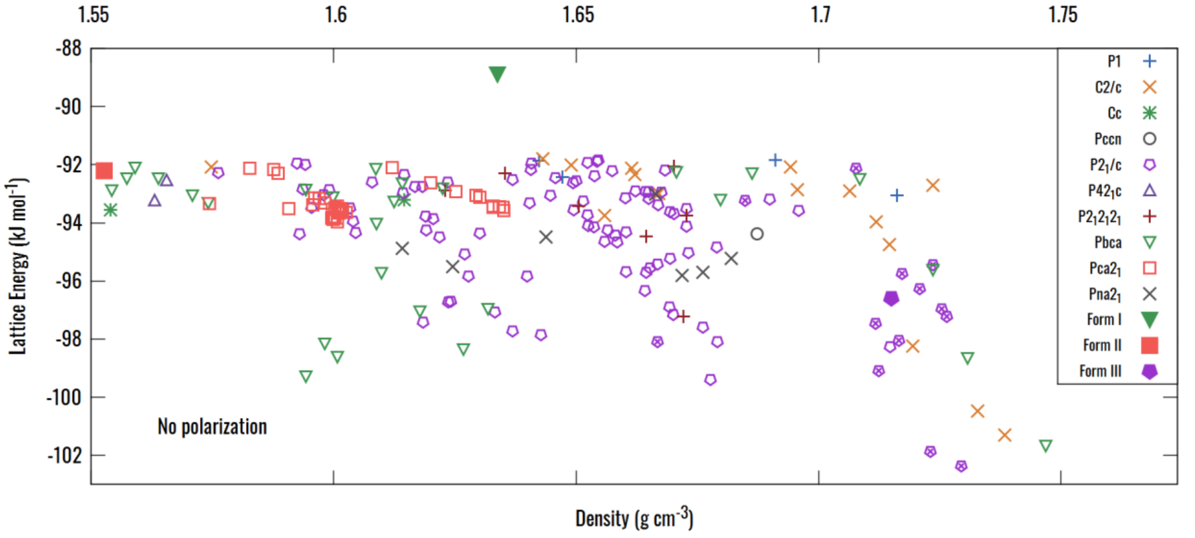
$\varepsilon_r = 3$



$\varepsilon_r = 7$



$\varepsilon_r = 11$



DIFF(no-pol)

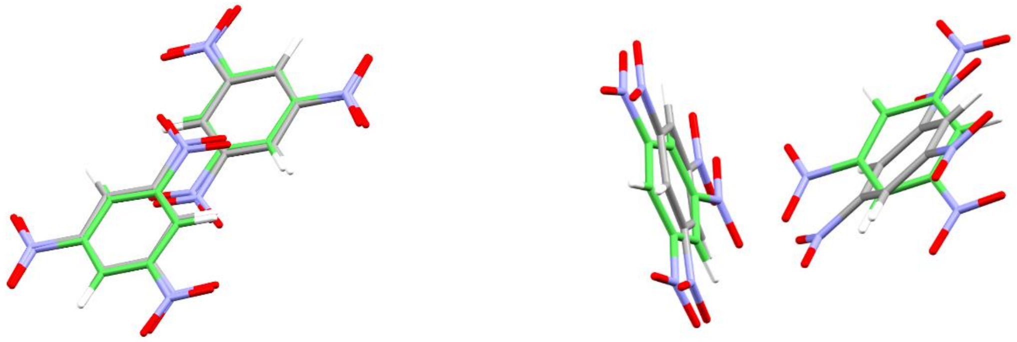
Figure 52: *Analysis of the 200 most stable CSP generated crystal structures of TNB ranked by lattice energy against its density, calculated with DIFF and different electrostatic models* ($\varepsilon_r = 3\ (a), 7\ (b), 11(c)$) *using a PCM across various dielectric constants and with the explicit polarization model removed and under no dielectric, DIFF(no-pol)* ($\varepsilon_r = 1\ (d)$)*. The xminopt structures of the three TNB forms are shown by the opaque symbols corresponding to their space group. Generated structures that had a 30/30 molecule overlay with form III are indicated by the pentagon and cross. The space group, lattice parameters and energy for the most stable structures can be found in the* ***Appendix CD****, linked to the structure identifier* $_{TNB}$*#.*

The explicit polarization model does not change the predicted cell geometries as the polarization forces are not calculated, but it does affect relative energies. ***Figure 52*** (d) shows the CSP landscape before the explicit polarization was added (DIFF(no-pol)) compared to the full DIFF in ***Figure 49***. DIFF(no-pol) predicts form III became more stable than form II (***Table 29***). The global minimum structure remained the same for both DIFF and DIFF(no-pol). However, it is noticeable that the addition of the explicit polarization energy not only significantly stabilises the lattice energy, but also increases the spread of energy values significantly ($10 - 15\ kJ\ mol^{-1}$), so that the DIFF has larger relative lattice energy differences. The explicit polarization contribution to the lattice energy is clearly very sensitive to the structure, as the spread of structures which have a 30/30 molecule overlay with form III has also changed by $5\ kJ\ mol^{-1}$.

The structural landscapes with the two most plausible dielectric constants $\varepsilon_r = 3$ and $\varepsilon_r = 1$ in ***Figure 52*** (a) & (d) suggest that the CSP results may not be significantly affected by neglecting the explicit polarization contribution when minimizing structures, if suitable approximations to the polarization energy are made. The PCM was found to not significantly affect the structures of the predicted lattice energy minima for the *xminopt* structure of form III (***Table 27***), even as $\varepsilon_r \to 11$. However, optimizing the CrystalPredictor generated structures (corresponding to the FIT with ISA atomic charge PES) with different PCM electrostatic models does make a notable difference to the set of low energy structures (***Figure 52***). The same global minimum structure found in the DIFF and DIFF(no-pol) was also found when $\varepsilon_r = 3$. The higher dielectric constants $\varepsilon_r = 7, 11$, give different global minimums, however, the 2nd most stable structure for the $\varepsilon_r = 7$ model and the 3rd most stable structure for the $\varepsilon_r = 11$ model are identical and have a 30/30 overlay with form III. The global minimum for both models is a Pbca structure that is substantially less dense than form III and may be an artefact of the potential, possibly due to the disfavouring of stacked configurations (**6.2.2.2**). This confirms the sensitivity of the CSP results to the modelling of the electrostatic and polarization model.

6.3.4 The most distinct low energy structures on the DIFF TNB lattice energy landscape

6.3.4.1 Prominent interactions and motifs



S4

T1

Figure 53: *The S4 and T1 gas-phase dimers (green) calculated with the DIFF compared with the two corresponding common two molecule contacts in the crystal structures calculated with the DIFF (grey), the S4 overlay is for the two-molecule motif found in the DIFF global minimum $_{TNB}$9955 and almost identical to the gas-phase dimer, compared to the skewed version of the T-shaped dimer found in $_{TNB}$5307.*

For both pyridine and TNB, the DIFF has transferred well to the crystalline state, predicting the observed most stable polymorphs as the lowest energy structures and other polymorphs within a competitive energy range. It is also important to decompose the lattice energies to determine, which contributions have been most significant. ***Table 29*** displays the most distinct structures (including the global minimum) found in the 200 most stable minima produced by the DIFF. The differences in space group and each contribution to the lattice energy is given, alongside a comparison with the *xminopt* minimized structures for the polymorphs of TNB (***Table 29***).

One can see that though the space groups differ significantly the percentage contributions of each term to the lattice energy are very similar. The only exception is for $_{TNB}$440, which is a relatively unstable, low density structure in a high-symmetry space group (Fdd2). Like pyridine (**Chapter 3**), the most dominant term is the dispersion energy, this is reflected in the most prominent gas-phase dimer motif, S4. Additionally, the electrostatic and polarization contributions are more prominent in the more electron rich TNB than pyridine. The polarization energy is approximately $11 - 18\%$ of the lattice energy.

The intermolecular interactions that dominate the most stable and distinct structures from the search are very different to those that dominate the gas-phase dimer structures (***Figure 36***). The most stable gas-phase minima S1, S2 and S3 are not even found in any of the 200 most stable structures in the DIFF and only a very select few crystal structures have the T1 motif. The most reoccurring dimer motifs in the crystal structures was actually the least stable stacked gas-phase dimer configuration S4 and a skewed version of the T-shaped T1 dimer (***Figure 53*** & ***Table 29***). The recurring presence of the S4 dimer, does suggest that even among different space groups many of the structures prefer particular interactions that favour certain contributions. The S4 dimer involves a mixture of strong electrostatic interactions between the poles on the electron rich nitro groups of TNB, and dispersive interactions between overlaying stacked units. This orientation also allows the monomers to slide across each other in the crystal lattice.[263]

Form III shares the S4 motif but the two other polymorphs of TNB, forms I and II, show none of the gas-phase motifs. This suggests that the interactions within the crystal are dominated by long-range forces like dispersion and polarization forces and crystal packing. In a flexible search, the rotation of the nitro-groups may allow for a denser, more stable packing as a variant on all the CSP structures generated with the rigid planar molecule, and maybe even more diverse structures would be found.

TNB#	I	II	III	9	55	119	195	1612	440	5307	9955 $(GM, {\sim}III)$	10109
Space group	Pbca	$Pca2_1$	$P2_1/c$	$Pna2_1$	Pbca	C2/c	Cc	$P2_12_12_1$	Fdd2	$P\bar{4}2_1c$	$P2_1/c$	$Pca2_1$
Motifs	~	~	S4	S4	S4	S4	~	~	S4	T1	S4	S4
$V_{elst}^{(1)}[DM]$	-35	-42	-37	-40	-41	-42	-40	-42	-26	-47	-42	-45
$V_{disp}^{(2)}[DM]$	-111	-105	-125	-124	-133	-132	-117	-114	-118	-110	-127	-114
V_{SR}	58	55	66	68	73	73	64	63	56	63	68	66
$V_{POL}^{(2-\infty)}[DM]$	-14	-19	-12	-21	-17	-14	-18	-20	-18	-17	-19	-17
E_{latt}	−103	−111	−109	−117	−118	−115	−111	−114	−106	−110	−121	−111

Table 29: *Solid-state energy contributions (in $kJ\,mol^{-1}$) for DIFF in the observed and the most diverse hypothetical structures. The lattice E_{latt}, electrostatic $V_{elst}^{(1)}[DM]$, dispersive $V_{disp}^{(2)}[DM]$, short-range V_{SR} and polarization $V_{POL}^{(2-\infty)}[DM]$ energies are given rounded up to their integer values.*

Visualising the crystal structures and examining the diversity of space-groups (**Appendix CD**) emphasises that this CSP study samples an enormous range of interactions and orientations. Many different orientations can stabilise different contributions to the energy, and the important contacts in the crystal structure differ significantly from the gas-phase.

6.3.4.2 The effects of model on the most unique structures

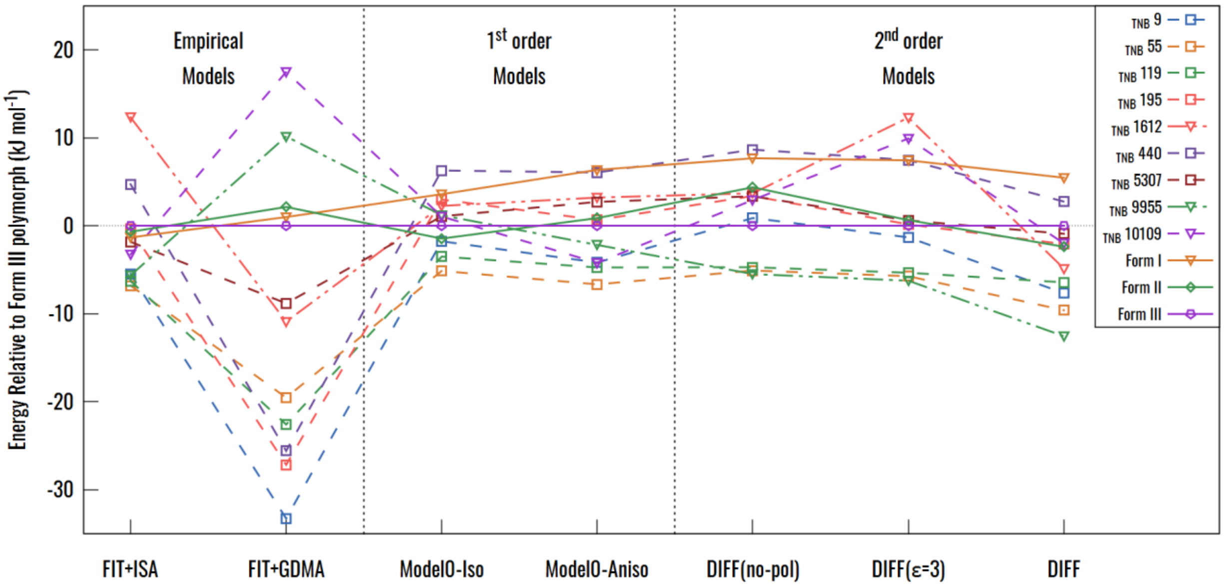


Figure 54: *Relative lattice energies of the observed and selected computer generated crystal structures of TNB, relative to form III as a function of force-field. The labels for the CSP generated structures* $_{TNB}$*# refer to their ranking after the CrystalPredictor*[35, 36] *stage, i.e. with the FIT+POINT model and further details on their structures are given in the* ***Appendix CD.***

The relative stabilities of the most stable, distinct structures alongside the polymorphs of TNB vary massively as a function of the force-field. Interestingly, the force-field with the most similar relative stabilities to the DIFF is the isotropic 1st order model (Model0-iso). The relative stabilities of the three polymorphs of TNB are the same in the DIFF and Model0-iso. Additionally, the global minimum in Model0-iso is closely related to that of the DIFF. All variations of the 2nd order models predict the DIFF global minimum $_{TNB}$9955 to be the most stable of the distinct structures. The other models predict $_{TNB}$9955 and the form III of TNB to be higher in energy. The range of relative energies is small for the 2nd order models and is even tighter for the 1st order models. The energy range increases dramatically when the wave-function quality and multipole moment partitioning method is changed in FIT+GDMA. The order of stabilities is also very different for the FIT+ISA model, however, the energy range for the most unique structures is not a huge as it is with FIT+GDMA (***Figure 54***).

6.4 Discussion

The optimized structure of TNB is planar, symmetrical, contains no strong hydrogen bonding, and yet contains polar and polarizable nitro-groups, in contrast to pyridine or many drug molecules with hydrocarbon chains, and hydrogen bond networks. Furthermore, the conformational differences between the molecules in the observed polymorphs and the optimized isolated molecular structure is greater for TNB (***Figure 44***) than in pyridine (***Figure***

14), which makes modelling the intermolecular interactions within the crystal using a rigid molecule model a greater approximation. Since TNB is an energetic molecule, there is a greater need for a force-field developed from firm theoretical footing for modelling the condensed phases, but also there is a lack of experimental data for validation. There is one polymorph, form I, that readily crystallizes, and two that have been found in attempted co-crystallization experiments, which gave only melting points and single-crystal crystal structure determinations of forms II & III.[263] Given the sensitivity of the lattice energy to NO_2 group torsion angles (**6.3.2.2**), the *xminexpt* minimizations are probably the more accurate. The *xminexpt* calculations gave the lattice energy difference between forms II and III to be very small ($\sim 0.008\ kJ\ mol^{-1}$) in (***Figure 48***), and form I to be very metastable. This is a better estimate of the relative lattice energies than that obtained using the Dreiding force-field,[263] but it would be interesting to compare with the best available electronic structure methods.[110, 130]

6.4.1 Crystal structures

6.4.1.1 CSP

Overall, the DIFF performed surprising well for the medium-sized energetic molecule TNB in these CSP studies. Even with the rigid molecule approximation, the global minimum was found to overlay with form III, the only observed $Z' = 1$ polymorph of TNB. The range of crystal structures generated in the CSP that have a 30 molecule overlay with form III and their energy range is also surprising. In many CSP studies, the generation of such similar structures would suggest the possibility of static or dynamic disorder in the crystal structures.[404-407] However, these additional structures may be an artefact of the rigid planar TNB molecule generating many lattice energy minima that would merge into one with conformational change. This seems possible if one compares the *xminexpt* lattice energies of form III (***Table 28***) with the lattice energies of the rigid hypothetical structures generated in the DIFF CSP study (***Figure 49***). The lattice energy of the DIFF global minimum, $_{\text{TNB}}$9955, is substantially lower than the *xminopt* form III structure, however, the *xminexpt* was calculated to be $-124\ kJ\ mol^{-1}$ (***Table 28***) against the global minimum's $-121\ kJ\ mol^{-1}$. The range of rigid structures that overlay with form III (pentagons with crosses in ***Figure 49***) with lattice energies between the *xminopt* and *xminexpt* structures suggest that there are transformation energy barriers between very similar configurations. These CSP searches are limited by the rigid molecule approximation, and show that the rigid molecule potential energy surface around the global minimum is bumpy.

The top 10 structures in the DIFF imply that there could be some unobserved polymorphs of TNB in the Pbca ($_{\text{TNB}}$55) and $Pna2_1$ ($_{\text{TNB}}$9) space groups (***Table 29***, ***Figure 49*** & ***Figure 54***). Though the overlay is loose with 8/30 molecules in common with form I, it is possible that $_{\text{TNB}}$55 may be found if the fast crystallization of form I with some sub-optimal packing could be suppressed. $_{\text{TNB}}$55 has a hypothetical density of $\sim 1.75\ g\ cm^{-1}$, which suggests it could be a high pressure form that has yet to be observed experimentally due to the obvious challenges of high pressure energetic studies. $_{\mathrm{TNB}}$55 also has 16 out of 30 molecules in common with form III, further implying that it could be the product of a high pressure transformation similar to form II → form III in the pyridine study (**3.3.4**), if a sufficient quantity of TNB form III could be generated. The less dense $_{\mathrm{TNB}}$9 has no molecules in common with form III and only 4 in common with form II, thus it may be an unobserved form that could be found via another crystallization route. These structures are sufficiently different to suggest that it would be worth confirming their relative stability (c.f. ***Figure 54***) before embarking on more experimental work on the solid state crystallization of TNB. The problems and safety issues of screening for new polymorphs of explosives may prevent this from happening, particularly, for high pressure experiments.

6.4.1.2 Cell geometry

The cell parameters of all forms are in reasonable agreement with the experimental structures, suggesting that contacts along the 3 axes are well-sampled and described by the non-empirical force-fields. The most dramatic changes are found to be in the lattice densities. This is likely due to the change in crystal packing that occurs when the monomers are held rigid. The range of torsion angles seen in the nitro-groups of form III are significantly smaller than that of forms I & II (***Figure 44***), which is why *xminopt* calculations with the non-empirical models perform better for form III than for the other two forms. There is a stark difference in the predicted and experimental densities when the molecule is held rigid and planar (*xminopt*) compared to the observed conformation (*xminexpt*). The predicted density of forms I and II increases notably in the *xminexpt* minimizations (***Table 28***), stressing the effects of conformation on crystal packing. The improvement of the overlays of forms I and III with the experimental determinations (***Table 28***) also reinforces the importance of conformation in the intermolecular interactions and the resultant minima. Considering the possible changes in cell geometry, due to experimental uncertainty, the inclusion of zero-point energy,[294] and temperature and pressure effects, the DIFF potential performs very well for form III (***Figure 48*** (c)). The empirical models do not model form III's lattice structure as well. The absorption and averaging over errors, approximations and other effects results in an unreliable prediction. In addition, changing the wave-function and the method of partitioning (GDMA vs ISA) changes the lattice density dramatically (***Figure 48*** (c)). Forms I and II could be very good tests of the ability of future flexible non-empirical models. However, it would be helpful to have more detailed experimental studies on the polymorphs of TNB, looking carefully for the presence of disorder and variability with sample crystal, as a function of temperature.

6.4.1.3 Polarization

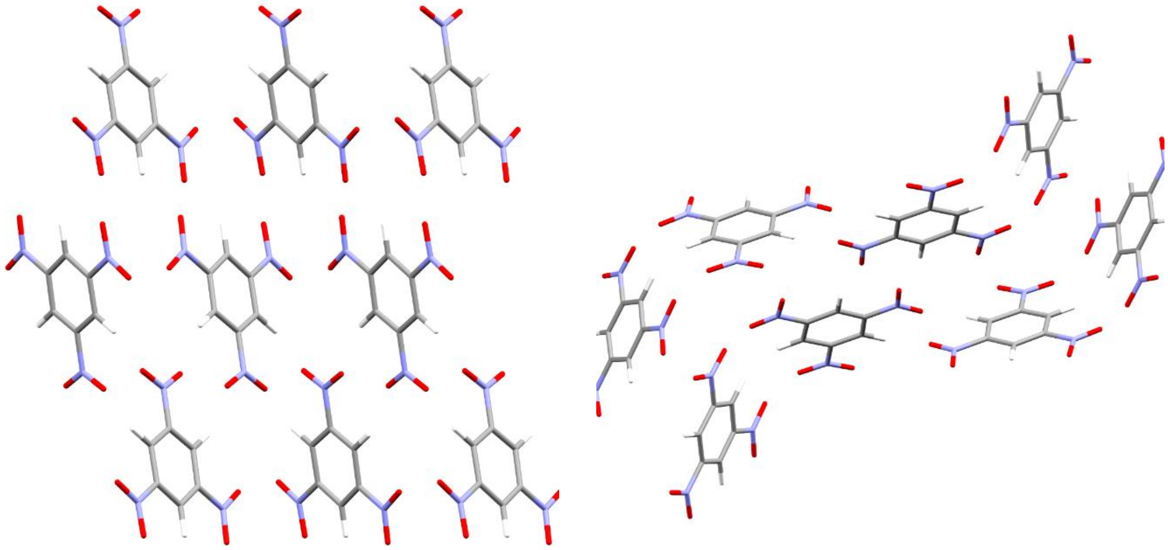
The non-additive nature of polarization is particularly challenging for modelling the crystalline phase because of the need to iterate to consistency (**2.6.1.3**). **Chapter 5** showed that the polarization energy is significant in dimers of TNB, and ***Table 29*** shows that it also makes a significant contribution to the lattice energy. This is because the induced dipoles are determined by the field around the atoms, and the local symmetry around atoms in these organic crystals is rarely much higher than in the dimer. This study has pioneered the use of distributed polarizability models in CSP studies, and shows that the polarization energy is very sensitive to crystal structure. The inclusion of this explicit polarization model has increased the relative lattice energies of fairly similar structures, such as those with a 30/30 molecule overlay with form III, and this may have been over-estimated (**6.4.4.1**).

The investigation of the use of the polarizable continuum model, as an approximation to modelling polarization during lattice energy minimization, appears successful, though it does not model the polarization effects in the dimer well (**6.2.2.2**). Including polarization via the PCM does increase attractive interactions and the computed lattice density but it does not improve on the predicted structure nor does it change the cell geometry significantly, as one can see from the RMSD_{30} values of the minimized form III (using different ε) overlaid with the observed structure in ***Table 27***. The potential energy surface does not change significantly either (***Figure 52***), however, it has to be stressed that TNB is an unusual medium-sized molecule in having high symmetry and few functional groups. Modelling with the PCM was introduced as a way to improve the relative energies of different hydrogen-bonding motifs in peptide crystal structures,[260] which are very different systems to TNB. Using a dielectric close to the calculated dielectric constant of TNB (2.72) and the proposed dielectric constant of organic solids[260, 261] ($\varepsilon_r = 3$) gives an energy landscape (***Figure 52***) similar to the DIFF with an explicit polarization model (***Figure 49***), and found a hypothetical structure that has a 30/30 molecule overlay with form III as the global minimum (***Figure 52***). Increasing the dielectric does not improve the model but instead generates artefact structures, as seen with the substantially less dense global minimums of the $\varepsilon_r = 7$ and 11 models (***Figure 52***), which may be to the maltreatment of certain two molecule contacts (**6.2.2.2**). Accordingly, it seems the proposed dielectric constant for approximating polarization effects in organic crystals over 10 years ago[260] is still a valuable model.

6.4.2 Lattice energies and intermolecular interactions

The lattice energy landscapes created by each potential vary noticeably, especially between empirical and non-empirical models. ***Figure 54*** highlights how dramatically the landscape changes, the relative energy rankings change between the non-empirical models, but the relative energy ranges are similar. The relative energy range and rankings change most notably for the empirical models. Especially for FIT+GDMA, which also uses a different wave-function to construct the electrostatic model. The energy range of the set of distinct crystal structures is an astounding $50\,kJ\,mol^{-1}$ with the FIT+GDMA model. The non-empirical models give a more reasonable qualitative crystal energy landscape than the empirical models (***Figure 49***, ***Figure 51*** & ***Figure 52***). Anisotropy in the short-range portion of the potential appears to stabilise the overall model, as the total lattice energy falls for the explicit Model0-aniso compared to Model0-iso. This may be because the shape (anisotropy) appears to also improve the description of the close-contacts in the crystal lowering the crystal density (Model0-iso vs Model0-aniso, ***Table 27***), and so increasing the stabilizing dispersion. It does seem that 1st order SAPT(DFT) dimer energies are not sufficient enough to provide an accurate enough potential energy surface for the anisotropic TNB, as the crystal structure overlays of form III are worse with the Model0s than the DIFF. Including the 2nd order dimer interaction energies in the DIFF seems to improve on this and lower the energy of the hypothetical structures (***Figure 49***).

It is clear the non-empirical models favour very different intermolecular geometries compared to the empirical models. The empirical models generate structures in a significantly larger number of space-groups (***Figure 51*** & ***Figure 55*** (b)). The range of space-groups and the increased presence of groups like Fdd2 and R3 highlights the benefit of minimizing the structures straight from their CrystalPredictor generation (**6.2.4**) and sampling more of the differences in the potential energy surfaces. When the lattice energy minima from a force-field with a good distributed multipole electrostatic model are used as starting points (**3.2.2.4**), many structures do not move far from their minima, once re-minimized with another potential. The FIT+ISA model in particular seems to favour the Pbca crystal configuration as its global minimum is the said space group and the FIT+ISA model is the only model that predicts form I, which is also Pbca, to be the most stable polymorph (***Figure 54***). The lattice energy landscape generated by FIT+ISA is quite reasonable, in that the three known polymorphs are close in energy, and not significantly less stable than the alternative structures generated.



(a) (b)

Figure 55: *The crystal motifs found in the $P2_1/c$ (a) and Pbca (b) space-groups, viewed along the b-axis.*

6.4.3 Flexibility

The relatively poor reproduction of the crystal structures by the rigid planar molecules in the *xminopt* calculations (***Table 28***) highlights the importance of conformation in nitrobenzene CSP studies. The stabilisation of the computed lattice energies of all three forms of TNB (***Table 28***) suggests crystal packing also plays a large role in the intermolecular energy. Keeping TNB in a rigid planar conformation seems to be the main limitation of the CSP results. It would be very difficult and expensive to include flexibility in this CSP study as the range of available conformations for the nitro-groups is large and the electrostatic model (and thus the polarization and dispersion models) changes with conformation (**Chapter 4** ***Figure 25***). The ISA-DMA multipole moments for the 5 different molecular conformations are very different (**see Appendix CD**) and have resulted in very different lattice energies (***Table 28***) even though the range of observed torsion angles is not enormous (***Figure 44***). Pressure would certainly change the NO_2 torsion angles and thus in order to conduct a pressure study like that done on pyridine in **Chapter 3,** a flexible search would be required. Analytically rotating the multipole moments may reproduce the crystal structures but does not replicate the relative energies (***Table 15***), thus using codes like DMAFlex[408] may not be suitable for a trustworthy flexible CSP study of TNB.

A future work could be to conduct a high quality flexible CSP study on TNB to see the effect of allowing the NO_2 groups to move would have on CSP predictions. This could easily be done with CrystalOptimizer[39] and the empirical FIT+GDMA potential to see if there was any improvement in prediction. However, it is also clear to see that using an empirical model with a cheaper inaccurate electrostatic model (FIT+GDMA) is not appropriate for anisotropic molecules like TNB. The FIT+GDMA model performed decently for pyridine in **Chapter 3,** but does not for the larger TNB. The oversimplification of functional form and the electrostatic energy fails to capture the shape and close-contacts of TNB. Improving the quality of the wave-function used in the electrostatic model and the partitioning method (FIT+ISA) greatly improves the predicted potential energy landscape. Although the FIT+ISA model fails to find form III in the search, it calculates the *xminopt* energy to be within the top 20 structures. On the other hand, re-calculating the ISA distributed multipoles of the required high quality wave-function would be computationally expensive to do for each conformation.

6.4.4 Omitted Terms

Alongside, approximating TNB as a rigid planar molecule, many other approximations have been made in the DIFF that may have been detrimental to the force-field. **Chapter 5** describes in detail the reasoning behind the omission of some higher-order and many-body terms in the DIFF. Contrary to the pyridine investigation (**Chapter 3**), omitting the many-body terms has not resulted in a denser predicted structure. This may be due to uncertainties in the experimental determinations of TNB, all three forms were from the same study, but also may be due to relative contribution of these terms to the intermolecular energy. The many-body dispersion should in theory be more considerable in the larger TNB but may not be as important as previously thought.[38, 167, 200, 409] The higher-order induction terms and charge delocalization contributions are likely to be small in the condensed phase of TNB, which is less polar than pyridine and substantially less polar than water, where these terms were found to be most important.[38, 46, 410] Conversely, form II is in a polar space group, and for a finite-sized crystal, there may be a destabilising energy term that depends on the crystal size and morphology.[411] The DIFF computed net dipole moment per unit cell of form II is $0.0026\ e\text{Å}$ and $0.09640\ e\text{Å}$ for the *xminopt* and *xminexpt* calculations respectively.

6.4.4.1 Charge delocalization and higher order induction terms

The effects of including the polarization implicitly via the PCM on the reproduction of the experimental configurations seemed negligible (***Table 27***), but there was a significant influence on the lattice energies (***Table 27***, ***Figure 52*** & ***Figure 54***). The implicit PCM potentials generated less polarization than the explicit distributed polarizability models. In this thesis, only dipolar polarizabilities have been used, and the single molecule damping parameter did not appear to be very accurate against the SAPT(DFT) polarization energies (**Chapter 5** ***Figure 38***). It is quite possible that the explicit polarization term implemented in all the non-empirical potentials is overestimating the effect of polarization on the relative lattice. It is possible that the inclusion of the charge delocalization and higher-order induction terms (via δ_{int}^{HF}), may have increased the stability of the lattice, lowering the energy and increasing the overall density of the structures. It was found that the non-empirical potentials underestimated the experimental densities (***Figure 48***) by a significant margin for forms I and II (even if one was to account for possible experimental uncertainties). δ_{int}^{HF} resulted in over-binding for the pyridine structures in **Chapter 3** because there was a strong presence of the most stable stacked dimer in the lowest energy crystal structures. However, in TNB one can see that the two molecule contacts in the lattice greatly differ from those found in the gas-phase (***Table 29*** & **5.5.3.2**), and it seems to not have a preference for stacked configurations. The only stacked gas-phase configuration present is the S4 dimer. Consequently, it is possible that the inclusion of δ_{int}^{HF} may not have been detrimental to the potential but it certainly would have been expensive. Further research into modelling the polarization of larger organic molecules in their low symmetry crystal structures is needed.

6.4.4.2 Non-additive many-molecule dispersion

Table 29 emphasises the importance of the dispersion contribution to the total intermolecular energy. The dispersion coefficients were derived from the molecular charge distribution as a pairwise-additive term, however, the importance of the many-body dispersion in the TNB phases was not investigated. TNB is a more electron-rich molecule than pyridine, thus, electron correlated contributions like the 3-molecule dispersion may actually have a considerable contribution to intermolecular interaction energy. Coupled cluster (CCSD(T)) (**1.3.2**) calculations on a TNB cluster would be considerably expensive but may shed some light on the importance of this contribution in the condensed phase. For example, stacked configurations like those with the S4 motif may be relatively destabilized if the non-additive many-molecule dispersion was included, which could significantly change the energy rankings.

6.4.4.3 The effects of temperature

The empirical FIT potential has absorbed some temperature effects in its parameterisation against experimentally observed crystal structures at various conditions.[67, 68] Thus using empirical force-fields in MD could lead to the double-counting of temperature effects. The FIT+ISA study gives sensible results, in that the observed structures are reasonably close in energy, and not implausibly higher in energy than the most stable structures (unlike FIT+GDMA). Moreover, the differences between the FIT+ISA and the non-empirical potentials are significant, most notably in the relative instability of form I. It seems likely that temperature will have a significant effect on the relative stabilities, particularly given the number of CSP generated structures that are very similar to form III. It is likely these similar structures are local minima on a bumpy PES with small energy barriers that would smooth out to one minimum if NO_2 rotation due to temperature was allowed. We cannot apply the rule of thumb for rigid hydrocarbon structures that thermal effects reduce relative energy differences, as the denser, more stable structures have lower amplitude and higher frequencies.[300] The DIFF has reversed the density order of forms I and III (***Figure 48***), which is partially because of neglecting conformational flexibility (***Table 28***). The librations of the NO_2 groups is likely to be a major contribution to the relative thermal stabilization. However, a definitive comparison with experimental data for judging the accuracy of a force-field should consider the effects of temperature.

6.4.5 Future work to reduction computational cost or accuracy

The DIFF outperforms the cheaper 1st order models but is this improvement worthwhile once the computational time spent in developing the force-field comes into question? The DIFF used only 327 $E_{int}^{(2)}[SAPT(DFT)]$ dimer interaction energies to relax Model0-aniso's $\alpha_{00}^{\iota\kappa}$ and $\rho_{00}^{\iota\kappa}$ short-range parameters and fit the polarization and dispersion model damping parameters. This is a substantially smaller dataset compared to pyridine's ~1000 2nd order points, which was considered "a tiny dataset".[38] With each TNB calculation taking approximately 15 hours compared to pyridine's 2 hours (on an older workflow, which hadn't even been optimized like TNB's workflow has been) it's understandable why the dataset is that size. For large molecules one has to be cunning with selecting configurations for fitting, and cannot assume that the gas-phase dimer configurations are going to dominate the crystal structures. Ideally, all of the configuration space would be sampled, without "diluting" important configurations, but as discovered in **Chapter 5**, this requires many configurations, and hence calculating sufficient second-order energies may not be possible for molecules larger than TNB. Future work would be to see how well Model0-aniso does in CSP using the approximate dispersion damping parameter of 1.8 (**5.3.4**) and an educated guess of the polarization damping parameters to see if the 1st order models are good enough to be used on large molecules as an alternative to empirical force-fields.

The main limitation of this TNB CSP was the neglect of flexibility in the nitro-groups. Minimizing the crystal structures whilst allowing the NO_2 groups to rotate (analytically rotating the multipole moments instead of re-calculating them) may have provided better crystal structures. However, the final energy evaluation would require re-calculating the ISA electrostatic models for the final conformation of each molecule within the low energy crystal structures in order for it to be reliable. The sensitivity of the quality of the electrostatic model (GDMA vs ISA), shows that this would be very expensive for the required quality of charge density.

6.5 Conclusion

To conclude, the DIFF was successful in modelling the TNB crystal lattice and has provided some important insight into the importance of conformation, force-field functional form and polarization effects in CSP. Even within the limitations of using a rigid planar TNB model, the $Z' = 1$ CSP of the non-empirical potentials showed the favourability of the form III packing, which is the motif favoured in TNB molecular complexes. Furthermore, the DIFF generated low energy hypothetical structures with similar motifs to those seen in forms I and III. This state-of-the-art model has really pushed the envelope with respect to what was considered previously possible with anisotropic force-fields for CSP. This is the third ever non-empirical anisotropic atom-atom potential used in CSP[60, 167] and only the second to include an explicit polarization model, the first being the pyridine DIFF (**Chapter 3**). One can see that for the larger more flexible trinitrobenzene, the DIFF performs very well in CSP, compared to the standard empirical FIT+GDMA model. The rigid molecule approximation coupled with a pairwise model and a many-body polarization model reproduces the lattice structure reasonably well (***Figure 48***) and can be used in CSP to model the potential energy surface of organic energetic materials (***Figure 49***). The CSP study raises many questions about the possible polymorphs of TNB, a system that warrants further experimental investigation in order to develop the modelling of energetic materials.

Appendix 6.A – Stability of observed & top 30 DIFF hypothetical structures

The lattice energies ($kJ\ mol^{-1}$) and cell geometries of the 30 lowest energy structures and the *xminopt* minimizations of the three forms of TNB using the DIFF.

Structure ($_{\text{pyr}}$#)	Lattice	Density	Space-Group	a	b	c	α	β	γ
$9955 \approx III$	-121.3867	1.7244	P21/c	12.8586	5.6599	11.3555	90	83.3271	90
9983	-120.7215	1.7208	P21/c	16.2124	5.6524	11.3901	90	52.0029	90
4917	-120.1334	1.7285	P21/c	12.85	5.6631	11.3444	90	82.7394	90
4888	-119.2635	1.7113	P21/c	24.9121	5.6993	11.3035	90	148.9775	90
30	-119.2343	1.7273	P21/c	12.8476	5.6631	11.3496	90	82.93	90
4891	-118.7367	1.7117	P21/c	17.961	5.7003	11.2982	90	45.6336	90
55	-118.4311	1.7469	Pbca	25.8181	5.5358	11.3385	90	90	90
2515	-118.4252	1.7294	P21/c	12.8406	5.6707	11.3367	90	82.5398	90
9	-116.4963	1.6762	Pna21	10.0116	9.1053	9.2636	90	90	90
5027	-116.3286	1.7138	P21/c	12.8672	5.9579	10.9892	90	78.6387	90
54	-116.1642	1.7307	Pbca	25.918	5.5876	11.295	90	90	90
4850	-116.1383	1.6689	P21/c	14.0265	5.8616	11.3202	90	114.3208	90
119	-115.268	1.7384	C2/c	26.329	5.5214	11.3921	90	79.5187	90
31	-115.2529	1.7105	P21/c	12.8955	5.6945	11.3198	90	84.5685	90
9961	-114.97	1.7091	P21/c	12.8739	5.9415	11.0174	90	79.3385	90
12374	-114.8589	1.6758	P21/c	15.5483	6.1539	11.3261	90	128.7907	90
4918	-114.5712	1.6689	P21/c	21.1977	5.8604	11.3228	90	37.0838	90
2301	-114.3375	1.6331	P21/c	10.5322	5.8758	14.0529	90	85.2796	90
5929	-114.3299	1.7329	C2/c	26.359	5.5079	11.4736	90	78.7359	90
10680	-114.0364	1.719	C2/c	38.3638	5.5914	11.3433	90	137.404	90
13	-113.9348	1.6791	P21/c	9.6005	9.7002	9.0901	90	84.7742	90
7	-113.7005	1.6719	Pna21	10.0313	9.0781	9.297	90	90	90
1612	-113.6915	1.623	P212121	11.0628	13.2627	5.9441	90	90	90
4916	-113.5758	1.6697	P21/c	14.0435	5.8522	11.3344	90	65.5166	90
8489	-113.5221	1.7217	C2/c	26.5095	5.591	11.3223	90	101.5273	90
33	-113.4227	1.7098	P21/c	12.8748	5.9372	11.0225	90	100.723	90
23	-113.3903	1.6236	P21/c	10.2745	9.3887	9.038	90	89.7458	90
12375	-113.3723	1.6796	P21/c	15.5998	6.1578	11.3589	90	50.5658	90

25	-113.3499	1.6766	P21/c	9.6436	9.749	9.0507	90	97.1621	90
39	-113.3389	1.6178	Pbca	10.0891	9.8364	17.6324	90	90	90
II	-111.2	1.5527	Pca21	10.1513	18.0981	9.9241	90	90.00	90
III	-108.8	1.7150	P21/c	12.5738	5.9370	11.3971	90	104.0449	90
I	-103.4	1.6338	Pbca	12.5870	9.6840	26.8600	90	90.00	90

The full structures can be found in the ***Appendix CD*** *in .res format.*

Appendix 6.B - DMACRYS Input Files for the DIFF

Table of Contents

The DMACRYS input files with axis definition files, can be found in the ***Appendix CD****. Further details on the formatting of the input files can be found in the DMACRYS manual.*[59]

11. Iterated Stockholder Atoms – Distributed Multipole Moments (ISA-DMA)

The ISA-DMA multipolar moments are given DMACRYS input file format. These detail the atom labels, atom types and multipole moment ranks used for TNB's electrostatic model. The atomic positions can also be found in this input file, adjacent to the corresponding atom label.

12. Iterated Stockholder Atoms – Distributed Polarizabilities

The ISA-DMA derived $Rank = 1$ distributed polarizabilities in DMACRYS input file format calculated in the molecule-fixed axis.

13. ISA-DMA under the PCM with $\varepsilon_r = 3$

The ISA-DMA multipolar moments under the polarizable continuum model (PCM) with a dielectric constant (ε_r) of 3, given DMACRYS input file format.

14. ISA-DMA under the PCM with $\varepsilon_r = 7$

Same as above, with $\varepsilon_r = 7$

15. ISA-DMA under the PCM with $\varepsilon_r = 11$

Same as above, with $\varepsilon_r = 11$

16. Model0-iso, Model0-aniso and DIFF DMACRYS input file

As the input files are identical to their ORIENT counterparts save for formatting and unit conversions.

7 Conclusions and Further Work

The non-empirical distributed intermolecular force-field (DIFF) developed for the gas-phase dimers of pyridine has been adapted for modelling crystal structures using DMACRYS. The rigid molecule approximation coupled with a non-empirical pairwise model with a many-body polarization model reproduces the lattice structure well (**Chapter 3**), and using a genuine potential energy surface showed the strengths and limitations of using the DIFF to predict properties at a range of pressures and temperatures (**Chapter 3**). The DIFF was even capable of finding the high-pressure polymorph of pyridine due to its accurate modelling of the repulsive wall of its potential energy surface (PES).

The extension of the DIFF approach to energetics started by assessing the importance of the flexibility of nitro-groups within crystal structures, the transferability of their charge distributions and its effect on electrostatic properties (**Chapter 4**). The advanced partitioning of high-quality molecular charge distributions of the closely related nitroaromatic molecules showed very limited transferability of the NO_2 charge distribution between themselves, and the variation with conformation. These results showed that molecule-specific force-fields were required and that conformational flexibility was an important issue. The conformational dependence of the charge would be a factor that would need to be modelled to predict energetic properties like impact sensitivity, in order to employ early empirical correlations (**4.3**).

A DIFF was developed for trinitrobenzene (TNB), as it considered an energetic building block,[148] and the polymorphic nitrobenzene that would be best modelled as a rigid molecule. This development produced new challenges as TNB is considerably larger than pyridine and the smaller molecules, which were used in developing the DIFF approach.[38, 60, 165, 227] Although the molecule has high symmetry, requiring only 5 atom-types, it is quite polarizable. The lessons learned about each component of the DIFF are discussed in **7.1**. Compared to pyridine, the much greater difference between the relative orientations sampled in the crystal and the gas-phase dimer structures for TNB make it challenging. Due to the larger molecule size and configuration space required to be sampled it was also computationally expensive to absorb errors in individual contributions into the models by fitting. In comparison to pyridine,[38] only a small number of total SAPT(DFT) intermolecular interaction energies could be calculated. This DIFF did provide a reasonable CSP study of TNB, further developing the CSP workflow (**7.2**), which raises some questions for further experimental work as well as theoretical study on crystalline TNB (**Chapter 6**).

The benefit of producing a model intermolecular potentials, such as the DIFF, is that in principle, it could be used to simulate a far wider range of properties than is possible with periodic electronic structure modelling, without double-counting temperature and other effects, which is implicit in using empirical force-fields. The future work needed to exploit the benefits of a DIFF are briefly discussed in **7.3** and **7.4**.

7.1 The theory of intermolecular forces and distributed intermolecular force-fields (DIFFs)

This thesis shows the value of computational studies on a molecule using CSP. One can conclude that a non-empirical model advances well beyond current empirical force-fields, and that building a force-field from solid theory[164] and state-of-the-art approaches provides unparalleled benefits. It allows for a more reliable description of the intermolecular potential energy surface, a better decomposition and understanding of the intermolecular interaction energy, and the ability to truly delve into the importance of various contributions to the intermolecular interactions in a crystalline lattice. Subsequently, a DIFF would allow one to derive a crystal's physical properties that would otherwise be costly or difficult to study experimentally with confidence that the underlying PES was on a sound theoretical footing with quantified errors. The uncertainty in using an empirical force-field for molecules or properties beyond the range of experimental data for which the force-field has been validated against is a common issue in the modelling done in specialty industries like energetics and pharmaceuticals.

7.1.1 Non-empirical vs empirical models

The non-empirical force-fields have a functional form based on the theory of intermolecular forces[164] with some chemical intuition (i.e. using symmetry to decide the number of distinct atom types). Pairing the simpler, isotropic exp-6 Buckingham functional forms of the empirical FIT & WILL potentials with a point-charge model or a less elaborate method of distributed multipole analysis can lead to uncertainties. Especially coupled with transferability assumptions, extrapolating to parts of the PES outside of the experimental conditions sampled by the dataset used for parameterisation can be unreliable at times. This was the case for pyridine (**Chapter 3**), where the DIFF was able to be used to model the effects of pressure on the relative stability of polymorphs, while the empirical models were not as successful. This is certainly a massive leap forward. The distributed intermolecular force-field (DIFF) was able to re-rank the pyridines crystal structures and suggest the correct lattice structure for form III, while the empirical FIT model was unable to accurately model the repulsive wall of the pyridine PES. This was due to the many errors, approximations and effects that have been averaged and absorbed into empirical potentials, deriving from their fitting to experimental data that has been obtained at various conditions.

Another point of caution is the comparison of relative stabilities and cell geometries of the static lattice to experimental determinations at ambient conditions. The empirical models have absorbed many phenomena like pressure and temperature effects in their parameterisation, and so can have expanded densities that might better match crystal determinations at higher temperatures. However, these are not accurate or reliable representations of a crystal's physical properties. We see this is the case when we compare the cell geometries of the polymorphs of pyridine in ***Figure 14*** in **Chapter 3**, and the cell geometries of the polymorphs of TNB in ***Figure 48***. Due to the absorption of many errors and approximations in empirical potentials, it is impossible to de-convolute the influence of each intermolecular contribution on physical properties.

The DIFF has the advantage that quantum and ambient effects can be included explicitly and so their influence can be properly observed. For example, **Chapter 3** also attempted to model the importance of free energy, though this was just a preliminary study, one can see the importance of isotope effects on polymorphism and the relative stabilities of polymorphs. In **Chapter 6**, the non-empirical models outperform empirical force-fields in CSP, with the DIFF and the 1st order Model0-iso finding TNB's form III as the global minimum, while Model0-aniso found form III as the 3rd most stable polymorph, within 3.5 $kJ\ mol^{-1}$ of the global minimum. The FIT+GDMA model that is typically used in CSP studies[23, 24] failed to find any approximation to TNB form III in its top 200 structures, and had the *xminopt* structure at 1600th. These simpler intermolecular force-fields, such as the FIT *exp*-6 model, although cheaper, do not consistently extend to molecules and conditions they have not been fitted to. For TNB this resulted in both FIT models generating artefact structures in their CSP that one could not expect to be possible unobserved polymorphs (**6.3.3.2**). Non-empirical force-fields have the advantage that the approximations used are known and can be systematically built upon to provide a realism unobtainable with empirical methods.

7.1.2 The electrostatic model

Deriving our long-range terms directly from the molecular charge density gives us confidence that the non-empirical force-fields are accurately capturing the importance of each long-range contribution. A rigorous partitioning of the molecular charge density using the Iterated Stockholders Atoms (ISA) approach allows one to obtain a more physical representation of the distributed charge density, which has minimal sensitivity to basis set and hence the artefacts of conformational change are minimized. Using the ISA partitioning in **Chapter 4** showed that the degree of observed conformational variations in nitro-groups, can cause sufficient changes in the molecular charge density to affect proposed empirical correlations between energetic properties and molecular electrostatic properties. Furthermore, these changes were significant enough to suggest that a single electrostatic model could not be used to dependably model the lattice energies of different conformers. Analytically rotating the multipole moments is a possible solution that could cheaply approximate finding low energy crystal structures but not relative energies (**Chapter 4** ***Table 15***). An inadequate description of the multipolar electrostatic energy has significant influences on the predicted intermolecular interactions and associated electrostatic and energetic properties (**4.4**).

In **Chapters 3** & **6**, we see that the influence of electrostatic model on the intermolecular interactions sampled and favoured on a potential energy surface is enormous. Using the more conventional partitioning method (GDMA) and a cheaper wave-function produces a very different intermolecular potential energy landscape compared to the more elaborate ISA partitioning alongside a high-quality basis set (**Chapter 3** ***Figure 18*****, Chapter 6** ***Figure 54***). Minimizing structures directly from the CrystalPredictor generated output allows us to really appreciate the influence of using distributed multipoles rather than point charges as the electrostatic model, and the method of partitioning and the quality of the wave-function on the PES (**6.3.3.2**). There is a more profound change in the re-ranking of structures, the FIT+ISA model manages to predict the observed forms of TNB within a reasonable energy range of the global minimum ($\sim 5\ kJ\ mol^{-1}$), while the FIT+GDMA model predicts the forms of TNB an astounding $35\ kJ\ mol^{-1}$ from the global minimum. For TNB, the GDMA (PBE0/6-31G(d, p)) electrostatic model is certainly not suitable for a reliable CSP study. Its widespread use has mainly been amongst pharmaceutical molecules; a series of CSP studies on six chloro-nitrobenzenes[412] used the original DMA method on a MP2/6-31G(d,p) charge density.

7.1.3 The polarization energy

Accurately modelling the polarization contribution to the intermolecular interaction energy has been a challenge. From the molecular charge density, we have derived state-of-the-art explicit polarization models for pyridine and TNB, which are non-pairwise additive and the only many-body term in the non-empirical force-fields. This is certainly pioneering work but it has raised many questions with regards to the importance of higher-order induction terms (**Chapter 3 & 5**) and the possible over-estimation of the polarization energy in the non-polar but polarizable TNB molecule (**6.4.4.1**). In **Chapter 5**, we found the damping of the polarization energy to be one of the largest sources of error in the DIFF as using a single parameter damping function fails to concisely model the polarization energy across all separations for every configuration sampled in both the gas-phase and the crystal lattice.

Computing the polarization forces to convergence is a computationally expensive calculation (**2.6.1.3**) that is not currently feasible for the number of structures generated from in a CSP study. After carrying out one-off calculations for the experimental forms of the polar pyridine in **Chapter 3**, the effects of polarization on the final cell geometry was suggested to be negligible. In **Chapter 6** the polarizable continuum model (PCM) was explored, which is normally used after obtaining the lowest energy structures, to observe changes in relative energies with polarization effects.[413] For TNB, it appeared that the polarization has minor influence on the crystal structures of lattice minima (**Chapter 6** ***Figure 52*** & ***Table 27***). The issues with the PCM lay in its unreliable prediction of the final lattice energy, however, in future CSPs, a non-empirical potential coupled with a PCM with an appropriate dielectric constant ($\varepsilon_r \approx 3$) could be used as an intermediary model for minimizing structures to find the most stable few. The lowest energy structures could then be re-minimized using a more expensive model that numerically iterates the forces due to an explicit polarization model to convergence.

7.1.4 Dispersion in weakly-bound organic crystals

In this thesis, we see that a genuine modelling of the dispersion is vital to reproducing the true PES of these organic systems. We find the dispersion to be a significant energy in both the gas-phase dimers and solid-state of pyridine (**Chapter 3**) and TNB (**Chapter 5** & **6**), making up approximately 150% and more of the intermolecular lattice energy. The non-empirical models attempt to accurately model the dispersion contribution over a range of separations by including the higher-order C_8 and C_{10} dispersion coefficients, which aims to better capture the dispersion at small R. The coverage of separation by the higher-order terms also better captures the influence of dispersion on density. The lattice summation used to calculate the lattice energy for crystal structures samples the potential over a much greater separation than for a gas-phase dimer or trimer, thus for empirical models like the FIT, a small error in the C_6 dispersion coefficient results in a large error due to lattice summations. This is also true for using higher-order multipole moments over a point-charge electrostatic model.

This improvement in the theoretical basis of the dispersion model coupled with a damping parameter derived from high quality 2nd order SAPT(DFT) calculations not only greatly increases the accuracy of dispersion energies but also their reliability. As with the polarization model, the largest source of error in our dispersion model is the use of a single parameter damping function (**5.5.4**). Single parameter damping functions are currently the only types of damping functions accepted in DMACRYS, however, we hope future codes will be built that are capable of employing atom-atom orientation dependent damping functions. We see in **Chapters 3, 5** & **6** that the range of orientations sampled in the crystal and gas-phase is far too rich for the repulsive wall to be truthfully modelled using a single damping parameter.

Nevertheless, the sound theoretical foundation of the non-empirical models compared to the conventional empirically-fitted C_6 models is certainly a step in the right direction for a genuine modelling of the PESs of future novel materials. The absorption of higher-order dispersion coefficients and many-body dispersion effects by empirical models makes it very difficult to decompose the influence of each dispersion contribution. For example, for pyridine (**Chapter 3**) by being able to study each contribution, it is possible to suggest that including the many-body dispersion may have destabilized the global minimum ($_{\text{pyr}}2$) and resulted in the observed forms being found as global minima in the CSP. The exclusion of the many-body dispersion has resulted in denser predicted structures, while the inclusion of the δ_{HF} resulted in the favouring of stacked dispersion dominated two molecule contacts in pyridine (**3.4.5**).

7.1.5 Creating a truthful description of the short-range interactions

Deriving the genuine potential energy surface of an organic molecular crystal is no easy feat. Particularly for larger, flexible more anisotropic molecules (**Chapter 5**). Through this thesis one can see that deriving a theoretically based model that accounts for all the nuances and contributions in the intermolecular interaction energy is very difficult for molecules larger than water. Water itself is a challenge.[46] Currently some approximations must be made (**Chapter 5**), and errors must be absorbed (**Chapter 3**) due to the computational and theoretical undertakings faced. For example, the 1st order short-range energies were fit using a Gaussian-log weighting function. However, this function could not be used for the total $E_{int}^{(2)}[SAPT(DFT)]$ energies, as it cannot go through 0, thus, a weighting function of a different shape had to be used. The short-range portion of the potential must be fitted as analytically deriving the terms is non-trivial (**2.9**), yet in fitting one runs into another problem. For the molecules of interest, the cost of high-quality, total intermolecular interaction energy calculations can be very expensive.

Over 2000 pseudo random dimer configurations were sampled in the fitting of the Model0s from the 1st order SAPT(DFT) interaction energies (**5.4.2**), compared to the 327 dimer configurations used to relax the final model. For a solid description of the PES both models need to sample an ample amount of the configuration space, which Model0 certainly does. ModelD demonstrates that (**5.5.3.3**) a poor sampling can result in a second-order model that actually performs worse than the cheaper well sampled first-order models because of an over-biasing of the parameters. It can be tough to combat poor sampling for larger, more expensive molecules. For TNB, each 2nd order SAPT(DFT) intermolecular dimer interaction energy calculation ($E_{int}^{(2)}[SAPT(DFT)]$) took a minimum 15 computing hours, and thus, adequately sampling the two molecule contact configuration space becomes an issue. How does one efficiently sample important interactions without biasing the potential? In the construction of these non-empirical force-fields one must be sure to sample the intermolecular potential energy surface extensively (**Chapter 5**). Gas-phase dimer configurations typically sample a limited region of configuration space, especially for large molecules. For larger molecules, crystal packing and long-range effects really play a part in the resultant close-contacts and crystal structure. A crystal structure samples a wider range of relative orientations, so in potential development one must try to account for this while including a test set (**5.5.3.2**) to challenge the resultant models. Furthermore, the task of interfacing and pushing current and new experimental programs for modelling can become a coding challenge, to say the least. These larger molecules can "break" codes, and demand careful treatment and monitoring. The development of a new methodology for CSP alone is a serious undertaking. It has been through great care and time that the novel distributed intermolecular force-fields (DIFFs) have been transferred to the pyridine solid-state (**Chapter 3**), and developed and applied to TNB for CSP, and in both cases they have outperformed current empirical treatments.

7.2 Changes to the CSP workflow

Another big development and progression within this thesis for the field of Crystal Structure Prediction (CSP), was the changes and updates that had been made to the specific CSP workflow and methodology (***Figure 1***) used in the Price group, to allow the use of more complex force-fields. In addition, the approach of minimizing the lattice energies of the generated structures directly with the high quality force-fields (**Chapter 6**) and not using an intermediate model (**Chapter 3**), is a new workflow that has allowed for some interesting structures to be found (**Chapter 6** ***Figure 49, Figure 51*** & ***Figure 52***). Previously, the majority of force-field approaches used completely isotropic force-fields (including an isotropic point-charge electrostatic model)[22-25, 93] to model the intermolecular forces between molecules as this functional form was easily transferable between codes and could be used in Molecular Dynamics (MD) simulations. The increasing desire to study larger more complex systems, the importance of atomic shape (anisotropy) and conformation for these systems, coupled with the improvement of MD codes like DL Multi[414] that can take anisotropic potentials, has resulted in a shift to developing more dependable intermolecular force-fields from ab initio methods.[60, 167, 415] Thanks to major leaps in computing capabilities and the optimization of calculations within codes like MolPro,[416] NWCHEM[402] and Psi4,[228] the ability to use SAPT(DFT), obtain adequate charge densities, and the reliability of novel non-empirical force-fields has truly grown. These newer codes are still very experimental and under constant development but have been vital in the progress made in this thesis. Molecule size is still a huge limiting factor, and many bugs, conversion and interface issues between quantum chemistry codes have been overcome in order to produce these non-empirical force-fields. The scripts and codes developed within this thesis for CSP can be found in the **Appendix CD**. However, this thesis shows that it is very possible to generate an anisotropic non-empirical atom-atom intermolecular force-field for a medium sized molecule using only its chemical diagram and attempt to model its genuine PES. The work here hopes to augment current foundational knowledge on force-field development, and really is just the tip of the ever-growing iceberg that is CSP. It is hoped that future work learns and develops on the challenges that have been faced within this thesis in order to generate transferable, inexpensive force-fields that rival the highest quality electronic structure methods.

7.3 Beyond the lattice energy

Chapters 3, 4 & **6** emphasise the importance of properties beyond the lattice energy in CSP. In **Chapter 3,** the importance of pressure is explored resulting in the unexpected discovery of the third form of pyridine.[293] Both **Chapters 3** & **6** highlighted the effects of temperature on crystal structure and the importance and influence of higher order induction terms and many-body dispersion terms. The conventional practice of computing the static lattice energy ($T = 0K$) is only an initial step in determining the crystal structures that may be synthesised in a lab. The most thermodynamically stable structures may not be the kinetically observed polymorphs.[417] This seems to be the case for TNB form I (**Chapter 6**).[263] However, to determine the possibility of disappearing polymorphs,[312] the prediction of unobserved more thermodynamically favourable structures is a major justification[30] for the development and testing[22] of innovative CSP treatments. **Chapter 3** highlights that a significant reshuffling of the energy ranking of structures can occur with changes in temperature or pressure. This may not occur through a phase transformation but instead via serendipitous recrystallization experiments in the presence of seeds of a polymorph or previously unexplored additive pathways.[263] Particularly for energetic materials it can be difficult to obtain and characterize different crystal structures of a molecule, especially as the experimental conditions can vary the kinetics of crystallization. Crystal structures that share many similar molecules within a coordination sphere with the observed polymorph (**Chapter 3** & **6**) may be thermodynamically more stable but not kinetically favoured. The DIFF suggests that both pyridine and TNB may have some unobserved thermodynamically competitive polymorphs that are simply waiting to be found.

Future uses of novel force-fields could go beyond lattice or free energies (**Chapter 4**), and also speculate the explosive properties of energetic materials (**Chapter 4**). By deriving the electrostatic, polarization and dispersion models directly from the molecular charge density, the effects of conformation on electrostatic properties like the electrostatic potential maximum V_{max} can be reliably analysed and used to extrapolate possible energetic properties using previously proposed empirical correlations.[137, 168, 236, 324, 355, 393, 418, 419] A poor definition of the molecular charge density using electrostatic models that are inaccurate for larger systems (**2.5.1**) can result in inaccurate predictions and will fail to model changes in charge density with crystalline environment. The iterated stockholder atoms (ISA) partitioning has proved to be the next generation treatment for a physically sound description of the distributed charge density. The ISA electrostatic models have not only produced tangible atomic charges but have permitted a deeper understanding of the influence of chemical environment and torsion angle on multipole moments, intermolecular interactions and energetic properties (**Chapter 4**).

7.4 Next generation modelling

The complexity of the processes that determine the impact sensitivity and other key properties of the polymorphs of energetic materials,[346] implies that multi-scale modelling, involving both periodic electronic structure calculations and MD simulations will be required. The MD codes will need to use the best non-empirical models for the inter- and intramolecular forces, derived directly from the molecular charge density. These models will be specific to the molecule, as the charge distribution around the nitro groups is clearly affected by changes in the rest of the molecule, particularly the adjacent functional groups or with the presence of an aromatic ring (**Chapter 4** ***Figure 25***).

7.4.1 Conformation

Considering the polymorphs of energetic molecules and CSP studies provides a good starting point for developing the next generation of molecule-specific flexible force-fields. This in turn will help contribute to the development of predictive computational modelling of their key physical properties such as free energies, thermal expansion, and other non-reactive properties. We find that for larger energetic molecules, molecular conformation should not be overlooked (**Chapter 4** & **6**). For every novel material studied, how conformation affects its molecular charge distribution and its intermolecular interactions should be thoroughly investigated. The difference in the diversity of gas-phase dimers found for pyridine (**Chapter 3**) and TNB (**Chapter 5**), emphasises that for larger molecules under a rigid molecule approximation, sampling the most favourable gas-phase dimer configurations is not enough to build an accurate representation of the intermolecular interactions in the solid state. Pyridine had 8 very different gas-phase dimers under 3 different classes of contact, compared to the 5 gas-phase dimers of TNB, 4 of which are in a stacked configuration. The percentage contribution of each intermolecular energy term for pyridine's gas-phase dimers also differed markedly (**Chapter 3** ***Figure 13***) compared to TNB's dimers (**Chapter 5** ***Figure 42***). Did the rigid molecule approximation limit the diversity of gas-phase minima as well? Possibly. It is likely that this is why so few of TNB's gas-phase dimers were seen in the lowest energy crystal structures (**Chapter 6** ***Table 29***), whereas the experimentally more rigid pyridine exhibited a richer range of gas-phase motifs (**Chapter 3** ***Table 7***).

Therefore, there may be a significant loss in accuracy if novel intermolecular potentials do not include an NO_2 torsion angle dependence. Future CSP studies of energetic materials should certainly attempt to include conformational flexibility using a high quality electrostatic model. It could be possible to fit analytical models to changes in atomic multipoles[420] as **Chapter 4** ***Figure 25*** shows that the ISA atomic multipoles change relatively smoothly with NO_2 torsion angle and therefore can be modelled using splines, or other functions of low complexity. However, while it is clear how these changes can be modelled as a function of a single degree of freedom (single torsion), the correlated conformational adjustments of neighboring groups on the aromatic ring may pose a challenge. The changes in multipole moments during the rotation of the methyl group in TNT, let alone with changes in the aliphatic ring seen for RDX (**Chapter 4** ***Figure 22***), will not be represented by a simple Fourier series term, and this will also apply to the conformational energy penalty.[166]

In **Chapter 4**, one sees the effects of conformation and approximations to conformation (analytical rotation, holding the molecule rigid etc.) on the electrostatic properties and final lattice energies. Even small geometric differences can result in significant changes in relative energies, and therefore, different CSP landscapes. Small differences in conformation may not change the intermolecular close-contacts significantly enough for dramatic changes in cell geometry but it can influence energy rankings. Yet, if the rigid molecule approximation can be made in force-fields for novel materials it should. It lowers the cost of the potential and required time for development by a considerable amount and is still capable of modelling the genuine potential energy surface of a weakly-bound organic molecular crystal. **Chapter 3**, exemplifies this in predicting the high pressure polymorph of pyridine in a CSP study, and in showing the effects of deuteration on the pyridine crystal structure are small but can result in different relative energies between its 3 polymorphs and their free energies. Isotope effects within polymorphism is not something to be ignored.

7.4.2 Next generation modelling of the polarization contribution

The explicit inclusion of the non-additive many-body polarization term in the DIFF has been a groundbreaking component of these non-empirical potentials, which has also proved challenging. Comparing DIFF(no-pol) and DIFF for both pyridine (**3.3.3.1**) and TNB (**6.3.3.3**), one can see that the influence of this energy on the relative energies of crystal lattices is huge. The long-range polarization contribution is vital in representing the genuine PES. Yet, how does one correctly model this term if many molecular mechanics codes cannot include explicit polarization models and if they do they only take a single parameter damping function?[59] How does one affordably include the forces due to polarization for more polarizable or polar molecules? It appears that future CSP studies will have to use a combination of the polarizable continuum model (**6.3.1**) and lattice energy minimizations that include the fully iterated polarization forces (***Table 6***) on important low energy structures in order to concisely determine the polarization contribution. Moreover, the polarization damping parameter should be adjusted iteratively through successive CSP studies in order to damp the multipolar polarization energy $V_{POL}^{(2-\infty)}[DM]$ appropriately, as not doing this could lead to an overestimation of the polarization in the lattice (**6.4.4.1**).

There are other terms that have not been as thoroughly investigated in this thesis, chiefly, the higher-order induction and many-body dispersion contributions, which have been found to be important for other molecules.[38, 40, 200, 421-423] But are they important for larger more anisotropic molecules like TNB? These terms are specific to the molecule, and as they are non-additive they also depend on the preferred configurations of the lowest energy crystal structures. For example, depending on configurations seen in the lattice co-ordination clusters, the many-body dispersion could be attractive or repulsive. Molecule specific terms will require concentrated investigations beforehand, like CCSD(T) calculations to obtain a precise description of these non-additive contributions to the intermolecular interaction energy.

7.4.3 Speed and accuracy

In comparison to empirically fit models or some isotropic non-empirical models,[93, 161] the DIFF is an expensive potential. Attempting to produce a non-empirical force-field that truthfully describes each contribution to intermolecular energy, with components that have a physically meaningful derivation, has been computationally expensive and will only be more challenging for the more complex systems investigated in industry. Recently, some alternative treatments for the fast development of cheap, accurate bespoke force-fields that rival periodic DFT accuracy have become prevalent. One of the most interesting alternatives are machine-learning or neural network (NN) approaches to CSP.[424-432]

It is almost guaranteed that some form of machine-learning methodology will be present in the next blind test, but currently these treatments face similar issues to empirical methods in that they are very specific to the data they have been parameterised to. More so than conventional empirical models, NN models are not transferable between molecules and or even between conformations.[431] Unlike the non-empirical models (DIFF and Model0), the interaction energies obtained from these approaches cannot be separated into any physically meaningful contributions because of the handling of the initial inputs into the NN. Furthermore, the functional forms used in the network layers have no physical link to the components of the energy and are purely mathematical representations of a property.[431] Nevertheless, the use of artificial intelligence (AI) in CSP is certainly exciting and re-enforces the importance and relevance of CSP. A truthful modelling of the genuine potential energy surface will always be a central goal in computational chemistry.

7.5 Conclusion

To conclude this thesis has found the gas and solid phases of pyridine and TNB can be modelled realistically using a rigid non-empirical anisotropic atom-atom intermolecular force-field. The distributed intermolecular force-fields (DIFFs) developed for pyridine and trinitrobenzene (TNB) are the first two non-empirical atom-atom anisotropic force-fields with explicit polarization models ever used in CSP. The success of the DIFF in predicting the solid-state of pyridine (a building block in the pharmaceuticals industry) and TNB (a building block in the energetics industry) is very promising in the quest for reliable and accurate force-fields for novel materials in specialty industries.

These non-empirical force-fields provide a realism that surpasses the empirical force-fields that have been used previously used for crystal structure prediction. Empirical models have the benefit of absorbing many errors and approximations, some degree of transferability and simplifying the functional form to enable the use of the models in many codes and simulation methods. The empirical force-fields are a sort of hit and miss approach that isn't consistently reliable but generally works. The DIFF, on the other hand, has been derived using the theory of intermolecular forces (**Chapter 2**) and SAPT(DFT) calculations using a suite of quantum chemistry programs (**Chapter 5**). The development and implementation of these force-fields is far more elaborate and time consuming. However, one can see that for the molecules in the energetic and pharmaceutical industries, this level of detail is vital if we wish to have a genuine intermolecular potential that can be used to simulate different phases over a range of pressures and temperatures. The DIFF demonstrates that an accurate potential energy surface can be obtained and applied to condensed phases, if we perceptively use molecule-specific knowledge and chemical intuition regarding the effects of certain approximations on modelling the crystal structure.

The rigid molecule approximation is a good approximation for pyridine, and when coupled with a non-empirical pairwise model and a many-body polarization model it reproduces the lattice structure well and can produce a genuine potential energy surface capable of predicting properties at non-ambient pressures and temperatures (**Chapter 3**). The DIFF was even capable of finding the high pressure polymorph of pyridine due to its accurate modelling of the repulsive wall of pyridine's PES. The extension of the DIFF workflow (**Chapter 1 *Figure 2***) to energetic materials brought new challenges and improved insights into the process. The flexibility of the NO_2 groups, and the dependence of its charge distribution on the molecule and conformation (**Chapter 4**) reduced the consistency of the DIFF CSP results, even for the most symmetrical, planar energetic nitroaromatic, TNB. Analytical rotation of the multipole moments will only assist with including flexibility in determining the minimized structures, as the changes due to re-distribution of charge with torsion angle affects the relative energies (**Chapter 4 *Table 15***).

The size of TNB brought challenges in performing an adequate enough number of SAPT(DFT) calculations with a suitable quality molecular charge density, particularly when calculating the second-order and higher energies (**Chapter 5**). TNB also illustrated the need to really cover a wide range of accessible relative orientations with second-order energy calculations, in order to ensure that the final relaxed and refined model provided a worthwhile improvement on anisotropic models derived from 1st order SAPT(DFT) calculations like Model0-aniso (**5.6.2**). The DIFF performed notably better than conventional empirical models in the CSP of the larger more flexible TNB, as well as pyridine. The two resultant publications[167, 168] and the impending publication on the DIFF for TNB are testament to the success of this thesis in the development of non-empirical force-fields for weakly bound organic molecular crystals.

References

1. Larentzos, J. P.; Rice, B. M.; Byrd, E. F.; Weingarten, N. S.; Lill, J. V., Parameterizing Complex Reactive Force Fields Using Multiple Objective Evolutionary Strategies (MOES). Part 1: ReaxFF Models for Cyclotrimethylene Trinitramine (RDX) and 1, 1-Diamino-2, 2-dinitroethene (FOX-7). *Journal of Chemical Theory and Computation* **2015,** *11* (2), 381-391.

2. Byrd, E. F.; Rice, B. M., Ab Initio Study of Compressed 1, 3, 5, 7-Tetranitro-1, 3, 5, 7-tetraazacyclooctane (HMX), Cyclotrimethylenetrinitramine (RDX), 2, 4, 6, 8, 10, 12-Hexanitrohexaazaisowurzitane (CL-20), 2, 4, 6-Trinitro-1, 3, 5-benzenetriamine (TATB), and Pentaerythritol Tetranitrate (PETN). *The Journal of Physical Chemistry C* **2007,** *111* (6), 2787-2796.

3. Neumann, M. A.; van de Streek, J., How many ritonavir cases are there still out there? *Faraday Discussions* **2018,** *211*, 441-458.

4. Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J., Ritonavir: An extraordinary example of conformational polymorphism. *Pharmaceutical Research* **2001,** *18* (6), 859-866.

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. *Organic Process Research & Development* **2000,** *4* (5), 413-417.

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

7. Cruz-Cabeza, A. J.; Reutzel-Edens, S. M.; Bernstein, J., Facts and fictions about polymorphism. *Chemical Society Reviews* **2015,** *44*, 8619-8635.

8. Alvarez-Quiceno, J. C.; Dalpian, G. M.; Osorio-Guillen, J. M., A systematic first-principles study of the tungsten trioxide polymorphs. *Physica Status Solidi B-Basic Solid State Physics* **2015,** *252* (10), 2290-2295.

9. Sun, C. C., Material Science Tetrahedron-A Useful Tool for Pharmaceutical Research and Development. *Journal of Pharmaceutical Sciences* **2009,** *98*, 1671-1687.

10. Lee, A. Y.; Erdemir, D.; Myerson, A. S., Crystal Polymorphism in Chemical Process Development. *Annual Review of Chemical and Biomolecular Engineering, Vol 2* **2011,** *2*, 259-280.

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

12. Best, D., Limited (Research Triangle Park, NC) and Novopharm Ltd.(Schaumberg, 111)(Glaxo Inc. v. Novopharm Ltd., 830 F. Suppl. 871, 29 USPQ 2d 1126 [EDNC. *Biotechnology Law Report* **1995,** *423* (3).

13. Wegner, H. C.; Center, D., PATENT LITIGATION. *Biotechnology Law Report* **1993,** *557* (5).

14. Bernstein, J., Polymorphism and Patents. In *Polymorphism in Molecular Crystals*, Oxford University Press: Oxford, 2002; Vol. 1, pp 297-307.

15. Andersen, W. H., Comments on "The chemistry of detonations". *Combustion and Flame* **1982,** *45*, 309-316.

16. Kamlet, M. J.; Jacobs, S. *THE CHEMISTRY OF DETONATIONS. 1. A SIMPLE METHOD FOR CALCULATING DETONATION PROPERTIES OF CHNO EXPLOSIVES*; DTIC Document: 1967.

17. Cooper, P. W.; Kurowski, S. R., *Introduction to the Technology of Explosives*. VCH New York: 1996.

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

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

20. Holden, J. R.; Du, Z. Y.; Ammon, H. L., Prediction of Possible Crystal-Structures For C-, H-, N-, O- and F-Containing Organic Compounds. *Journal of Computational Chemistry* **1993,** *14* (4), 422-437.

21. 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. *International Journal of Pharmaceutics* **2011,** *418* (2), 168-178.

22. Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A.; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J.

A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar, L.; Gatsiou, C.-A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Jankiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N. J. J.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. J.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; de Wijs, G. A.; Yang, J.; Zhu, Q.; Groom, C. R., Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B* **2016,** *72* (4), 439-459.

23. Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S. X. M.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; Della Valle, R. G.; Desiraju, G. R.; van Eijck, B. P.; Facelli, J. C.; Ferraro, M. B.; Grillo, D.; Habgood, M.; Hofmann, D. W. M.; Hofmann, F.; Jose, K. V. J.; Karamertzanis, P. G.; Kazantsev, A. V.; Kendrick, J.; Kuleshova, L. N.; Leusen, F. J. J.; Maleev, A. V.; Misquitta, A. J.; Mohamed, S.; Needs, R. J.; Neumann, M. A.; Nikylov, D.; Orendt, A. M.; Pal, R.; Pantelides, C. C.; Pickard, C. J.; Price, L. S.; Price, S. L.; Scheraga, H. A.; van de Streek, J.; Thakur, T. S.; Tiwari, S.; Venuti, E.; Zhitkov, I. K., Towards crystal structure prediction of complex organic compounds - a report on the fifth blind test. *Acta Crystallographica Section B-Structural Science* **2011,** *67*, 535-551.

24. 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 Crystallographica Section B - Structural Science* **2009,** *65* (2), 107-125.

25. 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 Crystallographica Section B - Structural Science* **2005,** *61* (5), 511-527.

26. 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 Crystallographica Section B - Structural Science* **2002,** *58*, 647-661.

27. 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 Crystallographica Section B - Structural Science* **2000,** *56*, 697-714.

28. Vasileiadis, M.; Pantelides, C. C.; Adjiman, C. S., Prediction of the crystal structures of axitinib, a polymorphic pharmaceutical molecule. *Chemical Engineering Science* **2015,** *121*, 60-76.

29. Pantelides, C. C.; Adjiman, C. S.; Kazantsev, A. V., General Computational Algorithms for Ab Initio Crystal Structure Prediction for Organic Molecules. *Topics in Current Chemistry* **2014,** *345*, 25-58.

30. Price, S. L.; Braun, D. E.; Reutzel-Edens, S. M., Can computed crystal energy landscapes help understand pharmaceutical solids? *Chemical Communications* **2016,** *52*, 7065-7077.

31. Sanner, M. F., Python: A programming language for software integration and development. *Journal of Molecular Graphics & Modelling* **1999,** *17* (1), 57-61.

32. Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C., The Cambridge Structural Database. *Acta Crystallographica Section B-Structural Science Crystal Engineering and Materials* **2016,** *72*, 171-179.

33. Kendrick, J.; Stephenson, G. A.; Neumann, M. A.; Leusen, F. J. J., Crystal Structure Prediction of a Flexible Molecule of Pharmaceutical Interest with Unusual Polymorphic Behavior. *Crystal Growth & Design* **2013,** *13* (2), 581-589.

34. Price, S. L.; Price, L. S., *Computational Polymorph Prediction*. 2011; p 427-450.

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

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

37. Sobol, I. M., *USSR Computational Mathematics and Mathematical Physics* **1967,** *7*, 86-112.

38. Misquitta, A. J.; Stone, A. J., Ab Initio Atom-Atom Potentials Using CAMCASP: Theory and Application to Many-Body Models for the Pyridine Dimer. *Journal of Chemical Theory and Computation* **2016,** *12* (9), 4184-4208.

39. 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.

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

41. Brandenburg, J. G.; Grimme, S., Organic crystal polymorphism: a benchmark for dispersion-corrected mean-field electronic structure methods. *Acta Crystallographica Section B-Structural Science Crystal Engineering and Materials* **2016,** *72*, 502-513.

42. Brandenburg, J. G.; Grimme, S., Accurate Modeling of Organic Molecular Crystals by Dispersion-Corrected Density Functional Tight Binding (DFTB). *Journal of Physical Chemistry Letters* **2014,** *5* (11), 1785-1789.

43. Iuzzolino, L.; McCabe, P.; Price, S. L.; Brandenburg, J. G., Crystal structure prediction of flexible pharmaceutical-like molecules: density functional tight-binding as an intermediate optimisation method and for free energy estimation. *Faraday Discussions* **2018,** *211*, 275-296.

44. Adjiman, C. S.; Brandenburg, J. G.; Braun, D. E.; Cole, J.; Collins, C.; Cooper, A. I.; Cruz-Cabeza, A. J.; Day, G. M.; Dudek, M.; Hare, A.; Iuzzolino, L.; McKay, D.; Mitchell, J. B. O.; Mohamed, S.; Neelamraju, S.; Neumann, M.; Nilsson Lill, S.; Nyman, J.; Oganov, A. R.; Price, S. L.; Pulido, A.; Reutzel-Edens, S.; Rietveld, I.; Ruggiero, M. T.; Schon, J. C.; Tsuzuki, S.; van den Ende, J.; Woollam, G.; Zhu, Q., Applications of crystal structure prediction - organic molecular structures: general discussion. *Faraday Discussions* **2018,** *211*, 493-539.

45. Rigby, M.; Smith, E. B.; Wakeham, W. A.; Maitland, G. C., *The Forces Between Molecules*. Oxford Science Publication: 1986.

46. Jankowski, P.; Murdachaew, G.; Bukowski, R.; Akin-Ojo, O.; Leforestier, C.; Szalewicz, K., Ab Initio Water Pair Potential with Flexible Monomers. *Journal of Physical Chemistry A* **2015,** *119* (12), 2940-2964.

47. Gilmore, R. A. J.; Dove, M. T.; Misquitta, A. J., First-Principles Many-Body Nonadditive Polarization Energies from Monomer and Dimer Calculations Only: A Case Study on Water. *Journal of Chemical Theory and Computation* **2020,** *16* (1), 224-242.

48. Cogoni, M.; D'Aguanno, B.; Kuleshova, L. N.; Hofmann, D. W. M., A powerful computational crystallography method to study ice polymorphism. *The Journal of Chemical Physics* **2011,** *134* (20), 204506-204506.

49. Liem, S. Y.; Popelier, P. L. A., Properties and 3D structure of liquid water: A perspective from a high-rank multipolar electrostatic potential. *Journal of Chemical Theory and Computation* **2008,** *4*, 353-365.

50. Ren, P. Y.; Ponder, J. W., Temperature and pressure dependence of the AMOEBA water model. *Journal of Physical Chemistry B* **2004,** *108* (35), 13427-13437.

51. Guillot, B.; Guissani, Y., Polyamorphism in low temperature water: A simulation study. *Journal of Chemical Physics* **2003,** *119* (22), 11740-11752.

52. Mas, E. M.; Bukowski, R.; Szalewicz, K., Ab initio three-body interactions for water. II. Effects on structure and energetics of liquid. *Journal of Chemical Physics* **2003,** *118* (10), 4404-4413.

53. Matsumoto, M.; Saito, S.; Ohmine, I., Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing. *Nature* **2002,** *416* (6897), 409-413.

54. DelleSite, L.; Alavi, A.; LyndenBell, R. M., The electrostatic properties of water molecules in condensed phases: an ab initio study. *Molecular Physics* **1999,** *96* (11), 1683-1693.

55. Astrand, P. O.; Ruud, K.; Mikkelsen, K. V.; Helgaker, T., Atomic charges of the water molecule and the water dimer. *Journal of Physical Chemistry A* **1998,** *102* (39), 7686-7691.

56. Astrand, P. O.; Linse, P.; Karlstrom, G., Molecular-Dynamics Study of Water Adopting a Potential Function With Explicit Atomic Dipole-Moments and Anisotropic Polarizabilities. *Chemical Physics* **1995,** *191* (1-3), 195-202.

57. Hodges, M. P.; Stone, A. J.; Xantheas, S. S., Contribution of many-body terms to the energy for small water clusters: A comparison of ab initio calculations and accurate model potentials. *Journal of Physical Chemistry A* **1997,** *101* (48), 9163-9168.

58. Mas, E. M.; Bukowski, R.; Szalewicz, K., Ab initio three-body interactions for water. I. Potential and structure of water trimer. *Journal of Chemical Physics* **2003,** *118* (10), 4386-4403.

59. Welch, G. W. A.; Karamertzanis, P. G.; Price, S. L.; Leslie, M. *DMACRYS*, 1.05, is a substantial revision of DMAREL; 2010.

60. Misquitta, A. J.; Welch, G. W. A.; Stone, A. J.; Price, S. L., A first principles solution of the crystal structure of C6Br2ClFH2. *Chemical Physics Letters* **2008,** *456* (1-3), 105-109.

61. 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? *Journal of Chemical Theory and Computation* **2008,** *4* (3), 522-532.

62. Illingworth, C. J.; Domene, C., Many-body effects and simulations of potassium channels. *Proceedings of the Royal Society A-Mathematical Physical and Engineering Sciences* **2009,** *465* (2106), 1701-1716.

63. Jiang, H.; Jordan, K. D.; Taylor, C. E., Molecular dynamics simulations of methane hydrate using polarizable force fields. *Journal of Physical Chemistry B* **2007,** *111* (23), 6486-6492.

64. Brenner, D. W., Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. *Physical Review B* **1990,** *42* (15), 9458.

65. Brenner, D. W.; Shenderova, O. A.; Harrison, J. A.; Stuart, S. J.; Ni, B.; Sinnott, S. B., A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. *Journal of Physics: Condensed Matter* **2002,** *14* (4), 783.

66. Stuart, S. J.; Tutein, A. B.; Harrison, J. A., A reactive potential for hydrocarbons with intermolecular interactions. *The Journal of chemical physics* **2000,** *112* (14), 6472-6486.

67. Coombes, D. S., Deriving intermolecular potentials for predicting the crystal structures of polar molecules. *Philosophical Magazine B-Physics of Condensed Matter Statistical Mechanics Electronic Optical and Magnetic Properties* **1996,** *73* (1), 117-125.

68. 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. *Journal of Physical Chemistry* **1996,** *100* (18), 7352-7360.

69. Williams, D. E., Improved intermolecular force field for crystalline oxohydrocarbons including O-H...O hydrogen bonding. *Journal of Computational Chemistry* **2001,** *22* (1), 1-20.

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

71. Buckingham, A. D.; Utting, B. D., Intermolecular Forces. *Annual Review of Physical Chemistry* **1970,** *21* (1), 287-316.

72. Williams, D. E.; Cox, S. R., Nonbonded Potentials For Azahydrocarbons: the Importance of the Coulombic Interaction. *Acta Crystallographica Section B - Structural Science* **1984,** *40* (8), 404-417.

73. Day, G. M.; Chisholm, J.; Shan, N.; Motherwell, W. D. S.; Jones, W., Assessment of lattice energy minimization for the prediction of molecular organic crystal structures. *Crystal Growth & Design* **2004,** *4* (6), 1327-1340.

74. 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. *Crystal Growth & Design* **2005,** *5* (3), 1023-1033.

75. Pertsin, A. J.; Kitaigorodsky, A. I., *The Atom-Atom Potential Method. Applications to Organic Molecular Solids*. Springer-Verlag: Berlin, 1987.

76. Lifson, S.; Hagler, A. T.; Dauber, S., Consistent Force Field Studies of Intermolecular Forces in Hydrogen-Bonded Crystals. 1. Carboxylic Acids, Amides, and the C=O...H- Hydrogen Bonds. *Journal of the American Chemical Society* **1979,** *101* (18), 5111-5121.

77. Hesselmann, A.; Jansen, G., Intermolecular dispersion energies from time-dependent density functional theory. *Chemical Physics Letters* **2003,** *367* (5-6), 778-784.

78. Hill, M. D., Recent Strategies for the Synthesis of Pyridine Derivatives. *Chemistry-a European Journal* **2010,** *16* (40), 12052-12062.

79. Metz, M.; Piszczatowski, K.; Szalewicz, K., Automatic Generation of Intermolecular Potential Energy Surfaces. *Journal of Chemical Theory and Computation* **2016,** *12* (12), 5895-5919.

80. Tulegenov, A. S.; Wheatley, R. J.; Hodges, M. P.; Harvey, A. H., Intermolecular potential and second virial coefficient of the water-nitrogen complex. *Journal of Chemical Physics* **2007,** *126* (9).

81. Bukowski, R.; Szalewicz, K.; Groenenboom, G.; van der Avoird, A., Interaction potential for water dimer from symmetry-adapted perturbation theory based on density functional description of monomers. *Journal of Chemical Physics* **2006,** *125* (04), 044301.

82. Hodges, M. P.; Wheatley, R. J.; Schenter, G. K.; Harvey, A. H., Intermolecular potential and second virial coefficient of the water-hydrogen complex. *Journal of Chemical Physics* **2004,** *120* (2), 710-720.

83. Hodges, M. P.; Wheatley, R. J.; Harvey, A. H., Intermolecular potential and second virial coefficient of the water-helium complex. *Journal of Chemical Physics* **2002,** *116* (4), 1397-1405.

84. Garberoglio, G.; Jankowski, P.; Szalewicz, K.; Harvey, A. H., All-dimensional H-2-CO potential: Validation with fully quantum second virial coefficients. *Journal of Chemical Physics* **2017,** *146* (5), 13.

85. Donoghue, G.; Wang, X. G.; Dawes, R.; Carrington, T., Computational study of the rovibrational spectra of CO2-C2H2 and CO2-C2D2. *Journal of Molecular Spectroscopy* **2016,** *330*, 170-178.

86. Wang, X. G.; Carrington, T.; Dawes, R., Computational study of the rovibrational spectrum of (CO2)(2). *Journal of Molecular Spectroscopy* **2016,** *330*, 179-187.

87. Klos, J.; Ma, Q. L.; Alexander, M. H.; Dagdigian, P. J., The interaction of NO(X-2 Pi) with H-2: Ab initio potential energy surfaces and bound states. *Journal of Chemical Physics* **2017,** *146* (11), 11.

88. Schewe, H. C.; Ma, Q. L.; Vanhaecke, N.; Wang, X. A.; Klos, J.; Alexander, M. H.; van de Meerakker, S. Y. T.; Meijer, G.; van der Avoird, A.; Dagdigian, P. J., Rotationally inelastic scattering of OH by molecular hydrogen: Theory and experiment. *Journal of Chemical Physics* **2015,** *142* (20), 13.

89. Kirste, M.; Wang, X. A.; Schewe, H. C.; Meijer, G.; Liu, K. P.; van der Avoird, A.; Janssen, L. M. C.; Gubbels, K. B.; Groenenboom, G. C.; van de Meerakker, S. Y. T., Quantum-State Resolved Bimolecular Collisions of Velocity-Controlled OH with NO Radicals. *Science* **2012,** *338* (6110), 1060-1063.

90. Walker, K. M.; Song, L.; Yang, B. H.; Groenenboom, G. C.; van der Avoird, A.; Balakrishnan, N.; Forrey, R. C.; Stancil, P. C., QUANTUM CALCULATION OF INELASTIC CO COLLISIONS WITH H. II. PURE ROTATIONAL QUENCHING OF HIGH ROTATIONAL LEVELS. *Astrophysical Journal* **2015,** *811* (1), 10.

91. Price, S. L.; Reutzel-Edens, S. M., The potential of computed crystal energy landscapes to aid solid-form development. *Drug Discovery Today* **2016,** *21* (6), 912-923.

92. Szalewicz, K., Determination of Structure and Properties of Molecular Crystals from First Principles. *Accounts of Chemical Research* **2014,** *47* (11), 3266-3274.

93. Podeszwa, R.; Rice, B. M.; Szalewicz, K., Crystal structure prediction for cyclotrimethylene trinitramine (RDX) from first principles. *Physical Chemistry Chemical Physics* **2009,** *11* (26), 5512-5518.

94. Klimes, J.; Michaelides, A., Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *The Journal of Chemical Physics* **2012,** *137* (12), 120901-120901.

95. Huang, Y.; Shao, Y.; Beran, G. J., Accelerating MP2C dispersion corrections for dimers and molecular crystals. *Journal of Chemical Physics* **2013,** *138* (22).

96. Van Vleet, M.; Misquitta, A.; Stone, A.; Schmidtt, J., Beyond Born-Mayer: Improved Models for Short-Range Repulsion in ab Initio Force Fields. *Journal of Chemical Theory and Computation* **2016,** *12* (8), 3851-3870.

97. Tafipolsky, M.; Ansorg, K., Toward a Physically Motivated Force Field: Hydrogen Bond Directionality from a Symmetry-Adapted Perturbation Theory Perspective. *Journal of Chemical Theory and Computation* **2016,** *12* (3), 1267-1279.

98. Podeszwa, R.; Szalewicz, K., Communication: Density functional theory overcomes the failure of predicting intermolecular interaction energies. *Journal of Chemical Physics* **2012,** *136* (16), 161102.

99. Misquitta, A. J.; Szalewicz, K., Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers. *Journal of Chemical Physics* **2005,** *122* (21), art-214103.

100. Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K., Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. *Journal of Chemical Physics* **2005,** *123* (21), 214103.

101. Szalewicz, K.; Patkowski, K.; Jeziorski, B., Intermolecular interactions via perturbation theory: From diatoms to biomolecules. In *Intermolecular Forces and Clusters Ii*, Wales, D. J., Ed. Springer-Verlag Berlin: Berlin, 2005; Vol. 116, pp 43-117.

102. Hesselmann, A.; Jansen, G.; Schutz, M., Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *Journal of Chemical Physics* **2005,** *122* (1), 17.

103. Podeszwa, R.; Rice, B. M.; Szalewicz, K., Predicting structure of molecular crystals from first principles. *Physical Review Letters* **2008,** *101* (11).

104. Podeszwa, R.; Bukowski, R.; Rice, B. M.; Szalewicz, K., Potential energy surface for cyclotrimethylene trinitramine dimer from symmetry-adapted perturbation theory. *Physical Chemistry Chemical Physics* **2007,** *9* (41), 5561-5569.

105. Totton, T. S.; Misquitta, A. J.; Kraft, M., A First Principles Development of a General Anisotropic Potential for Polycyclic Aromatic Hydrocarbons. *Journal of Chemical Theory and Computation* **2010,** *6* (3), 683-695.

106. Podeszwa, R.; Bukowski, R.; Szalewicz, K., Potential energy surface for the benzene dimer and perturbational analysis of pi-pi interactions. *Journal of Physical Chemistry A* **2006,** *110* (34), 10345-10354.

107. Misquitta, A. J.; Stone, A. J. *CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form*, University of Cambridge, http://www-stone.ch.cam.ac.uk/programs.html#CamCASP: 2007.

108. Stone, A. J.; Dullweber, A.; Engkvist, O.; Fraschini, E.; Hodges, M. P.; Meredith, A. W.; Nutt, D. R.; Popelier, P. L. A.; Wales, D. J. *ORIENT: a program for studying interactions between molecules*, 4.8.29; University of Cambridge, 2015.

109. Hirshberg, B.; Denekamp, C., First principles prediction of an insensitive high energy density material. *Physical Chemistry Chemical Physics* **2013,** *15* (40), 17681-17688.

110. Hoja, J.; Ko, H.-Y.; Neumann, M. A.; Car, R.; DiStasio, R. A.; Tkatchenko, A., Reliable and practical computational description of molecular crystal polymorphs. *Science Advances* **2019,** *5* (1), eaau-3338.

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

112. Dolgonos, G. A.; Hoja, J.; Boese, A. D., Revised values for the X23 benchmark set of molecular crystals. *Physical Chemistry Chemical Physics* **2019,** *21* (44), 24333-24344.

113. Nyman, J.; Yu, L.; Reutzel-Edens, S. M., Accuracy and reproducibility in crystal structure prediction: the curious case of ROY. *CrystEngComm* **2019,** *21* (13), 2080-2088.

114. Kresse, G.; Furthmuller, J., Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B* **1996,** *54* (16), 11169-11186.

115. Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalleri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D'Arco, P.; Noël, Y.; Causà, M.; Rérat, M.; Kirtman, B., CRYSTAL14: A program for the ab initio investigation of crystalline solids. *International Journal of Quantum Chemistry* **2014,** *114* (19), 1287-1317.

116. Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M., QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics-Condensed Matter* **2009,** *21* (39), 395502.

117. Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. J.; Refson, K.; Payne, M. C., First principles methods using CASTEP. *Zeitschrift fur Kristallographie* **2005,** *220* (5-6), 567-570.

118. Neumann, M. A.; Leusen, F. J. J.; Kendrick, J., A Major Advance in Crystal Structure Prediction. *Angewandte Chemie-International Edition* **2008,** *47* (13), 2427-2430.

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

120.Berland, K.; Cooper, V. R.; Lee, K.; Schröder, E.; Thonhauser, T.; Hyldgaard, P.; Lundqvist, B. I., van der Waals forces in density functional theory: a review of the vdW-DF method. *Reports on Progress in Physics* **2015,** *78* (6), 066501.

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

122.Grimme, S., Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry* **2006,** *27* (15), 1787-1799.

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

124.Becke, A. D.; Johnson, E. R., Exchange-hole dipole moment and the dispersion interaction revisited. *The Journal of Chemical Physics* **2007,** *127* (15), 154108.

125.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.

126.Gutowski, M.; Van Lenthe, J.; Verbeek, J.; Van Duijneveldt, F.; Chałasinski, G., The basis set superposition error in correlated electronic structure calculations. *Chemical Physics Letters* **1986,** *124* (4), 370-375.

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

128.Wei, X.; Zhang, A.; Ma, Y.; Xue, X.; Zhou, J.; Zhu, Y.; Zhang, C., Toward low-sensitive and high-energetic cocrystal III: thermodynamics of energetic–energetic cocrystal formation. *CrystEngComm* **2015,** *17* (47), 9037-9047.

129.Minenkov, Y.; Singstad, Å.; Occhipinti, G.; Jensen, V. R., The accuracy of DFT-optimized geometries of functional transition metal compounds: a validation study of catalysts for olefin metathesis and other reactions in the homogeneous phase. *Dalton Transactions* **2012,** *41* (18), 5526-5541.

130.Beran, G. J. O., Modeling Polymorphic Molecular Crystals with Electronic Structure Theory. *Chemical Reviews* **2016,** *116* (9), 5567-5613.

131.Stewart, E. L.; Foley, C. K.; Allinger, N. L.; Bowen, J. P., Ab-Initio Calculations with Electronic Correlation (Mp2) on the Nucleic-Acid Bases and Their Methyl-Derivatives. *Journal of the American Chemical Society* **1994,** *116* (16), 7282-7286.

132.Velders, G. J. M.; Feil, D., Effect of Electron Correlation on the Electron-Density Distribution and (Hyper)Polarizability of Molecules. *Journal of Physical Chemistry* **1992,** *96* (26), 10725-10735.

133.Sorescu, D. C.; Rice, B. M.; Thompson, D. L., Theoretical studies of the hydrostatic compression of RDX, HMX, HNIW, and PETN crystals. *Journal of Physical Chemistry B* **1999,** *103* (32), 6783-6790.

134.Byrd, E. F. C.; Scuseria, G. E.; Chabalowski, C. F., An ab initio study of solid nitromethane, HMX, RDX, and CL20: Successes and failures of DFT. *Journal of Physical Chemistry B* **2004,** *108* (35), 13100-13106.

135.Marchetti, O.; Werner, H.-J., Accurate Calculations of Intermolecular Interaction Energies Using Explicitly Correlated Coupled Cluster Wave Functions and a Dispersion-Weighted MP2 Method†. *The Journal of Physical Chemistry A* **2009,** *113* (43), 11580-11585.

136.Brandenburg, J. G.; Maas, T.; Grimme, S., Benchmarking DFT and semiempirical methods on structures and lattice energies for ten ice polymorphs. *Journal of Chemical Physics* **2015,** *142* (12).

137.Politzer, P.; Murray, J. S., Detonation Performance and Sensitivity: A Quest for Balance. In *Advances in Quantum Chemistry, Vol 69*, Sabin, J. R., Ed. Elsevier Academic Press Inc: San Diego, 2014; Vol. 69, pp 1-30.

138.Li, H. R.; Shu, Y. J.; Gao, S. J.; Chen, L.; Ma, Q.; Ju, X. H., Easy methods to study the smart energetic TNT/CL-20 co-crystal. *Journal of Molecular Modeling* **2013,** *19* (11), 4909-4917.

139.Bolton, O.; Matzger, A. J., Improved Stability and Smart-Material Functionality Realized in an Energetic Cocrystal. *Angewandte Chemie-International Edition* **2011,** *50* (38), 8960-8963.

140.Davidson, A. J.; Oswald, I. D.; Francis, D. J.; Lennie, A. R.; Marshall, W. G.; Millar, D. I.; Pulham, C. R.; Warren, J. E.; Cumming, A. S., Explosives under pressure—the crystal structure of γ-RDX as determined by high-pressure X-ray and neutron diffraction. *CrystEngComm* **2008,** *10* (2), 162-165.

141.Hakey, P.; Ouellette, W.; Zubieta, J.; Korter, T., Redetermination of cyclo-trimethylenetrinitramine. *Acta Crystallographica Section E: Structure Reports Online* **2008,** *64* (8), o1428-o1428.

142.Millar, D. I.; Oswald, I. D.; Francis, D. J.; Marshall, W. G.; Pulham, C. R.; Cumming, A. S., The crystal structure of β-RDX—an elusive form of an explosive revealed. *Chemical Communications* **2009,** (5), 562-564.

143. Millar, D. I.; Oswald, I. D.; Barry, C.; Francis, D. J.; Marshall, W. G.; Pulham, C. R.; Cumming, A. S., Pressure-cooking of explosives—the crystal structure of ε-RDX as determined by X-ray and neutron diffraction. *Chemical communications* **2010,** *46* (31), 5662-5664.

144. Vrcelj, R. M.; Sherwood, J. N.; Kennedy, A. R.; Gallagher, H. G.; Gelbrich, T., Polymorphism in 2-4-6 Trinitrotoluene. *Crystal growth & design* **2003,** *3* (6), 1027-1032.

145. Ccdc *Mercury*, 3.1; 2013.

146. McAdam, R., Urbanski, T - Chemistry and Technology of Explosives. *Chemistry in Britain* **1967,** *3* (12), 540-&.

147. Cruz-Cabeza, A. J.; Bernstein, J., Conformational Polymorphism. *Chemical Reviews* **2014,** *114* (4), 2170-2191.

148. Cooper, P. W.; Kurowski, S. R., Introduction to the Technology of Explosives, 1996. John Wiley & Sons, Inc., New York, USA.

149. Amelinckx, S., GROWTH FEATURES ON CRYSTALS OF LONG-CHAIN COMPOUNDS .2. *Acta Crystallographica* **1956,** *9* (1), 16-23.

150. Amelinckx, S., GROWTH FEATURES ON CRYSTALS OF LONG-CHAIN COMPOUNDS .3. *Acta Crystallographica* **1956,** *9* (3), 217-224.

151. Amelinckx, S., GROWTH FEATURES ON CRYSTALS OF LONG-CHAIN COMPOUNDS .1. *Acta Crystallographica* **1955,** *8* (9), 530-+.

152. Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J., Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of chemical physics* **1992,** *96* (9), 6796-6806.

153. 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. *Molecular Simulation* **2006,** *32* (12-13), 985-997.

154. 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. *Journal of Physical Chemistry B* **1997,** *101* (5), 798-808.

155. Smith, W.; Forester, T. R., DL_Poly 2.0 - A General Purpose Parallel Molecular Dynamics Simulation Package. *Journal of Molecular Graphics* **1996,** *14* (3), 136-141.

156. Rosen, J. M.; Dickinson, C., Vapor pressures and heats of sublimation of some high-melting organic explosives. *Journal of chemical and Engineering Data* **1969,** *14* (1), 120-124.

157. Olinger, B.; Roof, B.; Cady, H. In *The linear and volume compression of β-HMX and RDX*, Proc. Symposium (Intern.) on High Dynamic Pressures, 1978; pp 3-8.

158. Meents, A.; Dittrich, B.; Johnas, S.; Thome, V.; Weckert, E., Charge-density studies of energetic materials: CL-20 and FOX-7. *Acta Crystallographica Section B: Structural Science* **2008,** *64* (1), 42-49.

159. Evers, J.; Klapötke, T. M.; Mayer, P.; Oehlinger, G.; Welch, J., α-and β-FOX-7, Polymorphs of a High Energy Density Material, Studied by X-ray Single Crystal and Powder Investigations in the Temperature Range from 200 to 423 K. *Inorganic chemistry* **2006,** *45* (13), 4996-5007.

160. Latypov, N. V.; Bergman, J.; Langlet, A.; Wellmar, U.; Bemm, U., Synthesis and reactions of 1, 1-diamino-2, 2-dinitroethylene. *Tetrahedron* **1998,** *54* (38), 11525-11536.

161. Taylor, D. E.; Rob, F.; Rice, B. M.; Podeszwa, R.; Szalewicz, K., A molecular dynamics study of 1, 1-diamino-2, 2-dinitroethylene (FOX-7) crystal using a symmetry adapted perturbation theory-based intermolecular force field. *Physical Chemistry Chemical Physics* **2011,** *13* (37), 16629-16636.

162. Song, H. J.; Zhang, Y. G.; Li, H.; Zhou, T.; Huang, F. L., All-atom, non-empirical, and tailor-made force field for alpha-RDX from first principles. *RSC Advances* **2014,** *4* (76), 40518-40533.

163. Breneman, C. M.; Wiberg, K. B., Determining Atom-Centered Monopoles From Molecular Electrostatic Potentials - The Need For High Sampling Density in Formamide Conformational-Analysis. *Journal of Computational Chemistry* **1990,** *11* (3), 361-373.

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

165. Misquitta, A. J.; Stone, A. J.; Price, S. L., Accurate induction energies for small organic molecules. 2. Development and testing of distributed polarizability models against SAPT(DFT) energies. *Journal of Chemical Theory and Computation* **2008,** *4* (1), 19-32.

166.Uzoh, O. G.; Galek, P. T. A.; Price, S. L., Analysis of the conformational profiles of fenamates shows route towards novel, higher accuracy, force-fields for pharmaceuticals. *Physical Chemistry and Chemical Physics* **2015,** *17*, 7936-7948.

167.Aina, A. A.; Misquitta, A. J.; Price, S. L., From dimers to the solid-state: Distributed intermolecular force-fields for pyridine. *Journal of Chemical Physics* **2017,** *147* (16), 12.

168.Aina, A. A.; Misquitta, A. J.; Phipps, M. J. S.; Price, S. L., Charge Distributions of Nitro Groups Within Organic Explosive Crystals: Effects on Sensitivity and Modeling. *ACS Omega* **2019,** *4* (5), 8614-8625.

169.Andon, R. J. L.; Cox, J. D.; Herington, E. F. G.; Martin, J. F., THE 2ND VIRIAL COEFFICIENTS OF PYRIDINE AND BENZENE, AND CERTAIN OF THEIR METHYL HOMOLOGUES. *Transactions of the Faraday Society* **1957,** *53* (8), 1074-1082.

170.Price, S. L., Anisotropic Intermolecular Potentials for Nitrogen Determined from 2Nd Virial and Shear Viscosity Coefficient Data. *Chemical Physics Letters* **1981,** *79* (3), 553-558.

171.Sadtchenko, V.; Ewing, G. E.; Nutt, D. R.; Stone, A. J., Instability of ice films. *Langmuir* **2002,** *18* (12), 4632-4636.

172.Rudolph, W. W.; Mason, R.; Pye, C. C., Aluminium(III) hydration in aqueous solution. A Raman spectroscopic investigation and an ab initio molecular orbital study of aluminium(III) water clusters. *Physical Chemistry Chemical Physics* **2000,** *2* (22), 5030-5040.

173.Fellers, R. S.; Braly, L. B.; Saykally, R. J.; Leforestier, C., Fully coupled six-dimensional calculations of the water dimer vibration-rotation-tunneling states with split Wigner pseudospectral approach. II. Improvements and tests of additional potentials. *Journal of Chemical Physics* **1999,** *110* (13), 6306-6318.

174.de Matas, M.; Edwards, H. G. M.; Lawson, E. E.; Shields, L.; York, P., FT-Raman spectroscopic investigation of a pseudopolymorphic transition in caffeine hydrate. *Journal of Molecular Structure* **1998,** *440* (1-3), 97-104.

175.Abascal, J. L. F.; Sanz, E.; Fernandez, R. G.; Vega, C., A potential model for the study of ices and amorphous water: TIP4P/Ice. *Journal of Chemical Physics* **2005,** *122* (23).

176.Misquitta, A. J.; Bukowski, R.; Szalewicz, K., Spectra of Ar-CO2 from ab initio potential energy surfaces. *Journal of Chemical Physics* **2000,** *112* (12), 5308-5319.

177.Mas, E. M.; Szalewicz, K.; Bukowski, R.; Jeziorski, B., Pair potential for water from symmetry-adapted perturbation theory. *Journal of Chemical Physics* **1997,** *107* (11), 4207-4218.

178.Gora, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K., Interaction energies of large clusters from many-body expansion. *Journal of Chemical Physics* **2011,** *135* (22).

179.Riera, M.; Lambros, E.; Nguyen, T. T.; Götz, A. W.; Paesani, F., Low-order many-body interactions determine the local structure of liquid water. *Chemical Science* **2019,** *10* (35), 8211-8218.

180.Beran, G. J. O., Approximating quantum many-body intermolecular interactions in molecular clusters using classical polarizable force fields. *Journal of Chemical Physics* **2009,** *130* (16).

181.Kennedy, M. R.; McDonald, A. R.; DePrince, A. E.; Marshall, M. S.; Podeszwa, R.; Sherrill, C. D., Communication: Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory. *Journal of Chemical Physics* **2014,** *140* (12).

182.Engkvist, O.; Hobza, P.; Selzle, H. L.; Schlag, E. W., Benzene trimer and benzene tetramer: Structures and properties determined by the nonempirical model (NEMO) potential calibrated from the CCSD(T) benzene dimer energies. *Journal of Chemical Physics* **1999,** *110* (12), 5758-5762.

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

184.Dove, M. T., *Introduction to Lattice Dynamics*. Cambridge University Press: Cambridge, 1993.

185.Nyman, J.; Day, G. M., Modelling temperature-dependent properties of polymorphic organic molecular crystals. *Physical Chemistry Chemical Physics* **2016,** *18* (45), 31132-31143.

186.Crawford, S.; Kirchner, M. T.; Blaser, D.; Boese, R.; David, W. I. F.; Dawson, A.; Gehrke, A.; Ibberson, R. M.; Marshall, W. G.; Parsons, S.; Yamamuro, O., Isotopic Polymorphism in Pyridine. *Angewandte Chemie-International Edition* **2009,** *48* (4), 755-757.

187.Neumann, M. A.; de Streek, J. V.; Fabbiani, F. P. A.; Hidber, P.; Grassmann, O., Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. *Nature Communications* **2015,** *6*, 7793.

188.Gray, A. E.; Day, G. M.; Leslie, M.; Price, S. L., Dynamics in crystals of rigid organic molecules: contrasting the phonon frequencies calculated by molecular dynamics with harmonic lattice dynamics for imidazole and 5-azauracil. *Molecular Physics* **2004,** *102* (9-10), 1067-1083.

189.Price, S. L.; Willock, D. J.; Leslie, M.; Day, G. M. *DMAREL*, 4.1.1; 2004.

190.Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A., The Relaxation of Molecular Crystal Structures Using a Distributed Multipole Electrostatic Model. *Journal of Computational Chemistry* **1995,** *16* (5), 628-647.

191.Mitchell, J. B. O.; Price, S. L., A systematic nonempirical method of deriving model intermolecular potentials for organic molecules: Application to amides. *Journal of Physical Chemistry A* **2000,** *104* (46), 10958-10971.

192.Mitchell, J. B. O.; Price, S. L.; Leslie, M.; Buttar, D.; Roberts, R. J., Anisotropic repulsion potentials for cyanuric chloride (C3N3Cl3) and their application to modeling the crystal structures of azaaromatic chlorides. *Journal of Physical Chemistry A* **2001,** *105* (43), 9961-9971.

193.DAY, G.; PRICE, S.; LESLIE, M., Atomistic calculations of phonon frequencies and thermodynamic quantities for crystals of rigid organic molecules. *JOURNAL OF PHYSICAL CHEMISTRY B* **2003,** *107* (39), 10919-10933.

194.Day, G. M.; Price, S. L.; Leslie, M., Elastic constant calculations for molecular organic crystals. *Crystal Growth & Design* **2001,** *1* (1), 13-27.

195.Day, G. M.; Price, S. L.; Leslie, M., Atomistic calculations of phonon frequencies and thermodynamic quantities for crystals of rigid organic molecules. *Journal of Physical Chemistry B* **2003,** *107* (39), 10919-10933.

196.Cervinka, C.; Beran, G. J. O., Ab initio prediction of the polymorph phase diagram for crystalline methanol. *Chemical Science* **2018,** *9*, 4622-4629.

197.Hoja, J.; Tkatchenko, A., First-principles stability ranking of molecular crystal polymorphs with the DFT+MBD approach. *Faraday Discussions* **2018**.

198.Guo, R.; Uddin, M. N.; Price, L. S.; Price, S. L., Calculation of Diamagnetic Susceptibility Tensors of Organic Crystals: From Coronene to Pharmaceutical Polymorphs. *The Journal of Physical Chemistry A* **2020,** *124* (7), 1409-1420.

199.Aina, A. A.; Misquitta, A. J.; Price, S. L., From dimers to the solid-state: Distributed intermolecular force-fields for pyridine. *The Journal of Chemical Physics* **2017,** *147* (16), 161722.

200.Misquitta, A. J.; Stone, A. J., ISA-Pol: distributed polarizabilities and dispersion models from a basis-space implementation of the iterated stockholder atoms procedure. *Theoretical Chemistry Accounts* **2018,** *137* (11).

201.Boys, S. F.; Bernardi, F., Calculation of Small Molecular Interactions by Differences of Separate Total Energies - Some Procedures with Reduced Errors. *Molecular Physics* **1970,** *19* (4), 553.

202.Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965,** *140* (4A), A1133-A1138.

203.Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized gradient approximation made simple. *Physical Review Letters* **1996,** *77* (18), 3865-3868.

204.Adamo, C.; Barone, V., Toward reliable density functional methods without adjustable parameters: The PBE0 model. *Journal of Chemical Physics* **1999,** *110* (13), 6158-6170.

205.Perdew, J. P.; Emzerhof, M.; Burke, K., Rationale for mixing exact exchange with density functional approximations. *Journal of Chemical Physics* **1996,** *105* (22), 9982-9985.

206.Perdew, J. P., Density functional theory and the band gap problem. *International Journal of Quantum Chemistry* **1985,** *28* (S19), 497-523.

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

208.Tozer, D. J.; Handy, N. C., Improving virtual Kohn–Sham orbitals and eigenvalues: Application to excitation energies and static polarizabilities. *The Journal of chemical physics* **1998,** *109* (23), 10180-10189.

209.Casida, M. E.; Salahub, D. R., Asymptotic correction approach to improving approximate exchange–correlation potentials: Time-dependent density-functional theory calculations of molecular excitation spectra. *The Journal of Chemical Physics* **2000,** *113* (20), 8918-8935.

210.Hirata, S.; Zhan, C.-G.; Aprà, E.; Windus, T. L.; Dixon, D. A., A new, self-contained asymptotic correction scheme to exchange-correlation potentials for time-dependent density functional theory. *The Journal of Physical Chemistry A* **2003,** *107* (47), 10154-10158.

211.Moellmann, J.; Grimme, S., DFT-D3 Study of Some Molecular Crystals. *The Journal of Physical Chemistry C* **2014,** *118* (14), 7615-7621.

212.Bučko, T.; Lebègue, S.; Hafner, J.; Ángyán, J., Tkatchenko-Scheffler van der Waals correction method with and without self-consistent screening applied to solids. *Physical Review B* **2013,** *87* (6), 064110.

213.Tkatchenko, A.; DiStasio, R. A. J.; Car, R.; Scheffler, M., Accurate and efficient method for many-body van der Waals interactions. *Physical Review Letters* **2012,** *108* (23), 236402-236402.

214.Dunning, T. H., Gaussian-Basis Sets for Use in Correlated Molecular Calculations .1. the Atoms Boron Through Neon and Hydrogen. *Journal of Chemical Physics* **1989,** *90* (2), 1007-1023.

215.Williams, H. L.; Mas, E. M.; Szalewicz, K.; Jeziorski, B., On the Effectiveness of Monomer-Centered, Dimer-Centered, and Bond- Centered Basis Functions in Calculations of Intermolecular Interaction Energies. *Journal of Chemical Physics* **1995,** *103* (17), 7374-7391.

216.Dalgarno, A.; Drake, G., Rayleigh-Schrodinger multiple perturbation theory. *Chemical Physics Letters* **1969,** *3* (6), 349-350.

217.Kumar, M.; Parul, Methods for solving singular perturbation problems arising in science and engineering. *Mathematical and Computer Modelling* **2011,** *54* (1), 556-575.

218.Szalewicz, K., Symmetry-adapted perturbation theory of intermolecular forces. *WIREs Computational Molecular Science* **2012,** *2* (2), 254-272.

219.Jeziorski, B.; Moszynski, R.; Szalewicz, K., Perturbation-Theory Approach to Intermolecular Potential-Energy Surfaces of Van-Der-Waals Complexes. *Chemical Reviews* **1994,** *94* (7), 1887-1930.

220.Williams, H. L.; Chabalowski, C. F., Using Kohn-Sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions. *Journal of Physical Chemistry A* **2001,** *105* (3), 646-659.

221.Misquitta, A. J.; Jeziorski, B.; Szalewicz, K., Dispersion energy from density-functional theory description of monomers. *Physical Review Letters* **2003,** *91* (3), art. no.-033201.

222.Misquitta, A. J.; Szalewicz, K., Intermolecular forces from asymptotically corrected density functional description of monomers. *Chemical Physics Letters* **2002,** *357* (3-4), 301-306.

223.Misquitta, A. J.; Stone, A. J., Ab Initio Atom–Atom Potentials Using CamCASP: Theory and Application to Many-Body Models for the Pyridine Dimer. *Journal of Chemical Theory and Computation* **2016,** *ASAP*.

224.McDaniel, J.; Schmidt, J.; Johnson, M.; Martinez, T., Next-Generation Force Fields from Symmetry-Adapted Perturbation Theory. *Annual Review of Physical Chemistry, Vol 67* **2016,** *67*, 467-488.

225.Moszynski, R.; Heijmen, T. G. A.; Jeziorski, B., Symmetry-adapted perturbation theory for the calculation of Hartree-Fock interaction energies. *Molecular Physics* **1996,** *88* (3), 741-758.

226.Moszynski, R.; Wormer, P. E. S.; Jeziorski, B.; Vanderavoird, A., SYMMETRY-ADAPTED PERTURBATION-THEORY OF NONADDITIVE 3-BODY INTERACTIONS IN VAN-DER-WAALS MOLECULES .1. GENERAL-THEORY. *Journal of Chemical Physics* **1995,** *103* (18), 8058-8074.

227.Misquitta, A. J.; Stone, A. J., Accurate induction energies for small organic molecules: 1. Theory. *Journal of Chemical Theory and Computation* **2008,** *4* (1), 7-18.

228.Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.; Gonthier, J. F.; James, A. M.; McAlexander, H. R.; Kumar, A.; Saitow, M.; Wang, X.; Pritchard, B. P.; Prakash, V.; Schaefer, H. F.; Patkowski, K.; King, R. A.; Valeev, E. F.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, D., PSI4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *Journal of Chemical Theory and Computation* **2017,** *13* (7), 3185-3197.

229.Stone, R. J. A. T. a. A. J., Properties of the regular and irregular solid harmonics. *Journal of Physics A: Mathematical and General* **1977,** *10* (8), 1261.

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

231.Lillestolen, T. C.; Wheatley, R. J., Atomic charge densities generated using an iterative stockholder procedure. *Journal of Chemical Physics* **2009,** *131* (14), 144101.

232.Lillestolen, T. C.; Wheatley, R. J., Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *Chemical Communications* **2008,** *44* (45), 5909-5911.

233.Misquitta, A. J.; Stone, A. J.; Fazeli, F., Distributed Multipoles from a Robust Basis-Space Implementation of the Iterated Stockholder Atoms Procedure. *Journal of Chemical Theory and Computation* **2014,** *10* (12), 5405-5418.

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

235.Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

236.Rice, B. M.; Hare, J. J., A quantum mechanical investigation of the relation between impact sensitivity and the charge distribution in energetic molecules. *Journal of Physical Chemistry A* **2002,** *106* (9), 1770-1783.

237.Politzer, P.; Murray, J. S., Relationships between dissociation energies and electrostatic potentials of C-NO2 bonds: Applications to impact sensitivities. *Journal of Molecular Structure* **1996,** *376*, 419-424.

238.Murray, J. S.; Lane, P.; Brinck, T.; Politzer, P.; Sjoberg, P., Electrostatic Potentials on the Molecular-Surfaces of Cyclic Ureides. *Journal of Physical Chemistry* **1991,** *95* (2), 844-848.

239.Bader, R. F. W.; Larouche, A.; Gatti, C.; Carroll, M. T.; Macdougall, P. J.; Wiberg, K. B., Properties of Atoms in Molecules. Dipole Moments and Transferability of Properties. *Journal of Chemical Physics* **1987,** *87* (2), 1142-1152.

240.Mulliken, R. S., Electronic Population Analysis on Lcao-Mo Molecular Wave Functions .1. *Journal of Chemical Physics* **1955,** *23* (10), 1833-1840.

241.Price, S. L.; Richards, N. G. J., On the Representation of Electrostatic Fields Around Ab Initio Charge Distributions. *Journal of Computer-Aided Molecular Design* **1991,** *5* (1), 41-54.

242.Politzer, P.; Mulliken, R. S., COMPARISON OF 2 ATOMIC CHARGE DEFINITIONS, AS APPLIED TO HYDROGEN FLUORIDE MOLECULE. *Journal of Chemical Physics* **1971,** *55* (10), 5135-&.

243.Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M., The conformational sensitivity of iterative stockholder partitioning schemes. *Chemical Physics Letters* **2012,** *545*, 138-143.

244.Reed, A. E.; Weinstock, R. B.; Weinhold, F., Natural population analysis. *Journal of Chemical Physics* **1985,** *83* (2), 735-746.

245.Wiberg, K. B.; Rablen, P. R., Comparison of Atomic Charges Derived Via Different Procedures. *Journal of Computational Chemistry* **1993,** *14* (12), 1504-1518.

246.Lavery, R.; Etchebest, C.; Pullman, A., Calculation of the molecular electrostatic potential from a multipole expansion based on localized orbitals. *Chemical Physics Letters* **1982,** *85* (3), 266-270.

247.Stone, A. J., Distributed multipole analysis: Stability for large basis sets. *Journal of Chemical Theory and Computation* **2005,** *1* (6), 1128-1132.

248.Hirshfeld, F. L., Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theoretica Chimica Acta* **1977,** *44* (2), 129-138.

249.Ayers, P. W., Atoms in molecules, an axiomatic approach. I. Maximum transferability. *The Journal of Chemical Physics* **2000,** *113* (24), 10886-10898.

250.Matta, C. F.; Bader, R. F. W., An Experimentalist's Reply to “What Is an Atom in a Molecule?”. *The Journal of Physical Chemistry A* **2006,** *110* (19), 6365-6371.

251.Parr, R. G.; Ayers, P. W.; Nalewajski, R. F., What Is an Atom in a Molecule? *The Journal of Physical Chemistry A* **2005,** *109* (17), 3957-3959.

252.Bultinck, P.; van Alsenoy, C.; Ayers, P. W.; Carbo-Dorca, R., Critical analysis and extension of the Hirschfeld atoms in molecules. *Journal of Chemical Physics* **2007,** *126* (14), 144111.

253.Bultinck, P.; Cooper, D. L.; Van Neck, D., Comparison of the Hirshfeld-I and iterated stockholder atoms in molecules schemes. *Physical Chemistry Chemical Physics* **2009,** *11* (18), 3424-3429.

254.Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W., Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development. *Journal of Chemical Theory and Computation* **2016,** *12* (8), 3894-3912.

255.Misquitta, A. J., Charge Transfer from Regularized Symmetry-Adapted Perturbation Theory. *Journal of Chemical Theory and Computation* **2013,** *9* (12), 5313-5326.

256.Deng, S.; Wang, Q.; Ren, P., Estimating and modeling charge transfer from the SAPT induction energy. *Journal of computational chemistry* **2017,** *38* (26), 2222-2231.

257.Wheatley, R. J.; Lillestolen, T. C., Local polarizabilities and dispersion energy coefficients. *Molecular Physics* **2008,** *106* (12-13), 1545-1556.

258.KLAMT, A.; SCHUURMANN, G., COSMO - A NEW APPROACH TO DIELECTRIC SCREENING IN SOLVENTS WITH EXPLICIT EXPRESSIONS FOR THE SCREENING ENERGY AND ITS GRADIENT. *Journal of the Chemical Society-Perkin Transactions 2* **1993,** (5), 799-805.

259.York, D.; Karplus, M., A smooth solvation potential based on the conductor-like screening model. *Journal of Physical Chemistry a* **1999,** *103* (50), 11060-11079.

260.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. *Journal of Chemical Theory and Computation* **2008,** *4* (10), 1795-1805.

261.Piekara, A., Dielectric constant of liquid and solid nitrobenzene. *Nature* **1932,** *130*, 93-93.

262.Talebian, E.; Talebian, M., A general review on the derivation of Clausius-Mossotti relation. *Optik* **2013,** *124* (16), 2324-2326.

263.Thallapally, P. K.; Jetti, R. K. R.; Katz, A. K.; Carrell, H. L.; Singh, K.; Lahiri, K.; Kotha, S.; Boese, R.; Desiraju, G. R., Polymorphism of 1,3,5-trinitrobenzene induced by a trisindane additive. *Angewandte Chemie-International Edition* **2004,** *43* (9), 1149-1155.

264.Stone, A. J.; Misquitta, A. J., Charge-transfer in Symmetry-Adapted Perturbation Theory. *Chemical Physics Letters* **2009,** *473* (1-3), 201-205.

265.Jeziorska, M.; Jeziorski, B.; Cizek, J., DIRECT CALCULATION OF THE HARTREE-FOCK INTERACTION ENERGY VIA EXCHANGE PERTURBATION EXPANSION - THE HE = HE INTERACTION. *International Journal of Quantum Chemistry* **1987,** *32* (2), 149-164.

266.Moszynski, R.; Heijmen, T. G. A.; Jeziorski, B., Symmetry-Adapted Perturbation-Theory For the Calculation of Hartree- Fock Interaction Energies. *Molecular Physics* **1996,** *88* (3), 741-758.

267.Dunitz, J. D.; Gavezzotti, A., How molecules stick together in organic crystals: weak intermolecular interactions. *Chemical Society Reviews* **2009,** *38* (9), 2622-2633.

268.Casimir, H. B. G.; Polder, D., THE INFLUENCE OF RETARDATION ON THE LONDON-VANDERWAALS FORCES. *Physical Review* **1948,** *73* (4), 360-372.

269.McLachlan, A. D., RETARDED DISPERSION FORCES BETWEEN MOLECULES. *Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences* **1963,** *271* (1344), 387-+.

270.Pyzer-Knapp, E. O.; Thompson, H. P. G.; Day, G. M., An optimized intermolecular force field for hydrogen-bonded organic molecular crystals using atomic multipole electrostatics. *Acta Crystallographica Section B-Structural Science Crystal Engineering and Materials* **2016,** *72*, 477-487.

271.Tang, K. T.; Toennies, J. P., An Improved Simple-Model for the Vanderwaals Potential Based on Universal Damping Functions for the Dispersion Coefficients. *Journal of Chemical Physics* **1984,** *80* (8), 3726-3741.

272.Stone, A. J., Electrostatic Damping Functions and the Penetration Energy. *Journal of Physical Chemistry A* **2011,** *115* (25), 7017-7027.

273.Misquitta; Alston, ISA-Pol: Distributed polarizabilities and dispersion models from a basis-space implementation of the iterated stockholder atoms procedure. https://arxiv.org/abs/1806.06737v2, 2018.

274.Hodges, M.; Wheatley, R., Application of the overlap model to calculating correlated exchange energies. *Chemical Physics Letters* **2000,** *326* (3-4), 263-268.

275.Mootz, D.; Wussow, H. G., Crystal-Structures of Pyridine and Pyridine Trihydrate. *Journal of Chemical Physics* **1981,** *75* (3), 1517-1522.

276.Wheatley, P. J., The Crystal and Molecular Structure of Pyridine. *Acta Crystallographica* **1960,** *13*, 80-85.

277. Anghel, A. T.; Day, G. M.; Price, S. L., A study of the known and hypothetical crystal structures of pyridine: why are there four molecules in the asymmetric unit cell? *CrystEngComm* **2002,** *4* (62), 348-355.

278. Funnell, N. P.; Marshall, W. G.; Parsons, S., in preparation.

279. Middleton, B. A.; Partington, J. R., Dipole moment of pyridine. *Nature* **1938,** *141*, 516-517.

280. Podsiadlo, M.; Jakobek, K.; Katrusiak, A., Density, freezing and molecular aggregation in pyridazine, pyridine and benzene. *Crystengcomm* **2010,** *12* (9), 2561-2567.

281. Baker, C.; Grant, G., Modeling aromatic liquids: Toluene, phenol, and pyridine. *Journal of Chemical Theory and Computation* **2007,** *3* (2), 530-548.

282. van de Streek, J.; Neumann, M. A., Crystal-structure prediction of pyridine with four independent molecules. *CrystEngComm* **2011,** *13* (23), 7135-7142.

283. *NEIGHCRYS*, In house built program: University College London, 2010.

284. Gritsenko, O.; Baerends, E. J., Asymptotic correction of the exchange-correlation kernel of time-dependent density functional theory for long-range charge-transfer excitations. *Journal of Chemical Physics* **2004,** *121* (2), 655-660.

285. Misquitta, A. J.; Stone, A. J., Dispersion energies for small organic molecules: first row atoms. *Molecular Physics* **2008,** *106* (12-13), 1631-1643.

286. Misquitta, A. J.; Stone, A. J., Distributed polarizabilities obtained using a constrained density-fitting algorithm. *Journal of Chemical Physics* **2006,** *124* (2), 024111.

287. Williams, G. J.; Stone, A. J., Distributed dispersion: A new approach. *Journal of Chemical Physics* **2003,** *119* (9), 4620-4628.

288. Carter, A. H., *Classical and statistical thermodynamics*. Pearson Education India: 2007.

289. Cox, J. D.; Andon, R. J. L., THE 2ND VIRIAL COEFFICIENTS OF PYRIDINE AND BENZENE AND SOME METHYL HOMOLOGUES BETWEEN 135-DEGREES-C AND 165-DEGREES-C. *Transactions of the Faraday Society* **1958,** *54* (11), 1622-1629.

290. Gray, C. G.; Gubbins, K. E., *Theory of Molecular Fluids*. Clarendon Press: Oxford, 1984.

291. 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. *Journal of Applied Crystallography* **2006,** *39*, 453-457.

292. Gavezzotti, A., Towards a realistic model for the quantitative evaluation of intermolecular potentials and for the rationalization of organic crystal structures. Part II. Crystal energy landscapes. *CrystEngComm* **2003,** *5* (77), 439-446.

293. Giordano, N.; Beavers, C.; Campbell, B.; Eigner, V.; Gregoryanz, E.; Marshall, W.; Pena-Alvarez, M.; Teat, S.; Vennari, C.; Parsons, S., High-pressure polymorphism in pyridine. *Iucrj* **2020,** *7*, 58-70.

294. Heit, Y.; Beran, G., How important is thermal expansion for predicting molecular crystal structures and thermochemistry at finite temperatures? *Acta Crystallographica Section B-Structural Science Crystal Engineering and Materials* **2016,** *72*, 514-529.

295. McCullough, J. P.; Douslin, D. R.; Messerly, J. F.; Hossenlopp, I. A.; Kincheloe, T. C.; Waddington, G., PYRIDINE - EXPERIMENTAL AND CALCULATED CHEMICAL THERMODYNAMIC PROPERTIES BETWEEN 0-DEGREES-K AND 1500-DEGREES-K - A REVISED VIBRATIONAL ASSIGNMENT. *Journal of the American Chemical Society* **1957,** *79* (16), 4289-4295.

296. Buchholz, H. K.; Hylton, R. K.; Brandenburg, J. G.; Seidel-Morgenstern, A.; Lorenz, H.; Stein, M.; Price, S. L., Thermochemistry of Racemic and Enantiopure Organic Crystals for Predicting Enantiomer Separation. *Crystal Growth & Design* **2017,** *17* (9), 4676-4686.

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

298. Nyman, J.; Day, G., Modelling temperature-dependent properties of polymorphic organic molecular crystals. *Physical Chemistry Chemical Physics* **2016,** *18* (45), 31132-31143.

299. Brandenburg, J. G.; Potticary, J.; Sparkes, H. A.; Price, S. L.; Hall, S. R., Thermal Expansion of Carbamazepine: Systematic Crystallographic Measurements Challenge Quantum Chemical Calculations. *Journal of Physical Chemistry Letters* **2017,** *8* (17), 4319-4324.

300.Dunitz, J. D.; Filippini, G.; Gavezzotti, A., Molecular shape and crystal packing: A study of C12H12 isomers, real and imaginary. *Helvetica Chimica Acta* **2000,** *83* (9), 2317-2335.

301.Funnell, N. P.; Dawson, A.; Marshall, W. G.; Parsons, S., Destabilisation of hydrogen bonding and the phase stability of aniline at high pressure. *CrystEngComm* **2013,** *15* (6), 1047-1060.

302.Budd, L.; Ibberson, R.; Marshall, W.; Parsons, S., The effect of temperature and pressure on the crystal structure of piperidine. *Chemistry Central Journal* **2015,** *9*.

303.Fabbiani, F. P. A.; Allan, D. R.; David, W. I. F.; Davidson, A. J.; Lennie, A. R.; Parsons, S.; Pulham, C. R.; Warren, J. E., High-pressure studies of pharmaceuticals: An exploration of the behavior of piracetam. *Crystal Growth & Design* **2007,** *7* (6), 1115-1124.

304.Fabbiani, F. P. A.; Allan, D. R.; Marshall, W. G.; Parsons, S.; Pulham, C. R.; Smith, R. I., High-pressure recrystallisation - a route to new polymorphs and solvates of acetamide and parabanic acid. *Journal of Crystal Growth* **2005,** *275* (1-2), 185-192.

305.Fabbiani, F. P. A.; Allan, D. R.; David, W. I. F.; Moggach, S. A.; Parsons, S.; Pulham, C. R., High-pressure recrystallisation - a route to new polymorphs and solvates. *CrystEngComm* **2004,** *6*, 504-511.

306.Podeszwa, R.; Szalewicz, K., Accurate interaction energies for argon, krypton, and benzene dimers from perturbation theory based on the Kohn-Sham model. *Chemical Physics Letters* **2005,** *412* (4-6), 488-493.

307.Patkowski, K.; Szalewicz, K.; Jeziorski, B., Third-order interactions in symmetry-adapted perturbation theory. *Journal of Chemical Physics* **2006,** *125* (15).

308.Pernal, K.; Podeszwa, R.; Patkowski, K.; Szalewicz, K., Dispersionless Density Functional Theory. *Physical Review Letters* **2009,** *103* (26).

309.Patkowski, K.; Szalewicz, K.; Jeziorski, B., Orbital relaxation and the third-order induction energy in symmetry-adapted perturbation theory. *Theoretical Chemistry Accounts* **2010,** *127* (3), 211-221.

310.Podeszwa, R.; Szalewicz, K., Physical origins of interactions in dimers of polycyclic aromatic hydrocarbons. *Physical Chemistry Chemical Physics* **2008,** *10* (19), 2735-2746.

311.Kennedy, M.; McDonald, A.; DePrince, A.; Marshall, M.; Podeszwa, R.; Sherrill, C., Communication: Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory. *Journal of Chemical Physics* **2014,** *140* (12).

312.Bucar, D. K.; Lancaster, R. W.; Bernstein, J., Disappearing Polymorphs Revisited. *Angewandte Chemie-International Edition* **2015,** *54* (24), 6972-6993.

313.Srirambhatla, V. K.; Guo, R.; Price, S. L.; Florence, A. J., Isomorphous template induced crystallisation: a robust method for the targeted crystallisation of computationally predicted metastable polymorphs. *Chemical Communications* **2016,** *52*, 7384-7386.

314.Bujak, M.; Blaeser, D.; Katrusiak, A.; Boese, R., Conformational polymorphs of 1,1,2,2-tetrachloroethane: pressure vs. temperature. *Chemical Communications* **2011,** *47* (31), 8769-8771.

315.Neumann, M. A.; van de Streek, J.; Fabbiani, F. P. A.; Hidber, P.; Grassmann, O., Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. *Nat Commun* **2015,** *6*.

316.Megiel, E.; Kasprzycka-Guttman, T.; Jagielska, A.; Wroblewska, L., A theoretical and experimental N-14 NMR study of association of pyridine. *Journal of Molecular Structure* **2001,** *569* (1-3), 111-119.

317.Bako, I.; Radnai, T.; Palinkas, G., Investigation of the structure of liquid pyridine: A molecular dynamics simulation, an RISM, and an X-ray diffraction study. *Zeitschrift fur Naturforschung Section A-A Journal of Physical Sciences* **1996,** *51* (7), 859-866.

318.Hofmann, D.; Kuleshova, L., Are Crystal Phase Diagrams Predictable with Force Fields? Case of Benzene Polymorphs. *Crystal Growth & Design* **2014,** *14* (8), 3929-3934.

319.Albaugh, A.; Boateng, H.; Bradshaw, R.; Demerdash, O.; Dziedzic, J.; Mao, Y.; Margul, D.; Swails, J.; Zeng, Q.; Case, D.; Eastman, P.; Wang, L.; Essex, J.; Head-Gordon, M.; Pande, V.; Ponder, J.; Shao, Y.; Skylaris, C.; Todorov, I.; Tuckerman, M.; Head-Gordon, T., Advanced Potential Energy Surfaces for Molecular Simulation. *Journal of Physical Chemistry B* **2016,** *120* (37), 9811-9832.

320.Bereau, T.; Kramer, C.; Meuwly, M., Leveraging Symmetries of Static Atomic Multipole Electrostatics in Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2013,** *9* (12), 5450-5459.

321.Nessler, I.; Litman, J.; Schnieders, M., Toward polarizable AMOEBA thermodynamics at fixed charge efficiency using a dual force field approach: application to organic crystals. *Physical Chemistry Chemical Physics* **2016,** *18* (44), 30313-30322.

322.Brill, T. B.; James, K. J., Thermal-Decomposition of Energetic Materials .61. Perfidy in the Amino-2,4,6-Trinitrobenzene series of Explosives. *Journal of Physical Chemistry* **1993,** *97* (34), 8752-8758.

323.Wilson, W. S.; Bliss, D. E.; Christian, S. L.; Knight, D. J. *Explosive properties of polynitroaromatics*; Naval Weapons Center China Lake CA: 1990.

324.Chen, Z. X.; Xiao, H. M., Quantum Chemistry Derived Criteria for Impact Sensitivity. *Propellants Explosives Pyrotechnics* **2014,** *39* (4), 487-495.

325.Zeman, S.; Friedl, Z., A New Approach to the Application of Molecular Surface Electrostatic Potential in the Study of Detonation. *Propellants Explosives Pyrotechnics* **2012,** *37* (5), 609-613.

326.Owens, F. J., Relationship between Impact Induced Reactivity of Trinitroaromatic Molecules and their Molecular-Structure. *Journal of Molecular Structure-Theochem* **1985,** *22* (MAR), 213-220.

327.Delpuech, A.; Cherville, J., Relation between Shock Sensitiveness of Secondary Explosives and their molecular electronic-structure .2. Nitrate Esters. *Propellants and Explosives* **1979,** *4* (6), 121-128.

328.Delpuech, A.; Cherville, J., Relation between Shock Sensitiveness of Secondary Explosives and their molecular electronic-structure .3. Influence of Crystal Environment. *Propellants and Explosives* **1979,** *4* (3), 61-65.

329.Delpuech, A.; Cherville, J., Relation between Shock Sensitiveness of Secondary Explosives and their molecular electronic-structure .1. Nitroaromatics and Nitramines. *Propellants and Explosives* **1978,** *3* (6), 169-175.

330.Murray, J. S.; Lane, P.; Politzer, P.; Bolduc, P. R., A relationship between Impact Sensitivity and the Electrostatic Potentials at the Midpoints of C-NO2 bonds in Nitroaromatics. *Chemical Physics Letters* **1990,** *168* (2), 135-139.

331.Politzer, P.; Murray, J. S., C-NO2 Dissociation-Energies and Surface Electrostatic Potential Maxima in relation to the Impact Sensitivities of some Nitroheterocyclic Molecules. *Molecular Physics* **1995,** *86* (2), 251-255.

332.Murray, J. S.; Lane, P.; Politzer, P., Effects of strongly electron-attracting components on molecular surface electrostatic potentials: application to predicting impact sensitivities of energetic molecules. *Molecular Physics* **1998,** *93* (2), 187-194.

333.Owens, F. J.; Jayasuriya, K.; Abrahmsen, L.; Politzer, P., Computational Analysis of some properties associated with the Nitro-groups in Polynitroaromatic molecules. *Chemical Physics Letters* **1985,** *116* (5), 434-438.

334.Atalar, T.; Zeman, S., A New View of Relationships of the N-N Bond Dissociation Energies of Cyclic Nitramines. Part I. Relationships with Heats of Fusion. *Journal of Energetic Materials* **2009,** *27* (3), 186-199.

335.Zeman, S.; Liu, N.; Jungová, M.; Hussein, A. K.; Yan, Q.-L., Crystal lattice free volume in a study of initiation reactivity of nitramines: Impact sensitivity. *Defence Technology* **2018,** *14* (2), 93-98.

336.Atalar, T.; Jungova, M.; Zeman, S., A New View of Relationships of the N-N Bond Dissociation Energies of Cyclic Nitramines. Part II. Relationships with Impact Sensitivity. *Journal of Energetic Materials* **2009,** *27* (3), 200-216.

337.Zeman, S.; Atalar, T., A New View of Relationships of the N-N Bond Dissociation Energies of Cyclic Nitramines. Part III. Relationship with Detonation Velocity. *Journal of Energetic Materials* **2009,** *27* (3), 217-229.

338.Bondarchuk, S., Quantification of Impact Sensitivity Based on Solid-State Derived Criteria. *Journal of Physical Chemistry a* **2018,** *122* (24), 5455-5463.

339.Metz, M. P.; Piszczatowski, K.; Szalewicz, K., Automatic Generation of Intermolecular Potential Energy Surfaces. *Journal of Chemical Theory and Computation* **2016,** *12* (12), 5895-5919.

340.Moxnes, J. F.; Jensen, T. L.; Unneberg, E., Study of thermal instability of HMX crystalline polymorphs with and without molecular vacancies using reactive force field molecular dynamics. *Advanced Studies in Theoretical Physics* **2016,** *10* (7), 331-349.

341.Martinez, C.; Iverson, B., Rethinking the term "pi-stacking". *Chemical Science* **2012,** *3* (7), 2191-2201.

342.Storm, C. B.; Stine, J. R.; Kramer, J. F., Sensitivity Relationships in Energetic Materials. *Chemistry and Physics of Energetic Materials* **1990,** *309*, 605-639.

343. Pagoria, P. F.; Lee, G. S.; Mitchell, A. R.; Schmidt, R. D., A review of energetic materials synthesis. *Thermochimica Acta* **2002,** *384* (1-2), 187-204.

344. Kamlet, M. J.; Adolph, H. G., Relationship of Impact Sensitivity with Structure of Organic High Explosives .2. Polynitroaromatic Explosives. *Propellants and Explosives* **1979,** *4* (2), 30-34.

345. Song, X. L.; Wang, Y.; An, C. W.; Guo, X. D.; Li, F. S., Dependence of particle morphology and size on the mechanical sensitivity and thermal stability of octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine. *Journal of Hazardous Materials* **2008,** *159* (2-3), 222-229.

346. Kohno, Y.; Maekawa, K.; Tsuchioka, T.; Hashizume, T.; Imamura, A., A relationship between the impact sensitivity and the electronic-structures for the unique N-N bond in HMX Polymorphs. *Combustion and Flame* **1994,** *96* (4), 343-350.

347. Simpson, R. L.; Urtiew, P. A.; Ornellas, D. L.; Moody, G. L.; Scribner, K. F. J.; Hoffman, D. M., CL-20 performance exceeds that of HMX and its sensitivity is moderate. *Propellants Explosives Pyrotechnics* **1997,** *22* (5), 249-255.

348. Dobratz, B. M. *Ethylenediamine Dinitrate and its Eutectic Mixtures: A Historical Review of the Literature to 1982*; Los Alamos National Lab NM: 1983.

349. Jones, T. E., Role of Inter- and Intramolecular Bonding on Impact Sensitivity. *Journal of Physical Chemistry A* **2012,** *116* (45), 11008-11014.

350. Politzer, P.; Murray, J. S., Some Perspectives on Sensitivity to Initiation of Detonation. In *Green Energetic Materials*, Brinck, T., Ed. Blackwell Science Publ: Oxford, 2014; pp 45-62.

351. Murray, J. S.; Concha, M. C.; Politzer, P., Links between surface electrostatic potentials of energetic molecules, impact sensitivities and C-NO2/N-NO2 bond dissociation energies. *Molecular Physics* **2009,** *107* (1), 89-97.

352. Ostmark, H.; Wallin, S.; Ang, H. G., Vapor Pressure of Explosives: A Critical Review. *Propellants Explosives Pyrotechnics* **2012,** *37* (1), 12-23.

353. Chen, P. Y.; Zhang, L.; Zhu, S. G.; Cheng, G. B., Intermolecular Interactions, Thermodynamic Properties, Detonation Performance, and Sensitivity of TNT/CL-20 Cocrystal Explosive. *Chinese Journal of Structural Chemistry* **2016,** *35* (2), 246-256.

354. Feng, R. Z.; Zhang, S. H.; Ren, F. D.; Gou, R. J.; Gao, L., Theoretical insight into the binding energy and detonation performance of epsilon-, gamma-, beta-CL-20 cocrystals with beta-HMX, FOX-7, and DMF in different molar ratios, as well as electrostatic potential. *Journal of Molecular Modeling* **2016,** *22* (6), 14.

355. Politzer, P.; Murray, J. S., High Performance, Low Sensitivity: Conflicting or Compatible? *Propellants Explosives Pyrotechnics* **2016,** *41* (3), 414-425.

356. Politzer, P.; Murray, J. S., Some molecular/crystalline factors that affect the sensitivities of energetic materials: molecular surface electrostatic potentials, lattice free space and maximum heat of detonation per unit volume. *Journal of Molecular Modeling* **2015,** *21* (2), 11.

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

358. Akopyan, Z.; Struchkov, Y. T.; Dashevskii, V., Crystal and molecular structure of hexanitrobenzene. *Journal of Structural Chemistry* **1966,** *7* (3), 385-392.

359. Allen, F. H.; Motherwell, W. D. S., Applications of the Cambridge Structural Database in organic chemistry and crystal chemistry. *Acta Crystallographica Section B - Structural Science* **2002,** *58* (3), 407-422.

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

361. Allen, F. H.; Taylor, R., Research applications of the Cambridge Structural Database (CSD). *Chemical Society Reviews* **2004,** *33* (8), 463-475.

362. Ccdc *Mercury*, 2004.

363. Allen, F. H.; Kennard, O.; Watson, D. G., Tables of bond lengths determined by x-ray and neutron diffraction. Part 1. Bond lengths in Organic compounds. *Journal of the Chemical Society-Perkin Transactions 2* **1987,** (12), S1-S19.

364. Bader, R. F. W.; Preston, H. J. T., Determination of the charge distribution of methane by a method of density constraints. *Theoretica chimica acta* **1970,** *17* (5), 384-395.

365.Moxnes, J.; Hansen, F.; Jensen, T.; Sele, M.; Unneberg, E., A Computational Study of Density of Some High Energy Molecules. *Propellants Explosives Pyrotechnics* **2017,** *42* (2), 204-212.

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

367.Price, S. L.; Stone, A. J., Electrostatic Models For Polypeptides: Can We Assume Transferability? *Journal of the Chemical Society-Faraday Transactions* **1992,** *88* (13), 1755-1763.

368.Choi, C. S.; Abel, J. E., The Crystal Structure of 1,3,5-Trinitrobenzene by Neutron Diffraction. *Acta Crystallographica Section B - Structural Crystallography and Crystal Chemistry* **1972,** *28* (1), 193-201.

369.Carper, W. R.; Davis, L. P.; Extine, M. W., MOLECULAR-STRUCTURE OF 2,4,6-TRINITROTOLUENE. *Journal of Physical Chemistry* **1982,** *86* (4), 459-462.

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

371.Williams, D. E., Nonbonded Interatomic Potential-Energy Functions and Prediction of Crystal-Structures. *Acta Crystallographica Section A - Foundations of Crystallography* **1984,** *40*, C95-C95.

372.Liu, G.; Gou, R.; Li, H.; Zhang, C., Polymorphism of Energetic Materials: A Comprehensive Study of Molecular Conformers, Crystal Packing, and the Dominance of Their Energetics in Governing the Most Stable Polymorph. *Crystal Growth & Design* **2018,** *18* (7), 4174-4186.

373.Verma, A. R.; Krishna, P., *Polymorphism and polytypism in crystals*. John Wiley and Sons: 1966.

374.Politzer, P.; Lane, P.; Murray, J. S., Electrostatic Potentials, Intralattice Attractive Forces and Crystal Densities of Nitrogen-Rich C,H,N,O Salts. *Crystals* **2016,** *6* (1).

375.Juliano, T.; King, M.; Korter, T., Evaluating London Dispersion Force Corrections in Crystalline Nitroguanidine by Terahertz Spectroscopy. *Ieee Transactions on Terahertz Science and Technology* **2013,** *3* (3), 281-287.

376.Price, S. L., Is zeroth order crystal structure prediction (CSP_0) coming to maturity? What should we aim for in an ideal crystal structure prediction code? *Faraday Discussions* **2018,** *in press*.

377.van Dam, H.; de Jong, W.; Bylaska, E.; Govind, N.; Kowalski, K.; Straatsma, T.; Valiev, M., NWChem: scalable parallel computational chemistry. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2011,** *1* (6), 888-894.

378.Helgaker, T.; Jensen, H. J. A.; Joergensen, P.; Oslen, K.; Ruud, H.; Aagren, A.; Auer, A.; Bak, K.; Bakken, O.; Chuaqui, C. E. *DALTON, a molecular electronic structure program*, 2001.

379.Sadlej, A. J., Medium-Size Polarized Basis-Sets for High-Level Correlated Calculations of Molecular Electric Properties. *Collection of Czechoslovak Chemical Communications* **1988,** *53* (9), 1995-2016.

380.NIST NIST Chemistry WebBook. http://webbook.nist.gov/chemistry/ (accessed June 16, 2016).

381.Grüning, M.; Gritsenko, O. V.; van Gisbergen, S. J. A.; Baerends, E. J., Shape corrections to exchange-correlation potentials by gradient-regulated seamless connection of model potentials for inner and outer region. *The Journal of Chemical Physics* **2000,** *114* (2), 652-660.

382.Casida, M.; Salahub, D., Asymptotic correction approach to improving approximate exchange-correlation potentials: Time-dependent density-functional theory calculations of molecular excitation spectra. *Journal of Chemical Physics* **2000,** *113* (20), 8918-8935.

383.Misquitta, A. J.; Spencer, J.; Stone, A. J.; Alavi, A., Dispersion interactions between semiconducting wires. *Physical Review B* **2010,** *82* (7).

384.Misquitta, A. J.; Stone, A. J., Distributed Dipersion. *Molecular Physics* **2008,** *in preparation*.

385.Misquitta, A.; Stone, A., ISA-Pol: distributed polarizabilities and dispersion models from a basis-space implementation of the iterated stockholder atoms procedure. *Theoretical Chemistry Accounts* **2018,** *137* (11).

386.Hodges, M. P.; Wheatley, R. J., Flexible multipole models for hydrogen fluoride. *Physical Chemistry Chemical Physics* **2000,** *2* (8), 1631-1638.

387.Wales, D. J.; Scheraga, H. A., Review: Chemistry - Global optimization of clusters, crystals, and biomolecules. *Science* **1999,** *285* (5432), 1368-1372.

388.Farkas; Schlegel, H. B., Methods for optimizing large molecules. II. Quadratic search. *Journal of Chemical Physics* **1999,** *111* (24), 10806-10814.

389.Broyden, C. G., The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics* **1970,** *6* (1), 76-90.

390.Eastman, P.; Swails, J.; Chodera, J.; McGibbon, R.; Zhao, Y.; Beauchamp, K.; Wang, L.; Simmonett, A.; Harrigan, M.; Stern, C.; Wiewiora, R.; Brooks, B.; Pande, V., OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *Plos Computational Biology* **2017,** *13* (7).

391.Claus, A.; Becker, H., Ueber Trinitrotoluol und das flüssige Dinitrotoluol. *Berichte der deutschen chemischen Gesellschaft* **1883,** *16* (2), 1596-1598.

392.Radomska, M.; Radomski, R., Calorimetric studies of binary systems of 1,3,5-trinitrobenzene with naphthalene, anthracene and carbazole. I. Phase transitions and heat capacities of the pure components and charge-transfer complexes. *Thermochimica Acta* **1980,** *40* (3), 405-414.

393.Liu, G. R.; Gou, R. J.; Li, H. Z.; Zhang, C. Y., Polymorphism of Energetic Materials: A Comprehensive Study of Molecular Conformers, Crystal Packing, and the Dominance of Their Energetics in Governing the Most Stable Polymorph. *Crystal Growth & Design* **2018,** *18* (7), 4174-4186.

394.Zhang, C. Y.; Jiao, F. B.; Li, H. Z., Crystal Engineering for Creating Low Sensitivity and Highly Energetic Materials. *Crystal Growth & Design* **2018,** *18* (10), 5713-5726.

395.Choi, C. S.; Abel, J. E., The crystal structure of 1,3,5-trinitrobenzene by neutron diffraction. *Acta Crystallographica Section B* **1972,** *28* (1), 193-201.

396.Weissbuch, I.; Leiserowitz, L.; Lahav, M., Tailor-Made and Charge-Transfer Auxiliaries for the Control of the Crystal Polymorphism of Glycine. *Advanced Materials* **1994,** *6* (12), 952-956.

397.Caira, M. R., Crystalline polymorphism of organic compounds. *Design of Organic Solids* **1998,** *198*, 163-208.

398.Caira, M. R., Molecular complexes of sulfonamides. 3. Structure of 5-methoxysulfadiazine (form II) and its 1:1 complex with acetylsalicylic acid. *Journal of Chemical Crystallography* **1994,** *24* (10), 695-701.

399.Staab, E.; Addadi, L.; Leiserowitz, L.; Lahav, M., Control of polymorphism by 'tailor-made' polymeric crystallization auxiliaries. Preferential precipitation of a metastable polar form for second harmonic generation. *Advanced Materials* **1990,** *2* (1), 40-43.

400.Weissbuch, I.; Popovitzbiro, R.; Lahav, M.; Leiserowitz, L., Understanding and Control of Nucleation, Growth, Habit, Dissolution and Structure of 2-Dimensional and 3-Dimensional Crystals Using Tailor-Made Auxiliaries. *Acta Crystallographica Section B - Structural Science* **1995,** *51* (2), 115-148.

401.Mayo, S. L.; Olafson, B. D.; Goddard, W. A., Dreiding - A Generic Force-Field for Molecular Simulations. *Journal of Physical Chemistry* **1990,** *94* (26), 8897-8909.

402.van Dam, H. J. J.; de Jong, W. A.; Bylaska, E.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Valiev, M., NWChem: scalable parallel computational chemistry. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2011,** *1* (6), 888-894.

403.Bernstein, J., Crystal polymorphism. In *Engineering of Crystalline Materials Properties: State of the Art in Modeling, Design and Applications*, Novoa, J. J.; Braga, D.; Addadi, L., Eds. 2008; pp 87-109.

404.Woollam, G. R.; Neumann, M. A.; Wagner, T.; Davey, R. J., The importance of configurational disorder in crystal structure prediction: the case of loratadine. *Faraday Discussions* **2018,** *211*, 209-234.

405.Price, S. L., Why don't we find more polymorphs? *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2013,** *69*, 313-328.

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

407.van Eijck, B. P., Crystal structure predictions for disordered halobenzenes. *Physical Chemistry Chemical Physics* **2002,** *4* (19), 4789-4794.

408.Karamertzanis, P. G., CoCrystal prediction DMAflex. *Crystal Growth & Design* **2008,** *in preparation*.

409.Hollett, J. W., Non-pairwise additivity of the leading-order dispersion energy. *Journal of Chemical Physics* **2015,** *142* (8), 7.

410.Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R., Beyond Born-Mayer: Improved Models for Short-Range Repulsion in ab Initio Force Fields. *Journal of Chemical Theory and Computation* **2016,** *12* (8), 3851-3870.

411.van Eijck, B. P.; Kroon, J., Coulomb energy of polar crystals. *Journal of Physical Chemistry B* **1997,** *101* (6), 1096-1100.

412.Barnett, S. A.; Johnson, A.; Florence, A. J.; Price, S. L.; Tocher, D. A., A systematic experimental and theoretical study of the crystalline state of six chloronitrobenzenes. *Crystal Growth & Design* **2008,** *8* (1), 24-36.

413.Lucaioli, P.; Nauha, E.; Gimondi, I.; Price, L. S.; Guo, R.; Iuzzolino, L.; Singh, I.; Salvalaglio, M.; Price, S. L.; Blagden, N., Serendipitous isolation of a disappearing conformational polymorph of succinic acid challenges computational polymorph prediction. *CrystEngComm* **2018,** *20* (28), 3971-3977.

414.Leslie, M., DL_MULTI - A molecular dynamics program to use distributed multipole electrostatic models to simulate the dynamics of organic crystals. *Molecular Physics* **2008,** *106* (12-13), 1567-1578.

415.Van Vleet, M. J.; Misquitta, A. J.; Schmidt, J. R., New Angles on Standard Force Fields: Toward a General Approach for Treating Atomic-Level Anisotropy. *Journal of Chemical Theory and Computation* **2018,** *14* (2), 739-758.

416.Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schuetz, M., Molpro: a general-purpose quantum chemistry program package. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2012,** *2* (2), 242-253.

417.Kahr, B.; Tang, M.; Ye, H. M.; Shtukenberg, A. G., Polymorphism and Morphology Folklore. *Crystal Growth & Design* **2019,** *19* (11), 5999-6003.

418.Bondarchuk, S. V., Quantification of Impact Sensitivity Based on Solid-State Derived Criteria. *Journal of Physical Chemistry a* **2018,** *122* (24), 5455-5463.

419.Song, X. S.; Cheng, X. L.; Yang, X. D.; He, B., Relationship between the bond dissociation energies and impact sensitivities of some nitro-explosives. *Propellants Explosives Pyrotechnics* **2006,** *31* (4), 306-310.

420.Koch, U.; Stone, A. J., Conformational Dependence of the Molecular Charge Distribution and Its Influence On Intermolecular Interactions. *Journal of the Chemical Society-Faraday Transactions* **1996,** *92* (10), 1701-1708.

421.Reilly, A. M.; Tkatchenko, A., Role of Dispersion Interactions in the Polymorphism and Entropic Stabilization of the Aspirin Crystal. *Physical Review Letters* **2014,** *113* (5).

422.Marom, N.; DiStasio, R. A.; Atalla, V.; Levchenko, S.; Reilly, A. M.; Chelikowsky, J. R.; Leiserowitz, L.; Tkatchenko, A., Many-Body Dispersion Interactions in Molecular Crystal Polymorphism. *Angewandte Chemie-International Edition* **2013,** *52* (26), 6629-6632.

423.Tkatchenko, A.; Alfe, D.; Kim, K. S., First-Principles Modeling of Non-Covalent Interactions in Supramolecular Systems: The Role of Many-Body Effects. *Journal of Chemical Theory and Computation* **2012,** *8* (11), 4317-4322.

424.Podryabinkin, E. V.; Tikhonov, E. V.; Shapeev, A. V.; Oganov, A. R., Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Physical Review B* **2019,** *99* (6), 064114.

425.Ryan, K.; Lengyel, J.; Shatruk, M., Crystal Structure Prediction via Deep Learning. *Journal of the American Chemical Society* **2018,** *140* (32), 10158-10168.

426.Xin, D. Y.; Gonneila, N. C.; He, X. R.; Horspool, K., Solvate Prediction for Pharmaceutical Organic Molecules with Machine Learning. *Crystal Growth & Design* **2019,** *19* (3), 1903-1911.

427.McDonagh, D.; Skylaris, C. K.; Day, G. M., Machine-Learned Fragment-Based Energies for Crystal Structure Prediction. *Journal of Chemical Theory and Computation* **2019,** *15* (4), 2743-2758.

428.Taylor, R.; Wood, P. A., A Million Crystal Structures: The Whole Is Greater than the Sum of Its Parts. *Chemical Reviews* **2019,** *119* (16), 9427-9477.

429.Musil, F.; De, S.; Yang, J.; Campbell, J. E.; Day, G. M.; Ceriotti, M., Machine learning for the structure-energy-property landscapes of molecular crystals. *Chemical Science* **2018,** *9* (5), 1289-1300.

430.Handley, C. M.; Behler, J., Next generation interatomic potentials for condensed systems. *European Physical Journal B* **2014,** *87* (7).

431.Behler, J., Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *Journal of Chemical Physics* **2011,** *134* (7).

432.Li, Y.; Li, H.; Pickard, F. C.; Narayanan, B.; Sen, F. G.; Chan, M. K. Y.; Sankaranarayanan, S. K. R. S.; Brooks, B. R.; Roux, B., Machine Learning Force Field Parameters from Ab Initio Data. *Journal of Chemical Theory and Computation* **2017,** *13* (9), 4492-4503.