

PhD Thesis

Xiaozhou Li

Molecular dynamics with tailor-made force fields for materials science

This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen.

Submitted on: 21 December 2016

Supervisors

Professor Jukka Rantanen

Principal supervisor, December 2015 – December 2016

Co-supervisor, February 2015 – November 2015

Department of Pharmacy, University of Copenhagen, Denmark

Associate Professor Anders Ø. Madsen

Co-supervisor, February 2015 – December 2016

Department of Chemistry, University of Copenhagen, Denmark

Dr Jacco van de Streek*

Principal supervisor, January 2014 – November 2015

Department of Pharmacy, University of Copenhagen, Denmark

Dr Andrew D. Bond[†]

Co-supervisor, January 2014 – January 2015

Department of Pharmacy, University of Copenhagen, Denmark

Assessment committee

Professor Thomas Rades

Department of Pharmacy, University of Copenhagen, Denmark

Professor Thomas Vosegaard

Interdisciplinary Nanoscience Center (iNANO) and Department of Chemistry, Aarhus University, Denmark

Dr Stephan X. M. (Steef) Boerrigter

SSCI, a division of Albany Molecular Research Inc., West Lafayette, Indiana, United States

*Present affiliation: Avant-garde Materials Simulation Deutschland GmbH, Freiburg, Germany

[†]Present affiliation: Department of Chemistry, University of Cambridge, United Kingdom

Preface

The thesis is solely submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen for the degree of Doctor of Philosophy in pharmaceutical sciences. The work was conducted at the Section of Pharmaceutical Technology and Engineering, Department of Pharmacy, University of Copenhagen.

This PhD project was initiated in January, 2014 as part of the project “Changing scales in computational materials science”, funded by the Villum Foundation (project no. VKR023111). The materials simulation package *Materials Studio* (version 6.0) was funded by the Lundbeck Foundation (grant no. R49-A5604). The crystal structure prediction package *GRACE* was provided by Avant-garde Materials Simulation Deutschland GmbH.

The results in this thesis have been published or have been accepted in peer-reviewed journals. They are included in this thesis as Appendices I–IV.

- (I) X. Li, A. D. Bond, K. E. Johanssen & J. van de Streek, *Acta Crystallographica*, 2014, **C70**, 784-789.
- (II) S. D. Gumbert, M. Körbitzer, E. Alig, M. U. Schmidt, M. R. Chierotti, R. Gobetto, X. Li & J. van de Streek, *Dyes and Pigments*, 2016, **131**, 364-372.
- (III) X. Li, L. Tapmeyer, M. Bolte & J. van de Streek, *ChemPhysChem*, 2016, **17**, 2496-2502.
- (IV) X. Li, M. A. Neumann & J. van de Streek, *IUCrJ*, 2017, **4**, DOI: 10.1107/S2052252517001415.

The author has also contributed to the manuscripts listed below. However, they are out of the scope of this thesis and thus are not included:

1. I. Sovago, X. Li, S. W. Larsen, J. van de Streek & A. D. Bond. In preparation.
2. X. Li, J. Rantanen & A. Ø. Madsen. In preparation.

Acknowledgements

I would like to thank my supervisors, Jacco van de Streek, Andrew Bond, Anders Madsen and Jukka Rantanen for delivering exceptional support in all aspects during my PhD. I would like to thank Jacco and Andrew again for providing me the opportunity to carry out my PhD study at the Department of Pharmacy, University of Copenhagen.

I draw special attention to my colleagues in the old “Pharmaceutical Materials Science” group, in particular Kristoffer E. Johansson, Anders S. Larsen and Ioana Sovago, for their collaborations on research projects and assistance on practical issues.

Thanks all the colleagues in the Section of Pharmaceutical Technology and Engineering at the Department of Pharmacy and in the Section of Biophysical and Bioinorganic Chemistry at the Department of Chemistry for creating a great atmosphere for scientific discussions during the group meetings. A special thanks is given to Marianne Jørgensen for providing practical information with regard to the graduate programme. In addition I would like to thank Anders Madsen and Johan Bøtker for helping me translate the English abstract into Danish.

I am grateful for the five-month stay at the Solid State Chemistry group at Radboud University in Nijmegen, the Netherlands. A special thanks is extended to Elias Vlieg, Hugo Meekes, Herma Cuppen, Joost van den Ende and Mireille Smets.

The Villum foundation and the Department of Pharmacy, University of Copenhagen are acknowledged for the financial support of the project. The Lundbeck foundation is acknowledged for funding the *Materials Studio* package. Dr Marcus Neumann and his company Avant-garde Materials Simulation Deutschland GmbH are grateful acknowledged for making the crystal structure prediction package *GRACE* available. I would like to express gratitude to the support teams at the Danish Center for Scientific Computing and SUND-IT for maintaining the computer clusters.

Finally, I would like to thank my family and friends for their kindness, understanding and patience.

Xiaozhou Li
Copenhagen, December 2016

List of abbreviations

API	Active pharmaceutical ingredient
BLYP	Becke exchange functional and Lee-Yang-Parr correlation functional
CCDC	Cambridge Crystallographic Data Centre
CSD	Cambridge Structural Database
CPU	Central processing unit
DFT	Density functional theory
DFT-D	Density functional theory with dispersion-correction
FDA	(U.S.) Food and Drug Administration
GGA	Generalised gradient approximation
GIPAW	Gauge-including projector-augmented wave (method)
LDA	Local density approximation
MAD	Mean absolute deviation
MD	Molecular dynamics
MM	Molecular mechanics
MP2	Second-order Møller-Plesset perturbation theory
PAW	Projector-augmented wave (method)
PBE	Perdew-Burke-Ernzerhof (functional)
PES	Potential energy surface
PIMD	Path integral molecular dynamics
PXRD	Powder X-ray diffraction
P.Y. 138	Pigment Yellow 138
QM	Quantum mechanics
RAHB	Resonance-assisted hydrogen bond
RMSCD	Root-mean-square Cartesian displacement
RMSD	Root-mean-square deviation
SCXRD	Single crystal X-ray diffraction
SS-NMR	Solid-state nuclear magnetic resonance
TMFF	Tailor-made force field

Abstract

In our daily lives, we are surrounded by molecular crystals such as pharmaceuticals, pesticides, optical materials and pigments. Understanding their properties is vital for their optimal design and development. For example, the critical properties of solid-state pharmaceuticals such as bioavailability, solubility and stability are greatly dependent on the three-dimensional packing arrangements of the crystal structures.

The drug development process is typically based on experimental approaches. In recent years, it also receives benefits from computational techniques. Computational simulations are a powerful tool to interpret and predict the structures and properties of molecular crystals at the molecular level. Dispersion-corrected density functional theory (DFT-D) is a robust and reliable electronic structure method which reproduces both the energies and the structural details of molecular crystals within a reasonable time frame. Nevertheless, the practical use of DFT-D is usually limited to energy minimisations for representing molecular crystals at zero Kelvin. In order to consider the dynamic aspects and time, a computationally less-demanding energy function is required. Classical molecular dynamics (MD) simulations, which include time and temperature, conventionally employ transferable force fields such as COMPASS and DREIDING, because these are several orders of magnitude faster than DFT-D calculations, but they are also correspondingly less accurate. The tailor-made force field (TMFF) technique, which was originally developed for crystal structure prediction (CSP) in the *GRACE* software, is an option to overcome such limitations. The force fields are parameterised against DFT-D reference data for individual systems. With such customised force fields, greater accuracy can be achieved at the expense of loss of transferability.

The aim of this thesis is to evaluate the performance of the TMFF technique for MD simulations of crystalline materials. Firstly, benchmarks of existing methods were established; this includes static DFT-D calculations and MD simulations with the COMPASS force field. Secondly, the performance of a TMFF was assessed and compared with established benchmarks. Solid-state NMR (SS-NMR) chemical shift calculations were employed for quantifying the performance of the abovementioned approaches. With the availability of the gauge-including projector-augmented wave (GIPAW) algorithm in popular plane-wave DFT codes such as *CASTEP*, and with ever increasing access to fast and cheap computer hardware, the combination of SS-NMR experiments and calculations for elucidating the structures of molecular crystals is becoming attractive and has developed into the field of “NMR crystallography”.

The benchmark of static DFT-D calculations was established by two organic compounds, including a potential drug and a pigment. Their tautomeric states had been unclear for years

since powder X-ray diffraction-based methods could not discriminate the correct forms. Models of possible tautomers were prepared and subjected to DFT-D energy minimisations and *ab initio* SS-NMR calculations. The combination of these two computational approaches successfully ruled out the incorrect tautomers. Furthermore, the accessibility of the combination enables us to apply it for routine crystal structure validation. One drawback of static DFT-D calculations was that the dynamic effects cannot be represented with static configurations. This gives rise to inconsistency between the calculations and the experiments.

The capabilities of motional averaging with the COMPASS force field were investigated using four ^{13}C SS-NMR calibration compounds. The disordered phases of adamantane and hexamethylbenzene at room temperature were captured in the MD simulations. Together with glycine and 3-methylglutaric acid, the calculated ^{13}C chemical shifts of the four compounds yield a fairly good agreement with the experimental values, with a root-mean-square deviation of 2.0 ppm. The study reveals that the improvement introduced by motional averaging depends on the correct representations on both the atomic positions and the average over space and time, and in consequence on the accuracy of the energy potential applied for the MD simulations. The study also shows that transferable force fields are not always applicable for a wide range of systems; the COMPASS force field (version 2.8) cannot reproduce the conformation of glycine correctly.

Finally, the use of MD simulations with a TMFF for the motional averaging in SS-NMR calculations was studied. Free base cocaine was used as a model system. The TMFF provides better reproduction for the energies and the geometries of DFT-D predicted structures. In contrast, when the COMPASS force field was used, several predicted structures underwent significant conformational changes and phase transitions. With such a better force field like the TMFF, the accuracy of SS-NMR calculations for isotropic chemical shifts was, nevertheless, not significantly improved. Combined with recent investigations, the limitations of using motional averaging in the field of NMR crystallography were revealed. Rather than improving the description of the thermal motion, attention should be shifted to anisotropic chemical shifts and to the development of *ab initio* SS-NMR calculation methods.

In general, the TMFF technique is promising for studies of molecular crystalline systems which require incorporating motional effects. This technique is particularly valuable for different polymorphic forms of the same compound because the possible DFT-D energy minima, which are usually given by a crystal energy landscape, are well-retained during the MD simulations with the TMFF. The usefulness of the TMFF for solid-state pharmaceuticals can be further exploited by investigating systems such as hydrates, salts and even amorphous phases.

Resumé

I vores dagligdag er vi omgivet af molekulære krystaller, såsom lægemidler, pesticider, optiske materialer og pigmenter. Forståelsen af deres egenskaber er vigtig for deres optimale design og udvikling. For eksempel de kritiske egenskaber af faste lægemidler, såsom biotilgængelighed, opløselighed og stabilitet er stærkt afhængige af den tredimensionale pakning i krystalstrukturerne.

Lægemiddeludviklingsprocessen er typisk baseret på eksperimentelle fremgangsmåder. I de seneste år har computationelle teknikker også kunnet hjælpe. Computersimuleringer er et kraftfuldt værktøj til at fortolke og forudsige strukturer og egenskaber af molekulære krystaller på molekulært niveau. Dispersions-korrigered tæthedsfunktionalteori (dispersion-corrected density functional theory, DFT-D) er en kraftig og pålidelig elektronstrukturmetode, som giver både energi og strukturelle detaljer i molekulære krystaller inden for en rimelig tidshorisont. Den praktiske anvendelse af DFT-D er ikke desto mindre begrænset til energiminimeringer, som repræsenterer de molekulære krystaller ved nul grader Kelvin. For at inkludere de dynamiske aspekter og tid, kræves en energifunktion, som koster mindre tid for beregningerne. Klassiske molekulær dynamik simuleringer, som inkluderer tid og temperatur, bruger klassiske kraftfelter såsom COMPASS eller DREIDING, fordi disse er flere størrelsesordener hurtigere end DFT-D beregninger, men de er også tilsvarende mindre nøjagtige. Den skræddersyede kraftfelt-teknik (tailor-made force field, TMFF), som oprindeligt blev udviklet til krystalstrukturforudsigelse (crystal structure prediction, CSP) på GRACE, er en mulighed for at overvinde sådanne begrænsninger. Disse kraftfelter parameterises mod DFT-D referencedata for individuelle systemer. Større nøjagtighed kan således opnås - på bekostning af at tage overførbarheden - med sådanne tilpassede kraftfelter.

Formålet med denne afhandling er at undersøge resultaterne af TMFF-teknikken i forbindelse med MD simuleringer af krystallinske materialer. For det første blev der etableret referencepunkter for de eksisterende metoder; dette omfatter statiske DFT-D beregninger og MD simuleringer med COMPASS kraftfeltet. For det andet blev resultatet af en TMFF vurderet og sammenlignet med de etablerede referencepunkter. Faststof NMR (SS-NMR) kemisk skift beregninger blev brugt til at kvantificere ydeevnen af de ovennævnte metoder. Med tilgængeligheden af gauge-including projector-augmented wave (GIPAW) algoritmen i populær plane-wave DFT koder såsom CASTEP, og med stadigt stigende adgang til hurtig og billig computerhardware, er kombinationen af SS-NMR eksperimenter og beregninger for at belyse strukturerne af molekulære krystaller blevet attraktive og har udviklet sig til feltet NMR krystallografi.

Referencepunkterne for de statiske DFT-D beregninger blev etableret for to organiske

forbindelser, et potentielt lægemiddel og et pigment. Deres tautomere tilstande havde været uklar i ørrevis, fordi pulver røntgendiffraktion-baserede metoder ikke kunne diskriminere de korrekte former. Modeller af mulige tautomerer blev fremstillet og udsat for DFT-D energiminimeringer og *ab initio* SS-NMR-beregninger. Kombinationen af disse to beregningsmæs-sige metoder udelukkede de forkerte tautomerer. Endvidere tillader tilgængeligheden af denne kombination os at bruge den til rutinemæssig validering af krystalstrukturer. En ulempe ved statiske DFT-D beregninger var, at de dynamiske effekter ikke kan repræsenteres med statiske konfigurationer. Dette giver anledning til uoverensstemmelse mellem beregningerne og eksperimenterne.

Mulighederne for at beregne et bevægelsesgennemsnit med COMPASS kraftfeltet blev undersøgt ved hjælp af fire ^{13}C SS-NMR kalibreringsforbindelser. De uordnede faser af adamantan og hexamethylbenzen ved stuetemperatur blev fanget i MD simuleringerne. Sammen med glycine og 3-methylglutarsyre giver de beregnede ^{13}C kemiske skift en forholds-vis god overensstemmelse med de eksperimentelle værdier, med en rmsd (root mean square deviation) på 2,0 ppm. Undersøgelsen viser, at forbedringen opnået ved bevægelsesgen-nemsnittet afhænger af de korrekte repræsentationer på både de atomare positioner og gen-nemsnittet over tid og rum, og som følge heraf på nøjagtigheden af energipotentialet benyt-tet i MD simuleringerne. Undersøgelsen viser også, at overførbare kraftfelter ikke altid kan benyttes til en bred vifte af systemer. COMPASS kraftfeltet (version 2.8) kan således ikke gengive konformationen af glycine korrekt.

Endelig blev brugen af MD simuleringer baseret på TMFF i bevægelsesmidlede SS-NMR beregninger undersøgt. Fribase-kokain blev brugt som modelsystem. TMFF er bedre til at gengive energierne og geometrierne af de med DFT-D forudsagte strukturer. Ved brug MD simuleringer baseret på COMPASS kraftfeltet viste flere forudsagte strukturer signifikante konformationsændringer og faseovergange. Trods det bedre TMFF kraftfelt blev nøjagtighed-en af SS-NMR beregninger af isotropiske kemiske skift alligevel ikke væsentligt forbedret. Kombineret med de seneste undersøgelser afsløredes begrænsninger ved brug af bevægel-sensgennemsnittet inden for NMR krystallografi. Snarere end at forbedre beskrivelsen af termisk bevægelse, bør opmærksomheden rettes mod anisotropiske kemiske skift og mod udviklingen af *ab initio* SS-NMR beregningsmetoder.

Generelt er TMFF teknikken lovende for studier af molekulære krystallinske systemer, der kræver inkorporering af effekten af thermiske bevægelser. Denne teknik er særlig værdifuld for forskellige polymorfe former af den samme forbindelse, fordi de mulige DFT-D energiminima, som normalt gives af et krystal-energilandskab, bevares under MD simu-leringer med TMFF. Nutten af TMFF-er for faste lægemidler kan yderligere udnyttes ved at undersøge systemer såsom hydrater, salte og endda amorfne faser.

Contents

1 Molecular crystals	13
1.1 Crystal structure	13
1.2 Reciprocal space	14
1.2.1 Miller indices	14
1.2.2 Reciprocal lattice	15
1.2.3 Brillouin zone	16
1.3 Polymorphism	17
2 Computational methods	19
2.1 Energy minimisation	19
2.2 Density functional theory	19
2.2.1 Approximations of the exchange-correlation energy	21
2.2.2 Dispersion interactions and correction	22
2.2.3 Bloch's theorem	22
2.2.4 Plane-wave basis sets	23
2.2.5 Pseudopotentials	24
2.2.6 The application of dispersion-corrected DFT for molecular crystals . .	24
2.3 Molecular dynamics	25
2.3.1 Classical molecular dynamics	26
2.3.2 Tailor-made force fields	28
2.3.3 <i>Ab initio</i> MD	28
2.4 The role of <i>ab initio</i> calculations in NMR crystallography	29
2.4.1 NMR crystallography: a brief overview	29
2.4.2 <i>Ab initio</i> calculations of NMR chemical shifts for molecular crystals .	29
2.4.3 Expectations of <i>ab initio</i> SS-NMR calculations	30
3 Hypothesis and aim	31
3.1 Hypothesis of this thesis	31
3.2 Aim of this thesis	31
4 Benchmarking the performance of static DFT-D calculations: Case studies of a potential drug and a pigment	32
4.1 Introduction	32

4.1.1	UKIRAI	32
4.1.2	Pigment Yellow 138	33
4.2	Summary of methods	33
4.2.1	Modelling of the experimental crystal structures	33
4.2.2	DFT-D energy minimisations	36
4.2.3	SS-NMR calculations	37
4.3	Results and discussion	37
4.3.1	UKIRAI	37
4.3.2	P.Y. 138	40
4.4	Conclusions	41
5	Crystallography and dynamics of ^{13}C solid-state NMR standard compounds: DFT-D and molecular dynamics with the COMPASS force field	46
5.1	Introduction	46
5.1.1	Four calibration compounds	46
5.2	Brief summary of methods	47
5.2.1	Polymorph screening and crystal structure determination of 3-methylglutaric acid	47
5.2.2	Modelling of experimental crystal structures and DFT-D energy minimisations	49
5.2.3	Molecular dynamics simulations and SS-NMR calculations	49
5.3	Results and discussion	50
5.3.1	PBE-D2 energy minimisations	50
5.3.2	MD simulations	50
5.3.3	SS-NMR calculations: the importance of accurate motional averaging	52
5.4	Conclusions	54
6	Molecular dynamics with a tailor-made force field for molecular crystals: an <i>ab initio</i> NMR crystallography study	55
6.1	Introduction	55
6.2	Brief summary of methods	56
6.3	Results and discussion	57
6.3.1	Crystal energy landscape	57
6.3.2	The accuracy of the COMPASS force field and the TMFF	57
6.3.3	SS-NMR calculations	59
6.3.4	Discussion	61
6.4	Conclusions	63

CONTENTS

7 Concluding remarks	64
8 Future perspectives	65
9 References	66

Chapter 1

Molecular crystals

For the purpose of this thesis, a crystal is defined as a solid phase that is periodic in three dimensions. Crystals are widespread in our daily lives, ranging from pharmaceuticals and optical materials to pigments and even explosives. The novel and efficient design and development of molecular crystals depends crucially on the understanding of their three-dimensional structures and dynamics. In this chapter, a brief review of the basic concepts in crystallography is provided. The polymorphism phenomenon is discussed with an emphasis on the pharmaceutical industry.

1.1 Crystal structure

The smallest repeating “building block” of a crystal structure is called the unit cell. The three vectors, \mathbf{a} , \mathbf{b} and \mathbf{c} , are basis vectors, which describe the translation of the unit cell in three dimensions to form the entire crystal lattice. The lattice constants of a unit cell, a , b and c , are the lengths of the basis vectors. The angles α , β and γ are called cell angles, which are also defined by the lattice vectors. α is the angle between \mathbf{b} and \mathbf{c} , β is the angle between \mathbf{a} and \mathbf{c} and γ is the angle between \mathbf{a} and \mathbf{b} . An illustration of a unit cell is shown in Figure 1.1.

All crystal structures can be classified in seven crystal systems with restrictions on their geometries. Table 1.1 gives an overview of the crystal systems.

The symmetry of the molecules in a unit cell can be described by one of the 230 space groups. The most commonly used notation systems for space groups are the international

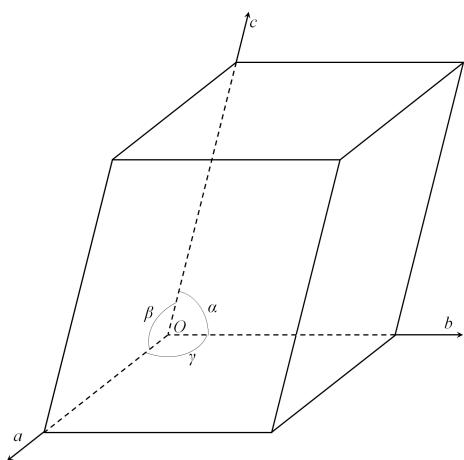


Figure 1.1: An illustration of the unit-cell parameters.

Crystal system	Cell edges	Cell angles
Triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90^\circ$
Monoclinic	$a \neq b \neq c$	$\alpha = \gamma = 90^\circ, \beta \neq 90^\circ$
Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Trigonal	$a = b \neq c$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$
Hexagonal	$a = b \neq c$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$
Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$

Table 1.1: The seven crystal systems.

symbol (Hermann-Mauguin notation) and the international short symbol. In the Hermann-Mauguin notation, the first symbol represents the lattice type and the centring of a unit cell. This includes *P* (primitive), *I* (body-centred), *F* (face-centred), *A*, *B* or *C* (base-centred) and *R* (rhombohedrally-centred). The next three symbols give the symmetry operations along **a**, **b** and **c**, such as inversion centres, *n*-fold axes, screw axes, mirror planes and glide planes. For example, a crystal structure which has space group *P* 1 2₁/*c* 1 (*P*2₁/*c*) has a primitive unit cell (*P*), a screw axis (2₁) along the *b*-axis and a *c*-glide which is perpendicular to the *b*-axis.

The asymmetric unit of a crystal is the smallest part from which a unit cell can be built by applying the space-group symmetry. The symbol *Z'* is conventionally used to describe the number of molecules in the asymmetric unit. The symbol *Z* is used to describe the number of chemical formula units per unit cell. For crystal structure prediction (see Section 2.2.6), *Z''* is used to indicate the total number of chemical fragments in the search space, e.g. *Z''* = 3 for sodium naproxen monohydrate with *Z'* = 1 (and *Z''* = 6 if *Z'* = 2).

1.2 Reciprocal space

1.2.1 Miller indices

In X-ray diffraction experiments, lattice planes reflect incident X-ray beams at certain angles. The diffracted beams are recorded using a detector for further analysis to solve crystal structures. Miller indices *hkl* are used to describe the positions of lattice planes. Consider a plane intersecting **a**, **b** and **c** at (x, 0, 0), (0, y, 0) and (0, 0, z). The Miller indices are calculated based on the reciprocal numbers of the three coordinates, namely

$$\begin{aligned} h &\sim \frac{1}{x} \\ k &\sim \frac{1}{y} \\ l &\sim \frac{1}{z} \end{aligned} \tag{1.1}$$

For example, there is a plane which intersects (3, 0, 0), (0, 2, 0) and (0, 0, 1). We then have

$$h \sim \frac{1}{3}, \quad k \sim \frac{1}{2}, \quad l \sim \frac{1}{1} \tag{1.2}$$

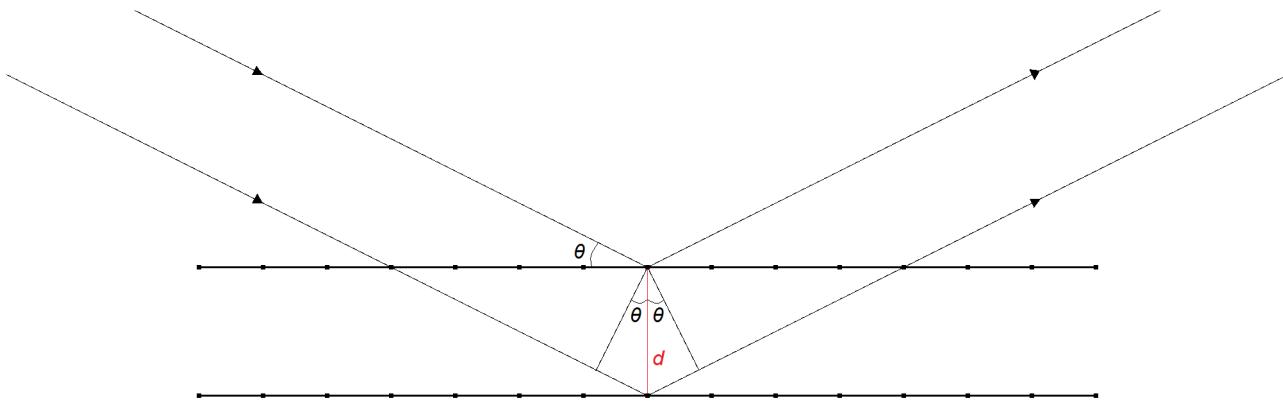


Figure 1.2: An illustration of Bragg's law.

Since Miller indices must be integers, the coordinates are then multiplied by their least common multiple, 6, leading to the Miller indices of the plane, (236).

1.2.2 Reciprocal lattice

Introducing the concept of reciprocal lattice facilitates the description of crystal structures in both X-ray diffraction experiments and *ab initio* plane-wave calculations (see Chapter 2).

Bragg's law

Bragg's law underpins the fundamentals of X-ray diffraction, giving the relationship between the lattice vectors and the diffraction geometry:

$$2d \sin \theta = n\lambda \quad (n = 1, 2, 3, \dots) \quad (1.3)$$

where d is the plane spacing, θ is the diffraction angle between the incident beam and the lattice plane, λ is the wavelength of the X-rays and n is an integer (see Figure 1.2). In practice, n can be eliminated because a reflection with $n > 1$ is equivalent to a reflection with $n = 1$ and $d = d_{n=1}/n$. Equation (1.3) can be rearranged in such a way that the diffraction angle is directly related to the reciprocal of the spacing between lattice planes:

$$\sin \theta = \left(\frac{\lambda}{2}\right) \cdot \left(\frac{1}{d}\right) \quad (1.4)$$

From the direct lattice to the reciprocal lattice

In X-ray diffraction experiments, it is important to index the reflections of (hkl) planes and understand the relation between the d -spacing and (hkl) planes. The planes can be described by a vector \mathbf{d} , which is normal to the planes. The length of \mathbf{d} is the plane spacing. (hkl) planes are obtained based on reciprocals and it is therefore difficult to describe \mathbf{d} using real-space vectors.*

*Real-space, or direct space, is the space in which we observe and describe the physical properties of molecular crystals. Therefore, basis vectors \mathbf{a} , \mathbf{b} and \mathbf{c} are called direct basis vectors.

Herein, \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* are used to represent the reciprocal basis vectors. They are normals of the real-space planes (100), (010) and (001) in geometry, respectively. The mathematical relation between the direct basis vectors \mathbf{a} , \mathbf{b} and \mathbf{c} and their corresponding reciprocal basis vectors \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* is listed below:

$$\begin{pmatrix} \mathbf{a} \cdot \mathbf{a}^* & \mathbf{a} \cdot \mathbf{b}^* & \mathbf{a} \cdot \mathbf{c}^* \\ \mathbf{b} \cdot \mathbf{a}^* & \mathbf{b} \cdot \mathbf{b}^* & \mathbf{b} \cdot \mathbf{c}^* \\ \mathbf{c} \cdot \mathbf{a}^* & \mathbf{c} \cdot \mathbf{b}^* & \mathbf{c} \cdot \mathbf{c}^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.5)$$

The smallest “building block” in reciprocal space is called the reciprocal unit cell. Replicating the reciprocal unit cell in three-dimension in reciprocal space generates the reciprocal lattice.

It can be proved that the reciprocal vector of \mathbf{d} , denoted \mathbf{d}^* , can be easily constructed by \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* :

$$\mathbf{d}^* = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \quad (1.6)$$

The d -spacing can be calculated based on \mathbf{d}^* :

$$\frac{1}{d^2} = \mathbf{d}^* \cdot \mathbf{d}^* = (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) \cdot (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) \quad (1.7)$$

Where $1/d^2$ can be calculated based on Equation (1.4). The wavelength λ is known and θ can be read from the diffractometer. The relation between d -spacing and (hkl) can thus be established via reciprocal space.

1.2.3 Brillouin zone

Wigner-Seitz cell

Consider a unit cell in real space. If there is only one lattice point in a unit cell, it is called a “primitive unit cell”. A Wigner-Seitz cell is a primitive unit cell which is constructed in such a way that any point in the Wigner-Seitz cell is closer to the lattice point of the cell than to any other lattice points. Figure 1.3 shows the construction of a Wigner-Seitz cell. It starts with selecting a lattice point (in red). Lines (in blue) are then connected between the chosen lattice point and the nearest lattice points (black dots). Next, planes are drawn in the middle and perpendicular to the lines. The polyhedron generated by the planes with a lattice point in the centre is a Wigner-Seitz cell.

The first Brillouin zone

The reciprocal unit cell of the Wigner-Seitz cell is called the first Brillouin zone. Its concept is important in plane-wave calculations because the properties of a molecular crystal can be characterised from the first Brillouin zone. Further discussions regarding the computational methods are given in Chapter 2.

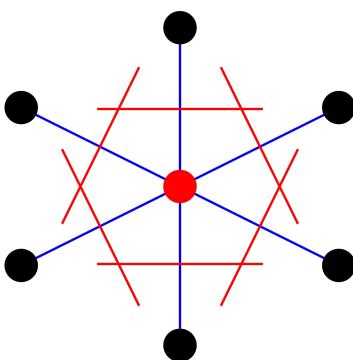


Figure 1.3: An illustration of constructing a Wigner-Seitz cell. Source: https://en.wikipedia.org/wiki/Wigner%20Seitz_cell. Used under the Creative Commons Attribution-Share Alike 3.0 Unported licence.

1.3 Polymorphism

Polymorphism is a common phenomenon for molecular crystals. As defined by McCrone, “a polymorph is a solid crystalline phase of a given compound resulting from the possibility of at least two different arrangements of the molecules of that compound in the solid state”.¹ For solid-state pharmaceuticals, different polymorphic forms of a drug compound result in differences in properties such as density, melting point, stability, solubility and bioavailability. It is therefore of paramount importance to understand and explore the polymorphic landscape and possible phase transitions during the preformulation phase in drug development. Indeed, the Food and Drug Administration (FDA) expects pharmaceutical manufacturers “to have adequate knowledge about drug substance polymorphs”.² Considering different polymorphic forms in the patent portfolio is also important for the intellectual property protection. It should be noted that the polymorphism phenomenon along with other factors such as crystal morphology, particle size, solvation and hydration have an impact on the bulk properties of a drug, such as flowability, compactability and filterability.³

The ritonavir case is perhaps the most infamous crisis as a result of polymorphism. Ritonavir is an antiviral drug developed by Abbott Laboratories.⁴ There was only one form identified, known as Form I, from the drug discovery stage in 1992 to the FDA approval in March, 1996. A less soluble form, Form II, turned up in 1998, which presented a higher solid-state stability and much lower bioavailability.⁵ The packings of the two polymorphs are shown in Figure 1.4. Ritonavir was then temporarily withdrawn from the market. The reformulation was carried out and approved in 2000 by the FDA under a new drug name “Kaletra”; the estimated loss was over 250 million U.S. dollars.⁶

Small molecule drugs predominate the market, although biologics have received considerable attention in recent years. Based on the approvals of new molecular entities (NMEs) and biologics licence applications (BLAs) from FDA, 84% of approvals from 1993 to 2016 are NMEs,^{7,8} as shown in Figure 1.5. Nevertheless, the development of small molecule drugs is becoming challenging. A survey conducted by Benet *et al.*⁹ found that approximately 35% of the top 200 marketed drugs in the U.S. fall in class II (low solubility, high permeability)

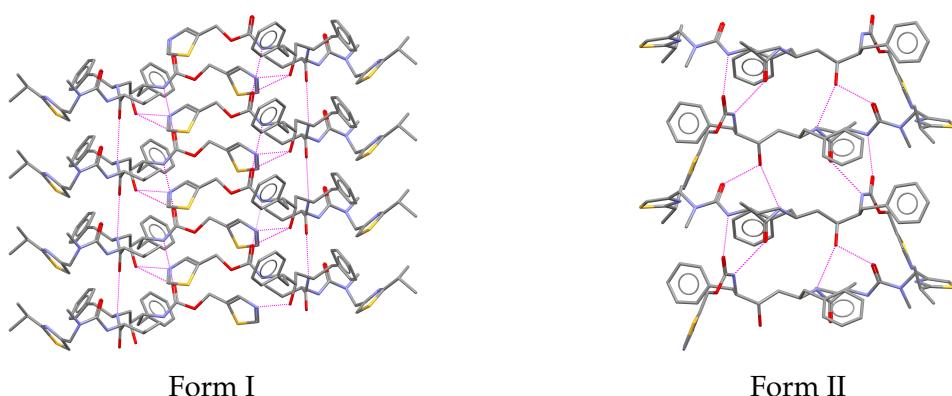


Figure 1.4: Two polymorphs of ritonavir.¹² Hydrogen bonds are indicated in purple dashed lines. Hydrogen atoms are removed for clarity. According to the study of Form II, "... all the strong hydrogen bond donors and acceptors have been satisfied".⁵

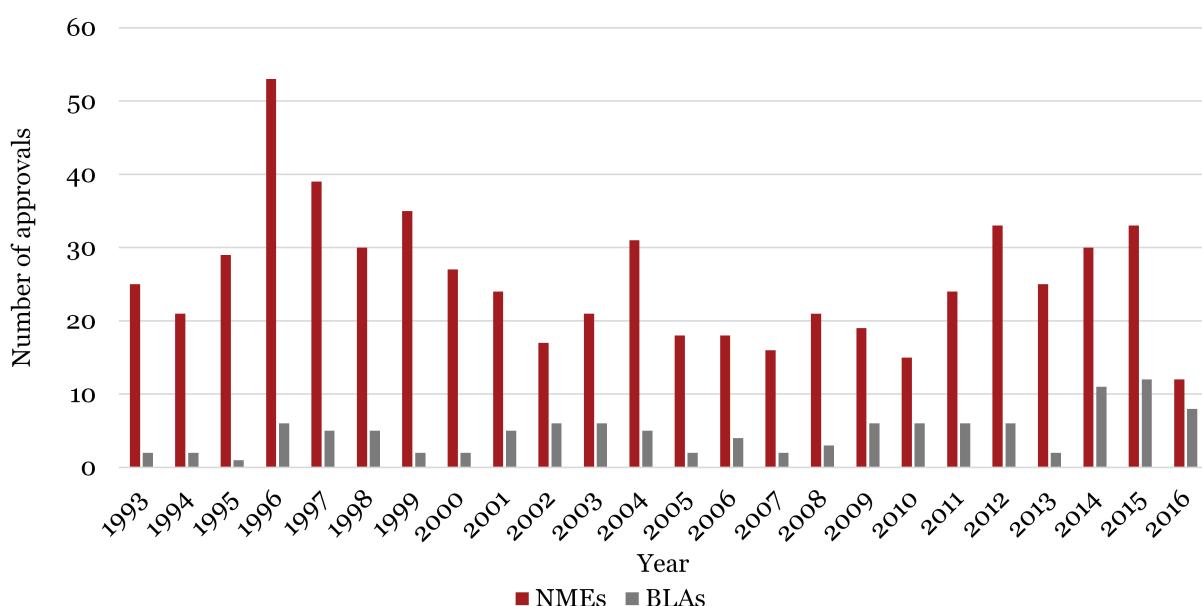


Figure 1.5: FDA approvals of new molecular entities (NMEs) and biologics licence applications (BLAs) from 1993 to 2016 (Date accessed: 19 December 2016).

or class IV (low solubility, low permeability) of the Biopharmaceutics Classification System (BCS);¹⁰ it was estimated that ca. 90% of the NMEs in the pipeline are categorised as class II or class IV. Enormous effort and investment were devoted in order to get a new drug to the market.¹¹ Therefore, accompanying with state-of-the-art experimental techniques, there is an urgent need for better, more efficient and more reliable complements to understand, interpret and predict the structure-property relationship of molecular crystals. This is driven not only by industrial aftermaths such as ritonavir, which were painful trial-and-errors, but also by the demands of future drug development.

Chapter 2

Computational methods

Molecular modelling has had a flourishing development over the past decades. It has been successfully applied in many natural science disciplines, including pharmaceutical sciences. Using computational modelling in the drug discovery stage has a long and fruitful history; on the other hand, the emerging trend of modelling in the drug development stage has been seen in recent years.¹³ Figure 2.1 provides a summary of experimental and computational methods that are applied in different areas in the development of pharmaceuticals. As Rantanen and Khinast demonstrated, one of the trends in future pharmaceutical manufacturing is “experimental techniques should, where possible, be supported by computational approaches”.¹⁴ In this chapter, we review computational methods that are commonly used for the modelling of solid-state pharmaceuticals.

2.1 Energy minimisation

Energy minimisation, also known as geometry optimisation, is a process of finding stationary points on a potential energy surface (PES), i.e. points on which the first derivatives are zero.¹⁵ The points on the PES correspond to arrangements of particles, such as the geometries of molecules.

In this thesis, unless otherwise specified, we only discuss minima, i.e. stationary points on which the second derivatives are all positive.* The minimum with the lowest energy is called *global minimum* (point C in Figure 2.2), whilst the other minima (points A, E and G in Figure 2.2) are called *local minima*.

2.2 Density functional theory

Density functional theory (DFT) is a practical and accurate electronic structure method that has attracted extraordinary attention over the past fifty years.¹⁶ Walter Kohn was awarded the Nobel prize in chemistry in 1998 for “his development of the density functional theory”.¹⁷ Herein, a very brief review of the DFT method is presented.

For quantum systems, the time-independent, non-relativistic Schrödinger equation is:

$$\mathbf{H}\Psi = E\Psi \quad (2.1)$$

*In some cases, the first-order saddle points are of interest. Those points have one negative second derivative in one direction and all positive second derivatives in other directions. Their corresponding geometries are called “transition states”. The topic of transition state is out of the scope of this thesis.

CHAPTER 2. COMPUTATIONAL METHODS

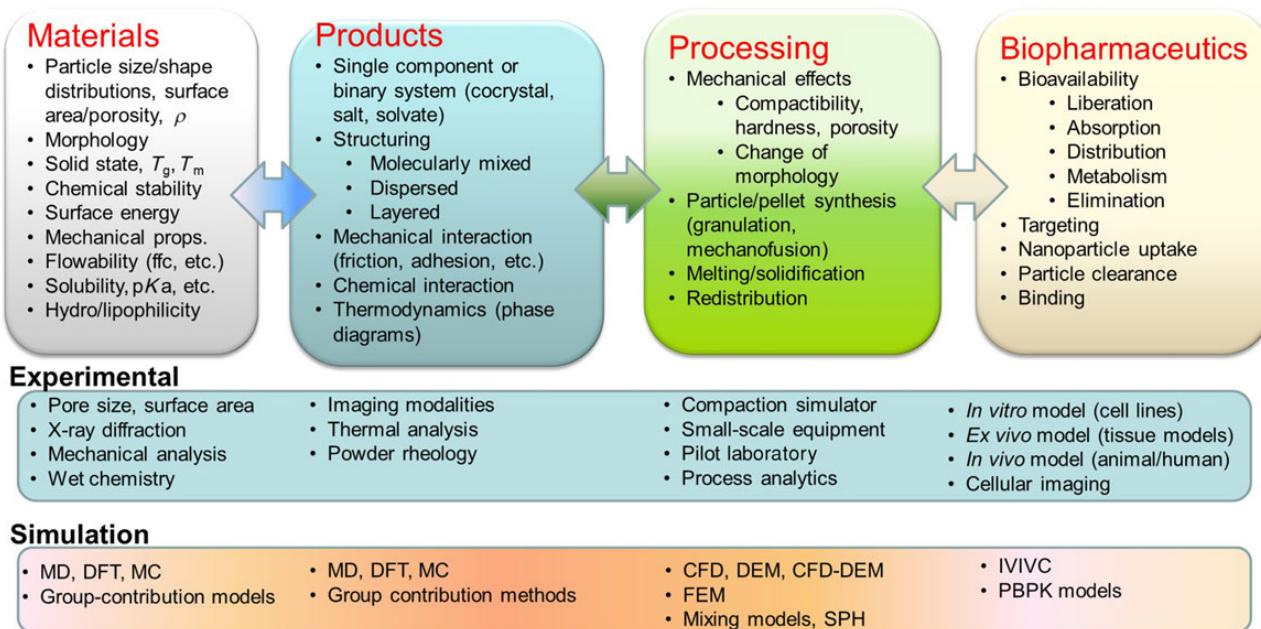


Figure 2.1: The crucial properties and related characterisation methods in different areas of pharmaceutical development. Acronyms (in alphabetical order): CFD, computational fluid dynamics. DEM, discrete element method. DFT: density functional theory. FEM, finite element method. MC, Monte Carlo (simulation). IVIVC, *in vitro-in vivo* correlations. MD, molecular dynamics (simulation). PBPK, physiologically-based pharmacokinetics. SPH, smoothed particle hydrodynamics. *Reprinted from Rantanen and Khinast¹⁴ under the Creative Commons Attribution (CC-BY) 4.0 Licence.*

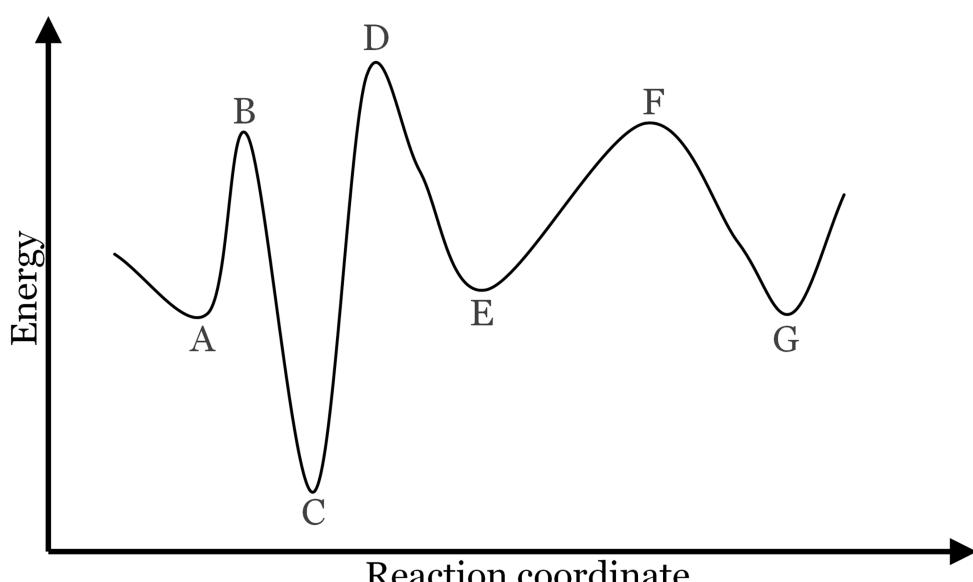


Figure 2.2: An illustration of a potential energy surface.

where \mathbf{H} is the Hamiltonian operator, which characterises the total energy of a wavefunction Ψ . The form of \mathbf{H} depends on the system being described by the Schrödinger equation. In this thesis, we draw attention to the interactions between multiple electrons and multiple nuclei. The Hamiltonian operator in this case contains the kinetic energy operator \mathbf{T} and the potential energy operator \mathbf{V} :

$$\mathbf{H} = \mathbf{T} + \mathbf{V} \quad (2.2)$$

in which

$$\begin{aligned} \mathbf{T} &= - \sum_{i=1}^N \frac{1}{2m_i} \nabla_i^2 \\ \nabla_i^2 &= \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \end{aligned} \quad (2.3)$$

Ψ stands for wavefunctions (sometimes called “state functions”); it can be interpreted as eigenstates (a set of possible solutions) of the Hamiltonian. For each eigenstate, there is an associated eigenvalue, E , which is referred to as the energy of the state and must be a real number.

Within the time-independent Kohn-Sham DFT framework,¹⁸ the ground state energy E_0 of a many-body system is solely a functional (a function of another function) of the electron density, $\rho(\mathbf{r})$:

$$E_0[\rho(\mathbf{r})] = T_s[\rho(\mathbf{r})] + E_{ne}[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + E_{XC}[\rho(\mathbf{r})] \quad (2.4)$$

where $T_s[\rho(\mathbf{r})]$ is the kinetic energy, $E_{ne}[\rho(\mathbf{r})]$ is the nuclear-electron interaction energy and $J[\rho(\mathbf{r})]$ is the energy of electron-electron interactions (i.e. the Coulomb energy). These three terms are known and can be solved. The term $E_{XC}[\rho(\mathbf{r})]$ is the exchange-correlation energy, which is unknown and requires approximations.

2.2.1 Approximations of the exchange-correlation energy

The local density approximation (LDA) is the simplest approximation method for the exchange-correlation energy in which the local electron density is treated as a uniform electron gas. It was proposed along with the Kohn-Sham DFT scheme.¹⁸ The strategy was applied successfully for many systems with slowly varying electron densities; however, it failed to describe certain systems. For example, Becke¹⁹ observed an overbinding tendency for diatomic systems. For noble gas atoms, it has been found that the atomic exchange energies were underestimated by ca. 10% and the atomic correlation energies were overestimated by ca. 100%.¹⁶

The generalised gradient approximation (GGA) is a widely-used improvement over LDA. In GGA, the exchange-correlation energy of a system depends on both the electron density and the gradient of the electron density, which allows the consideration of the non-homogeneity of the true electron density. Examples of popular GGA methods are B88 (correction to the exchange energy of LDA),²⁰ Lee-Yang-Parr (LYP) (correction to the correlation energy of LDA),²¹ Perdew-Wang 1991 (PW91)²² and Perdew-Burke-Ernzerhof (PBE).²³ In

CHAPTER 2. COMPUTATIONAL METHODS

this thesis, the PBE functional was used for most of the cases. Occasionally, the BLYP functional (the combination of B88 and LYP) was used.

Hybrid functionals, which mix the exact exchange from the Hartree-Fock theory with the exchange and correlation from other sources, are extensively used among the DFT community. The concept was first proposed by Becke.²⁴ B3LYP^{21,25–27} is perhaps the most well-known hybrid functional, which is primarily used for non-periodic molecular systems. DFT calculations of periodic solid-state systems are usually carried out using plane-wave DFT codes, in which the use of hybrid functionals are not common due to the demanding computational cost for evaluating the Fock exchange term under the implementation of plane-wave basis sets (see Section 2.2.4).²⁸

2.2.2 Dispersion interactions and correction

The long-range dispersion interactions (van der Waals interactions) have long been recognised as important in the binding of molecular crystals.²⁹ Traditional density functional approximations cannot describe van der Waals interactions correctly.^{30–33} In 2005 Neumann and Perrin³⁴ studied 21 molecular crystals using DFT calculations with the GGA-type PW91 functional.²² They identified that the relative volume of the molecular crystals were overestimated by an average of 20.4%. Corrections for dispersion interactions within the Kohn-Sham DFT scheme were started from 2002 with an empirical treatment by Wu and Yang.³⁵ Grimme has reviewed commonly used dispersion-correction methods in the literature.³⁶

In this thesis, the Grimme-2006 (-D2)³⁷ and the Grimme-2010 (-D3)³⁸ dispersion-correction schemes were used. The choices were made based on both the quality of the schemes^{39,40} and the existing implementations in the utilised simulation packages.

2.2.3 Bloch's theorem

Bloch's theorem enables us to compute the properties of a periodic system, such as a molecular crystal, within a single unit cell.⁴¹ The potential energy of a molecular crystal also shows periodicity:

$$V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r}) \quad (2.5)$$

where \mathbf{R} stands for lattice translation vectors:

$$\mathbf{R} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c} \quad (2.6)$$

where u , v and w are integers and \mathbf{a} , \mathbf{b} and \mathbf{c} are primitive lattice vectors.

Bloch's theorem states that the eigenstates Ψ of the Hamiltonian operator \mathbf{H} (see Equation (2.1)) can be chosen to have the form of a plane wave times a function with the periodicity of the Bravais lattice:

$$\Psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}) \quad (2.7)$$

where \mathbf{k} are Bloch wavevectors that lie within the reciprocal unit cell and $u_{\mathbf{k}}(\mathbf{r})$ is a periodic

function which has the same periodicity as the potential energy V :

$$u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{\mathbf{k}}(\mathbf{r}) \quad (2.8)$$

The functions $e^{i\mathbf{k}\cdot\mathbf{r}}$ are called plane waves (phase factors); calculations which use this strategy are called “plane-wave calculations”. Combining Equation (2.7) and Equation (2.8), we have

$$\Psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\Psi(\mathbf{r}) \quad (2.9)$$

2.2.4 Plane-wave basis sets

In computational chemistry, a basis set refers to a set of functions which are combined linearly to describe molecular orbitals. Plane-wave basis sets are commonly used for solid-state systems due to their inherent periodic nature.

The periodic function $u_{\mathbf{k}}(\mathbf{r})$ in Equation (2.7) can be expanded in a summation of a set of plane waves:

$$u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \quad (2.10)$$

where $c_{\mathbf{G}}$ are Fourier coefficients and \mathbf{G} are reciprocal lattice vectors, which can be represented as:

$$\mathbf{G} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \quad (2.11)$$

where h, k and l are integers and \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* are primitive reciprocal lattice vectors. Combining Equation (2.10) with Equation (2.7) yields

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \quad (2.12)$$

A wavefunction can be expanded in a plane-wave basis set. Based on Equation (2.12), at a point in reciprocal space, the summation is made over infinite \mathbf{G} vectors, which is not possible in reality. We now compare Equation (2.12) with the solution of the kinetic energy of the Schrödinger equation:

$$E = \frac{\hbar^2}{2m} |\mathbf{k} + \mathbf{G}|^2 \quad (2.13)$$

where m is the mass of the particle, \hbar is the reduced Planck constant, or Dirac constant. The relationship between the Planck constant h and \hbar is $\hbar = h/2\pi$. Since the solutions with lower kinetic energies are typically more important than those with higher energies, it makes sense to introduce a cut-off truncating the infinite summation and only consider the solutions which are lower than a certain value:

$$E_{cut} = \frac{\hbar^2}{2m} G_{cut}^2 \quad (2.14)$$

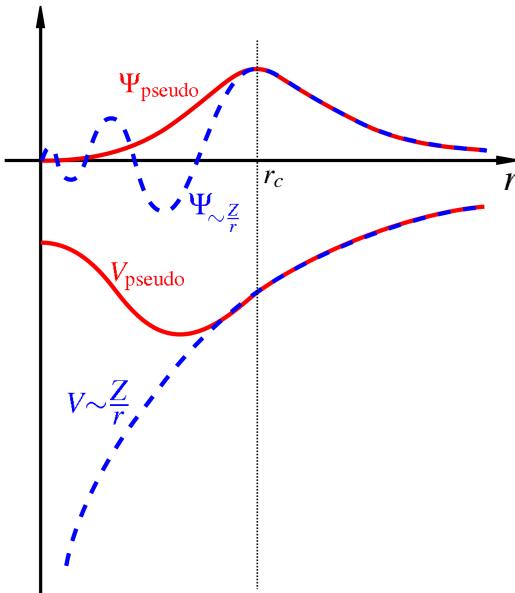


Figure 2.3: An illustration of the idea of pseudopotential. r_c is the core cut-off radius. Source: <https://en.wikipedia.org/wiki/Pseudopotential>.

Combining Equation (2.14) and Equation (2.12) gives the eigenstates as:

$$\Psi_{\mathbf{k}(\mathbf{r})} = \sum_{|\mathbf{k}+\mathbf{G}| < G_{cut}} c_{\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G}) \cdot \mathbf{r}} \quad (2.15)$$

The k -points are sampled in the first Brillouin zone in reciprocal space. In this thesis, the Monkhorst-Pack grid method⁴² was used to sample the k -points.

2.2.5 Pseudopotentials

Most of the chemical properties of a system are dominated by the electrons in the valence region, whereas the core region of an atom, which involves complicated effects of the electron motion, does not have significant impact on the chemical properties. The pseudopotential approach replaces the complicated, rapidly oscillating wavefunction in the core region by a smoothed wavefunction such that the all-electron potential is replaced by an effective potential. It still retains the accurate description in the valence region (Figure 2.3); the electrons in the core region are considered as fixed (also known as the *frozen core approximation*). In this case, the Coulomb term in the Schrödinger equation is substituted by an effective modified pseudopotential term, which is easier to be described by a plane-wave basis set.

In this thesis, the Vanderbilt-type ultrasoft pseudopotentials,⁴³ the norm-conserving pseudopotentials^{44,45} and the on-the-fly generated ultrasoft pseudopotentials⁴⁶ were used.

2.2.6 The application of dispersion-corrected DFT for molecular crystals

As discussed in Section 2.2, the Schrödinger equation computes the total energy of a system, from which other properties such as optical and magnetic properties can be derived. Herein, two applications of DFT-D employed in this thesis were discussed. It should be

noted that DFT, along with other electronic structure methods such as coupled cluster theory and periodic second-order Møller-Plesset perturbation theory (MP2), can be used for predicting phase diagrams,^{47,48} vibrational spectra^{49,50} and NMR spectra (see Section 2.4.1) of molecular crystals.

Crystal structure prediction

Crystal structure prediction (CSP)[†] is a process that can explore possible crystal packings of an organic compound of interest, in particular with the aid of *ab initio* calculations. The input is a two-dimensional (2D) chemical diagram of the compound.⁵¹ The predicted structures are generally ranked based on their thermodynamic stability. The remarkable progress of CSP is evidenced by a series of CSP blind tests organised by the Cambridge Crystallographic Data Centre (CCDC). In the fourth blind test,⁵² one group has successfully captured all four targets in the CSP using *GRACE*,⁵³ which employed DFT-D calculations³⁴ throughout the CSP process. In the fifth⁵⁴ and the sixth⁵⁵ blind tests, *GRACE*,⁵⁶ along with other methods developed by Day *et al.*^{57–63} and Price *et al.*,^{57,60,61,64–66} achieved high success rates.

Furthermore, the development of DFT-D dispersion-correction schemes were revealed in the sixth blind test, as shown by the reranking results from Tkatchenko *et al.* using the many-body dispersion-correction scheme.^{67–70} CSP has already been applied in the industrial development of pharmaceuticals as *ab initio* polymorph screening.^{71,72}

In this thesis, the crystal structure prediction package *GRACE* (Avant-garde Materials Simulation Deutschland GmbH, Freiburg, Germany) was employed to parameterise tailor-made force fields (see Section 2.3.2) and predict crystal structures.

Validation of crystal structures determined from experiments

In 2010, van de Streek and Neumann validated 241 crystal structures solved from single-crystal X-ray diffraction (SCXRD) using DFT-D energy minimisations.⁷³ In 2014, the same DFT-D approach was applied to 215 crystal structures solved from powder X-ray diffraction (PXRD).⁷⁴ These two studies illustrate the robustness of DFT-D to reproduce experimental structures. Moreover, they established criteria for identifying the correctness of experimental crystal structures using root-mean-square Cartesian displacements (RMSCDs) between energy minimised and experimental structures, which are summarised in Table 2.1. Since then, validating experimental crystal structures using DFT-D is becoming a general routine, particularly for structures determined from PXRD.^{75,76} Figure 2.4 illustrates an overlay of a crystal structure, representing the average reproduction of experimental structures using DFT-D.⁷⁴

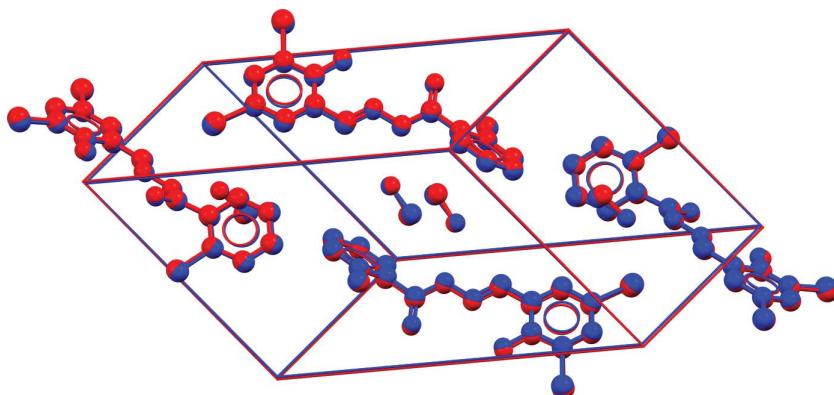
2.3 Molecular dynamics

The DFT-D method has tremendous applications for molecular crystals; however, the method only represents static structures at zero Kelvin. On the other hand, all crystal structures and

[†]In this thesis we only focus the scope of CSP on small organic molecular crystals.

Source of determination	RMSCD	Correctness
Single-crystal X-ray diffraction	$\leq 0.25 \text{ \AA}$	Correct
	$> 0.25 \text{ \AA}$	Incorrect
Powder X-ray diffraction	$\leq 0.30 \text{ \AA}$	Correct
	$0.30 \text{ \AA} < \text{RMSCD} \leq 0.40 \text{ \AA}$	Grey zone
	$> 0.40 \text{ \AA}$	Incorrect

Table 2.1: The criteria of the correctness of a crystal structure determined from experiment.

Figure 2.4: Overlay of an experimentally determined structure (in red) and the DFT-D energy minimised structures (in blue) showing an RMSCD of 0.084 Å. Reprinted from van de Streek and Neumann⁷⁴ under the Creative Commons Attribution (CC-BY) 2.0 Licence.

their properties are determined at a temperature higher than zero Kelvin. A series of molecular crystals exhibit phase transitions or disorder with the change of external conditions such as temperature. For solid-state pharmaceuticals, the conditions during manufacturing and storage such as temperature, pressure, humidity *etc.* are key factors as they may affect the polymorphic form and consequently the bioavailability of a drug. In this case, the dynamics of molecular crystals is important to take into account in computational modelling.

2.3.1 Classical molecular dynamics

According to Frenkel and Smit, molecular dynamics (MD) simulation is “a technique for computing the equilibrium and transport of a classical many-body system”.⁷⁷ The term “classical” means that a system is treated based on classical mechanics. The time evolution of a system is described by Newton’s equation of motion. The potential energy of a system can be described using a set of classical mechanical functional forms and parameters, which is known as the *force field* method. In general, the potential energy has two components: bonded and non-bonded interactions.

$$E_{FF} = E_{bonded} + E_{non-bonded} \quad (2.16)$$

The bonded interactions are usually composed by bond, angle and dihedral interactions. In addition, cross terms, which describe the couplings among bonded interactions (e.g. bond-angle coupling, angle-dihedral coupling and bond-angle-dihedral coupling) are included in

2.3. MOLECULAR DYNAMICS

some force fields. Figure 2.5 gives an overview of bonded and non-bonded interactions.

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral} + E_{cross-terms} \quad (2.17)$$

The non-bonded interactions typically contain electrostatic interactions and van der Waals (vdW) interactions:

$$E_{non-bonded} = E_{electrostatic} + E_{vdW} \quad (2.18)$$

The functional form of the potential energy given by the COMPASS force field⁷⁸ is listed below:

$$E_{COMPASS} = E_{bond} + E_{angle} + E_{dihedral} + E_{cross-terms} + E_{electrostatic} + E_{vdW} \quad (2.19a)$$

$$E_{bond} = \sum_b [k_2(b - b_0)^2 + k_3(b - b_0)^3 + k_4(b - b_0)^4] \quad (2.19b)$$

$$E_{angle} = \sum_\theta [k_2(\theta - \theta_0)^2 + k_3(\theta - \theta_0)^3 + k_4(\theta - \theta_0)^4] \quad (2.19c)$$

$$\begin{aligned} E_{dihedral} &= E_{torsion\ angle} + E_{out-of-plane\ angle} \\ &= \sum_\phi [k_1(1 - \cos\phi) + k_2(1 - \cos 2\phi) + k_3(1 - \cos 3\phi)] + \sum_\chi k_2 \chi^2 \end{aligned} \quad (2.19d)$$

$$\begin{aligned} E_{cross-terms} &= E_{bond-bond} + E_{bond-angle} + E_{bond-torsion} \\ &\quad + E_{angle-torsion} + E_{angle-angle} + E_{angle-angle-torsion} \\ &= \sum_{b,b'} k(b - b_0)(b' - b'_0) + \sum_{b,\theta} k(b - b_0)(\theta - \theta_0) \\ &\quad + \sum_{b,\phi} (b - b_0)[k_1(1 - \cos\phi) + k_2(1 - \cos 2\phi) + k_3(1 - \cos 3\phi)] \\ &\quad + \sum_{\theta,\phi} (\theta - \theta_0)[k_1(1 - \cos\phi) + k_2(1 - \cos 2\phi) + k_3(1 - \cos 3\phi)] \\ &\quad + \sum_{\theta,\theta'} k(\theta - \theta_0)(\theta' - \theta'_0) + \sum_{\theta,\theta',\phi} k(\theta - \theta_0)(\theta' - \theta'_0) \cos\phi \end{aligned} \quad (2.19e)$$

$$E_{electrostatic} = \sum_{i,j} \frac{q_i q_j}{r_{ij}} \quad (2.19f)$$

$$E_{vdW} = \epsilon_{ij} \left[2 \left(\frac{r_{ij}^0}{r_{ij}} \right)^9 - 3 \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 \right] \quad (2.19g)$$

Note that the COMPASS force field uses many anharmonic functional forms to describe atomic interactions. Here we take E_{bond} as an example. The simplest model of E_{bond} can be described by harmonic oscillation:

$$E_{bond} = \frac{1}{2} k(b - b_0)^2 \quad (2.20)$$

The harmonic approximation is sufficient for describing a range of systems. In some cases, anharmonic functional forms are used to improve the accuracy for predicting the properties such as vibrational frequencies. This can be achieved using Taylor expansion, as shown in

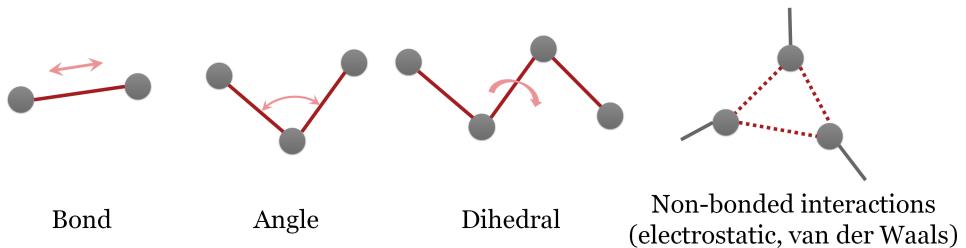


Figure 2.5: An illustration of the force field parameters.

Equation (2.19b).

It is also worthwhile to note that the COMPASS force field uses the Lennard-Jones 9-6 form to describe van der Waals interactions, as seen in Equation (2.19g).

2.3.2 Tailor-made force fields

The successful reproduction and prediction of the dynamics of molecular crystals rely on the accuracy of the potential energies given by the force fields. Traditional transferable force fields such as COMPASS,⁷⁸ AMBER,⁷⁹ OPLS,⁸⁰ CHARMM⁸¹ etc. may be suitable for describing certain systems, but are not always reliable for a wide spectrum of molecules.^{82,83} It is desirable to have an approach that can study the dynamics with high speed yet retain the accuracy.

Tailor-made force fields (TMFFs) are a promising approach to overcome abovementioned limitations of transferable force fields. TMFFs are parameterised against DFT-D reference data. In other words, TMFF parameters represent the potential energies given by DFT-D calculations. The TMFF technique was originally developed for generating possible crystal structures (usually thousands) in *GRACE*. The remarkable performance of TMFFs has been seen in a series of CCDC blind tests (see Section 2.2.6). Details of the functional form and the parameterisation protocol of TMFFs are documented in the literature.⁸⁴ Using the latest version of *GRACE*, TMFFs can be exported to third-party MD simulation packages such as *GROMACS* (in the AMBER⁷⁹ format) and *Materials Studio* (in the DREIDING⁸⁵ format). This expands the useability of TMFFs to the dynamics of, in principle, every molecular crystal.

2.3.3 *Ab initio* MD

We mention *ab initio* MD in this part for completeness, although it was not used in this thesis. In classical MD simulations, the forces are calculated based on the force field method (see Section 2.3.1). In *ab initio* MD simulations, the forces acting on a system of particles are calculated at each time step (on-the-fly) using electronic structure methods to evolve the dynamics of the system in time.⁸⁶ Introducing *ab initio* calculations increases both the accuracy and computational cost of the *ab initio* MD technique. For molecular crystals, the time scale of an *ab initio* MD simulation is usually a few picoseconds and the size of the simulation cell is generally limited to $1 \times 1 \times 1$ (up to ca. 200 atoms). In contrast, classical MD can easily cope with a system which has 10^5 atoms; the simulation time is typically in the order of nanoseconds.

2.4 The role of *ab initio* calculations in NMR crystallography

2.4.1 NMR crystallography: a brief overview

The nuclear magnetic resonance (NMR) crystallography method uses solid-state NMR (SS-NMR) spectroscopy and *ab initio* calculations, often in combination with PXRD, to elucidate the structures and dynamics of solid-state materials.⁸⁷ The chemical shifts in SS-NMR spectra are sensitive to the local environments, which allows us to detect subtle differences and changes in molecular crystals. SS-NMR spectroscopy has been applied for studying polymorphs,⁸⁸ crystallisation process,⁸⁹ disorder⁹⁰ and so forth.

2.4.2 *Ab initio* calculations of NMR chemical shifts for molecular crystals

There are challenging cases in which the SS-NMR chemical shifts are difficult to interpret, such as the ambiguity in the assignment of chemical shifts due to broadened lineshapes or sophisticated chemical environments. *Ab initio* calculations of SS-NMR chemical shifts provide guidance for understanding the experimental data.

The effective Hamiltonian of a spin 1/2 nucleus is:

$$\mathbf{H} = - \sum_K \gamma_K \mathbf{I}_K (1 - \overleftrightarrow{\sigma}_K) \mathbf{B}_{ext} \quad (2.21)$$

Where γ_K is the magnetogyric ratio of the nucleus K , \mathbf{I}_K is the spin angular momentum and \mathbf{B}_{ext} is an externally applied magnetic field.

The electrons then respond to the applied magnetic field, \mathbf{B}_{ext} , which gives rise to orbital currents $\mathbf{j}(\mathbf{r}')$ and consequently an induced magnetic field, \mathbf{B}_{in} :

$$\mathbf{B}_{in}(\mathbf{R}_K) = - \overleftrightarrow{\sigma}_K \mathbf{B}_{ext} \quad (2.22)$$

Where \mathbf{R}_K is the position of the nucleus. The isotropic magnetic shielding is given by:

$$\sigma_{iso} = \frac{Tr(\overleftrightarrow{\sigma}_K)}{3} \quad (2.23)$$

For a diamagnet,

$$\mathbf{B}_{in}(\mathbf{r}) = \frac{1}{c} \int d^3 r' \mathbf{j}(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \quad (2.24)$$

Where c is the speed of light in vacuum. Therefore, calculating the magnetic shielding is a matter of calculating the electronic currents induced by an applied magnetic field. This can be achieved using perturbation theory because the electronic changes by an applied field are known to be small.⁹¹

As discussed in Section 2.2.5, the pseudopotential approach adopts the frozen core approximation. For the calculations of NMR parameters, the electronic structure in the core region of a nucleus is important. The introduction of the projector augmented wave (PAW)

	RMSD [ppm]	Standard deviation (σ_{SD}) of RMSD [ppm]
^{13}C	1.9	0.4
^1H	0.33	0.16

Table 2.2: The expectations of the RMSDs between the DFT-GIPAW calculated and the experimental chemical shifts for ^{13}C and ^1H .

method⁹² allows rebuilding the all-electron wavefunction, Ψ_{AE} , based on the pseudised wavefunction, Ψ_{PP} , using a transformation operator, \mathbf{T} .

$$\Psi_{AE}(\mathbf{r}) = \mathbf{T}\Psi_{PP}(\mathbf{r}) \quad (2.25)$$

The gauge-including projector augmented wave (GIPAW) method⁹³ extended the formalism of the PAW approach, which enables the calculations of NMR parameters using DFT. Details of calculating the induced currents (see Equation 2.24) are described in the literature.⁹³

The GIPAW method computes absolute shielding constants, σ , which need to be converted to chemical shifts (δ) to compare with experimental SS-NMR spectra. These two quantities are related by

$$\delta = \frac{\sigma_{ref} - \sigma}{1 - \sigma_{ref}} \quad (2.26)$$

where σ_{ref} is the reference shielding. For ^1H and ^{13}C chemical shifts, $|\sigma_{ref}| \ll 1$, Equation (2.26) is then approximated to a linear function

$$\delta \approx \sigma_{ref} - \sigma \quad (2.27)$$

Based on Equation (2.27), σ_{ref} can be obtained by fitting the calculated shielding constants against experimental chemical shifts in a linear function with the slope constrained to -1. The extrapolation to zero shift gives σ_{ref} .

2.4.3 Expectations of *ab initio* SS-NMR calculations

The expectations of the root-mean-square deviations (RMSDs) between the DFT-GIPAW calculated and the experimental chemical shifts for ^{13}C and ^1H are listed in Table 2.2. The data were obtained based on a study for ca. 30 organic compounds.^{94,95}

Chapter 3

Hypothesis and aim

3.1 Hypothesis of this thesis

Tailor-made force fields (TMFFs), which are parameterised against dispersion-corrected density functional theory (DFT-D) reference data for individual molecules, are more accurate than transferable force fields in describing the dynamics of molecular crystals.

3.2 Aim of this thesis

This thesis had a clear aim to evaluate the performance of TMFFs for molecular crystals. In order to do so, benchmarks of existing methods were established and compared with TMFFs. SS-NMR chemical shift calculations were used in the project to quantify the performance of static DFT-D calculations and motional averaging with the COMPASS force field and a TMFF. This was done by measuring the deviations between the calculated and the experimental chemical shifts. More specifically, the project was split into the sub-objectives listed below.

- Assess the performance of static DFT-D calculations. A potential drug compound and a pigment are used as examples.
- Examine the performance of a transferable force field, COMPASS,⁷⁸ in the motional averaging for four ¹³C SS-NMR calibration compounds.
- Evaluate the performance of static DFT-D, motional averaging with the COMPASS force field and motional averaging with a TMFF for a drug compound.

Chapter 4

Benchmarking the performance of static DFT-D calculations: Case studies of a potential drug and a pigment

This chapter is based on Appendix I “Distinguishing tautomerism in the crystal of (Z)-N-(5-ethyl-2,3-dihydro-1,3,4-thiadiazol-2-ylidene)-4-methylbenzenesulfonamide using DFT-D calculations and ^{13}C solid-state NMR” and Appendix II “Crystal structure and tautomerism of Pigment Yellow 138 determined by X-ray powder diffraction and solid-state NMR”.

4.1 Introduction

X-ray diffraction-based approaches are commonly used to determine crystal structures. PXRD is an alternative when obtaining single crystals is challenging. One drawback of the PXRD technique is locating hydrogen atoms due to their low X-ray scattering power. Hydrogen atoms, on the other hand, have a crucial role in molecular crystals, particularly in the formation of hydrogen bonds. Therefore, additional validation approaches such as *ab initio* calculations are desired to solve a crystal structure including the positions of hydrogen atoms.

Prototropic tautomerism is the transfer of a proton between two atoms that can occur spontaneously.⁹⁶ The tautomeric state has a strong impact on the properties of molecular crystals such as chemical stability, melting point and colour.⁹⁷ If the tautomeric state can be understood and controlled properly, it will facilitate the development of novel functional materials or pharmaceuticals. For example, drug compounds which have ring-closed intramolecular hydrogen bonds show higher lipophilicity than that of the ring-open forms, which is important for the absorption in the gastrointestinal tract.⁹⁸

In this chapter, two tautomeric compounds were revisited, in which the protonation states had been unclear for years. PXRD, SS-NMR experiments and static DFT-D calculations were combined to study the crystal structures of these two compounds at the atomic level; the combination fits the scope of “NMR crystallography”.⁸⁷

4.1.1 UKIRAI

The first case is *N*-(5-ethyl-[1,3,4]-thiadiazole-2-yl)toluenesulfonamide (CSD reference code: UKIRAI).⁹⁹ This compound is referred to as UKIRAI in the following context for simplification. UKIRAI contains an *N*-substituted heterocyclic sulfonamide group, which is found

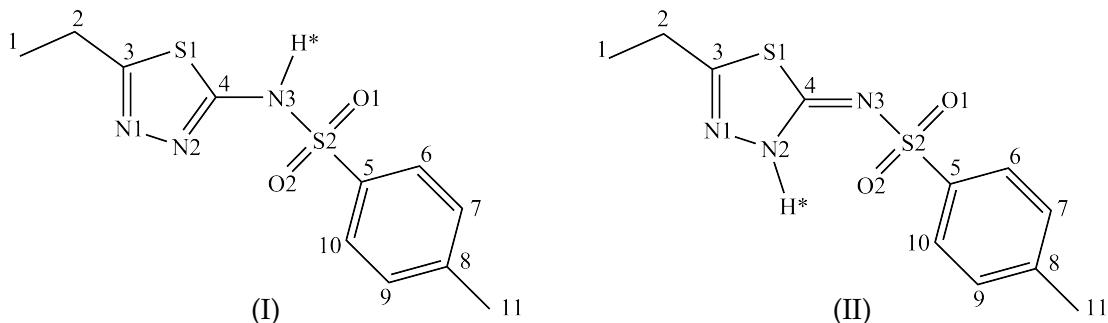


Figure 4.1: Two possible tautomers of the compound. (I): the amine tautomer (II): the imine tautomer. *Reprinted from Appendix I.*

in a series of antibacterial drugs.¹⁰⁰ UKIRAI can form either the amine tautomer (I) with the asterisked H-atom positions on the sulfonamide nitrogen, or the imine tautomer (II) with the asterisked H-atom positions on one nitrogen atom of the 1,3,4-thiadiazole ring, as shown in Figure 4.1. The crystal structure of UKIRAI was solved from a powder sample and reported as the amine tautomer, which was supported by ¹³C SS-NMR experiments and calculations.⁹⁹ Nevertheless, the root-mean-square deviation (RMSD) between the calculated and the experimental chemical shift is much larger than the expected deviation (see Section 2.4.3), indicating that the reported structure may be incorrect. When analysing the hydrogen bonds in the crystal structure, the hydrogen atom in the amine tautomer is pointing at a methyl group, whereas in the imine tautomer, a N—H···O hydrogen bond is formed at a reasonable site along the α axis (Figure 4.2). Based on the preliminary analysis on the hydrogen bond network, we postulated that the correct tautomer should be the imine form. This was subsequently validated by DFT-D energy minimisations and *ab initio* SS-NMR calculations.

4.1.2 Pigment Yellow 138

C.I. Pigment Yellow 138 (P.Y. 138) is a quinophthalone pigment with a greenish yellow colour that exhibits tautomerism. It is used primarily for the colouring of plastics and industrial finishes.^{101,102} Despite its commercial application for decades, the tautomeric state of P.Y. 138 in the solid state remained ambiguous for a long time (Figure 4.3). The compound was reported as CH-form,^{101,102} OH-form^{103–106} and NH-form¹⁰⁷ in different textbooks, review articles and patents. Single crystals are typically desired to determine a crystal structure; however, obtaining a single crystal of P.Y. 138 was a failure. This was expected, because pigments are designed to have a very low solubility in most solvents.¹⁰² Alternatives were thus adopted to determine the correct tautomeric state for P.Y. 138.

4.2 Summary of methods

4.2.1 Modelling of the experimental crystal structures

UKIRAI

The crystal structure of UKIRAI deposited in the Cambridge Structural Database (CSD)¹⁰⁸ has no hydrogen atoms. Therefore, two models (I) and (II) were made, based on Figure 4.1; hydrogen atoms were added to the two models accordingly. These two model struc-

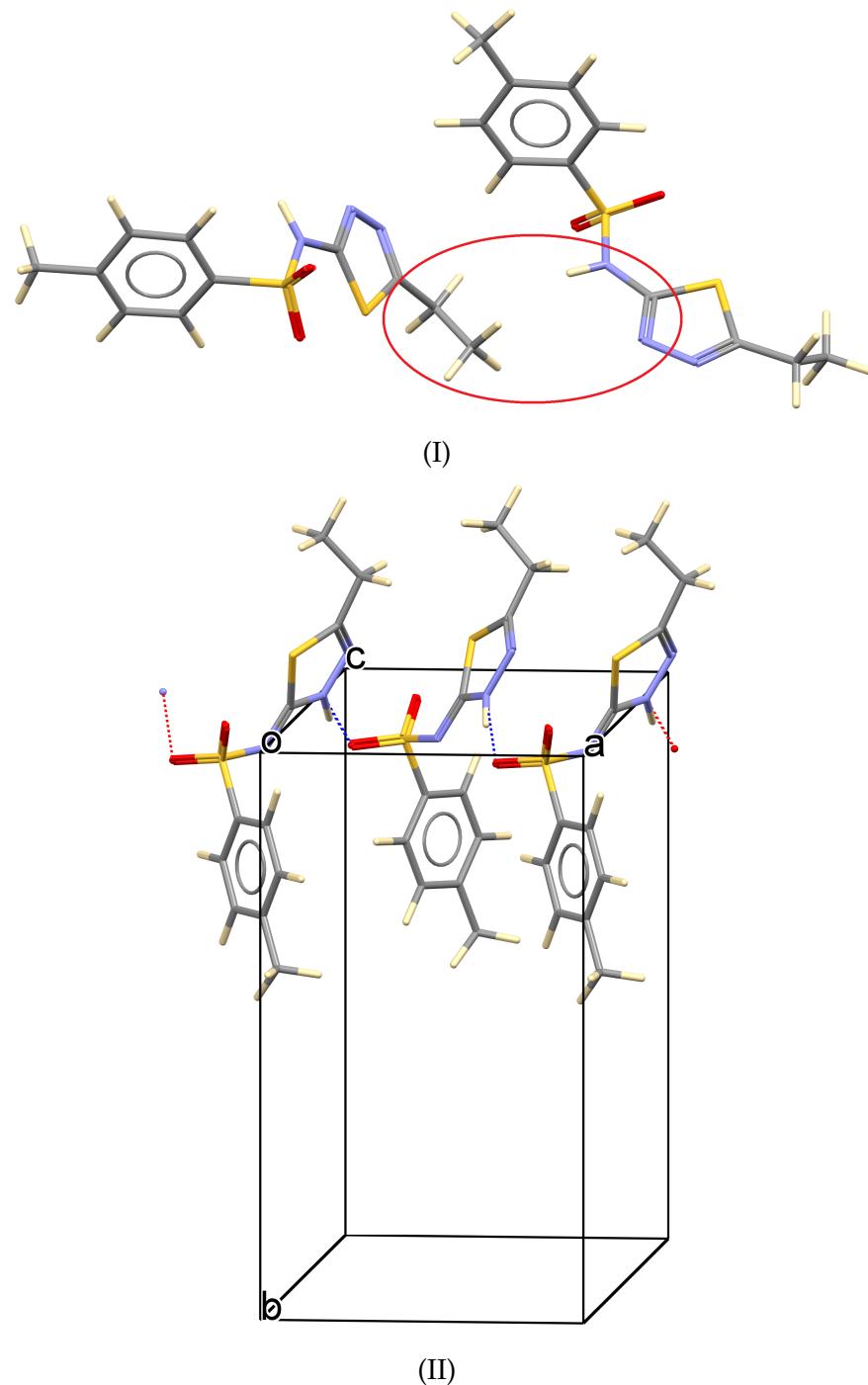


Figure 4.2: Intermolecular interactions of (I) the amine tautomer, in which the proton is pointing towards a methyl group. (II) the imine tautomer. Hydrogen bond interactions are indicated in dashed lines.

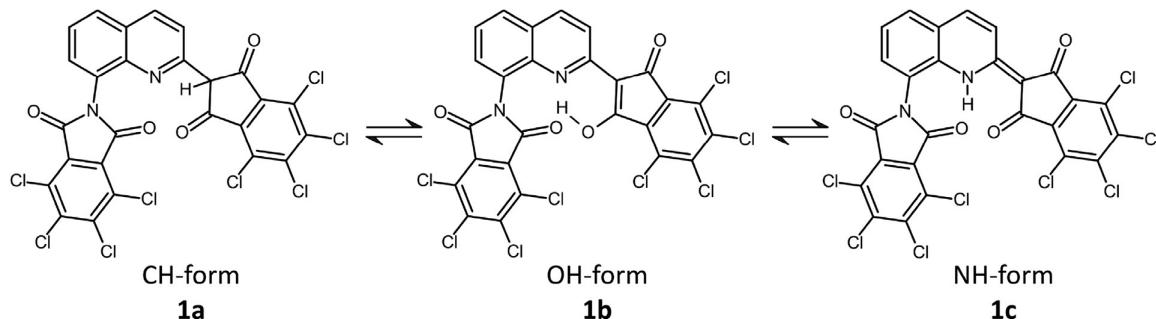


Figure 4.3: The possible tautomers of P.Y. 138. *Reprinted from Appendix II with permission.*

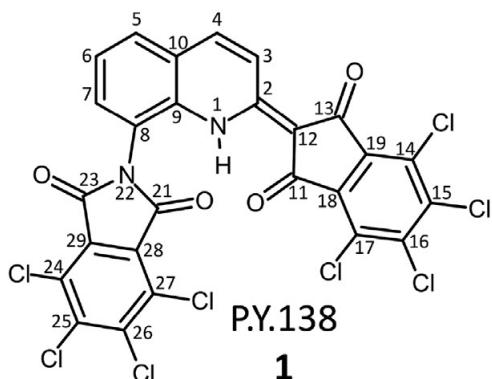


Figure 4.4: The atomic labels of P.Y. 138. *Reprinted from Appendix II with permission.*

tures were subjected to preliminary DFT-D energy minimisations using CASTEP,¹⁰⁹ with the space group *Pbca* constrained and both the atomic positions and the lattice parameters allowed to vary. The optimised geometries of models (I) and (II) provided restraints on the bonded interactions. They were then used for Rietveld refinement against the published PXRD data.⁹⁹ The final experimental structures for models (I) and (II) adopted the cell parameters and positions of non-hydrogen atoms from the Rietveld refinement. The positions of hydrogen atoms were obtained from the DFT-D energy minimisations.

An “average model” based on the experimental structures (I) and (II) was introduced to separate the effect of H-atom positions on calculated ¹³C chemical shifts. The averaging was conducted for the lattice parameters and all atoms of the two structures except the tautomeric hydrogen atom. According to the tautomeric states of (I) and (II), the hydrogen atoms were then added to the average model and formed two new structures, ‘average (I)’ and ‘average (II)’. DFT-D energy minimisations were carried out for the two averaged structures with only the hydrogen atoms allowed to move; the unit-cell parameters and non-H atoms were fixed.

P.Y. 138

The lattice parameters of P.Y. 138 were obtained by indexing the PXRD pattern. The space group, *P2₁/c*, was obtained by identifying the extinction symbol from electron diffraction intensities and analysing the probability of possible space groups. The coordinates of P.Y. 138 in three dimensions were then solved by simulated annealing in DASH;¹¹⁰ three possible tautomers were used as input. The NH and the OH tautomers achieved good figures of

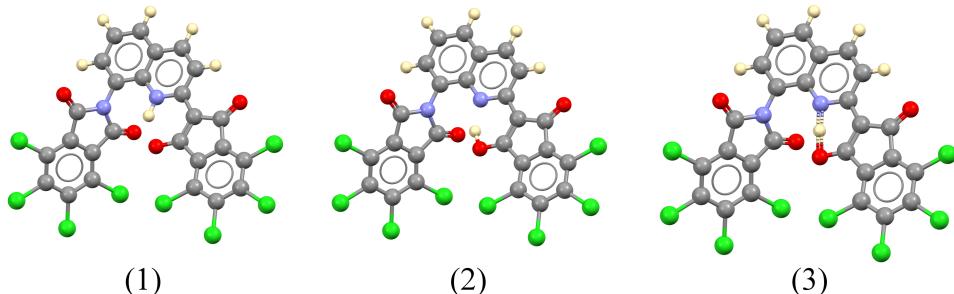


Figure 4.5: Three models of P.Y. 138 prepared for ^{13}C solid-state NMR calculations. (1) The NH-model. (2) The OH-model. (3) The halfway model.

merit against the experimental data. The Rietveld refinement failed to generate the CH model due to a non-planar sp^3 -carbon atom, C12 (Figure 4.4). Therefore, only the NH and the OH tautomers were considered in further calculations.

4.2.2 DFT-D energy minimisations

UKIRAI

The plane-wave DFT code *CASTEP*¹⁰⁹ was used for energy minimisations of the experimental structures (I) and (II). The PBE functional²³ was used with the Grimme-2006 semi-empirical dispersion-correction scheme,³⁷ this is known as PBE-D2. To speed up the calculations and to avoid the structures getting trapped in undesired minima, the energy minimisations were split into three steps with increased degrees of freedom and kinetic energy cut-offs. The details are documented in Appendix I.¹¹¹

P.Y. 138

The experimental NH- and OH-forms of P.Y. 138 were initially subjected to energy minimisations using *CASTEP*¹⁰⁹ with PBE-D2. Both tautomers converged to the NH form. However, the OH-form cannot be readily excluded because the PBE functional is known to show overbinding for hydrogen bonds in ice.¹¹² The two experimental structures underwent energy minimisations with the *ab initio* simulation code *VASP*^{113–116} *via GRACE*. The BLYP functional^{20,21} was used with the Grimme-2010 dispersion-correction scheme,³⁸ the combination is referred to as BLYP-D3.

To investigate how ^{13}C chemical shifts vary under the influence of prototropic tautomerism, three models were built with everything being equal but the positions of the proton. The lattice parameters and the atomic coordinates were taken from the average of the BLYP-D3 energy minimised NH- and OH-forms. The three models are: (1) the NH-model, in which the proton was positioned near N using the N–H bond length from neutron data; (2) the OH-model, in which the proton was positioned near O using the O–H bond length from neutron data and (3) the halfway model, in which the proton was positioned halfway between N and H. Figure 4.5 provides a visualisation of the three models.

	Energy cut-off [eV]	Energy minimisation protocol	All-atom RMSCD	Non-H-atom RMSCD
(I)	520	Unit cell fixed	0.4309	0.3487
(I)	520	Unit cell free	0.4356	0.3561
(I)	1200	Unit cell free	0.6972	0.5749
(II)	520	Unit cell fixed	0.1494	0.1283
(II)	520	Unit cell free	0.1569	0.1329
(II)	1200	Unit cell free	0.1626	0.1282

Table 4.1: Root-mean-square Cartesian displacements (RMSCDs) (\AA) between the PBE-D2 energy minimised and the experimental structure of tautomers (I) and (II). *Reproduced from Appendix I.*

4.2.3 SS-NMR calculations

The GIPAW method⁹³ implemented in CASTEP¹⁰⁹ was employed with the ultrasoft pseudopotentials generated on-the-fly⁴⁶ to calculate the shielding constants for all the models, including energy-minimised models (I) and (II), average (I) and average (II) of UKIRAI, plus the NH-model, the OH-model and the halfway model of P.Y. 138. The kinetic energy cut-off for SS-NMR calculations was 1,200 eV. Integrals taken over the first Brillouin zone were done on a Monkhorst-Pack grid⁴² with a spacing of approximately 0.05 \AA^{-1} in the reciprocal space and at least two k -points along each direction.

4.3 Results and discussion

4.3.1 UKIRAI

Rietveld refinement against PXRD data

The Rietveld refinement plots of the experimental structures of the amine tautomer (I) and imine tautomer (II) in the present study are shown in Figure 4.6. Both (I) and (II) yielded essentially the same goodness of fits, $\chi^2_{(I)} = 1.53$ and $\chi^2_{(II)} = 1.53$. The R -factors of the two tautomers are very close, $R(F)_{(I)} = 3.4\%$ and $R(F)_{(II)} = 3.3\%$. Consequently, it is impossible to discriminate the correct tautomer based on the Rietveld refinement against the PXRD data.

PBE-D2 energy minimisations

Figure 4.7 provides overlays of the PBE-D2 energy minimised and the experimental structures of (I) and (II). The amine tautomer (I) shows a large distortion after the energy minimisation. The RMSCDs between the energy minimised and the experimental structures are listed in Table 4.1. According to the benchmark of RMSCDs (Table 2.1), the imine tautomer (II) is likely to be correct, whereas the amine tautomer (I) would be incorrect.

^{13}C SS-NMR calculations

The assignment of the chemical shifts of UKIRAI was taken from Hangan *et al.*,⁹⁹ in which the ^{13}C chemical shifts were assigned based on solution NMR experiments. There are two

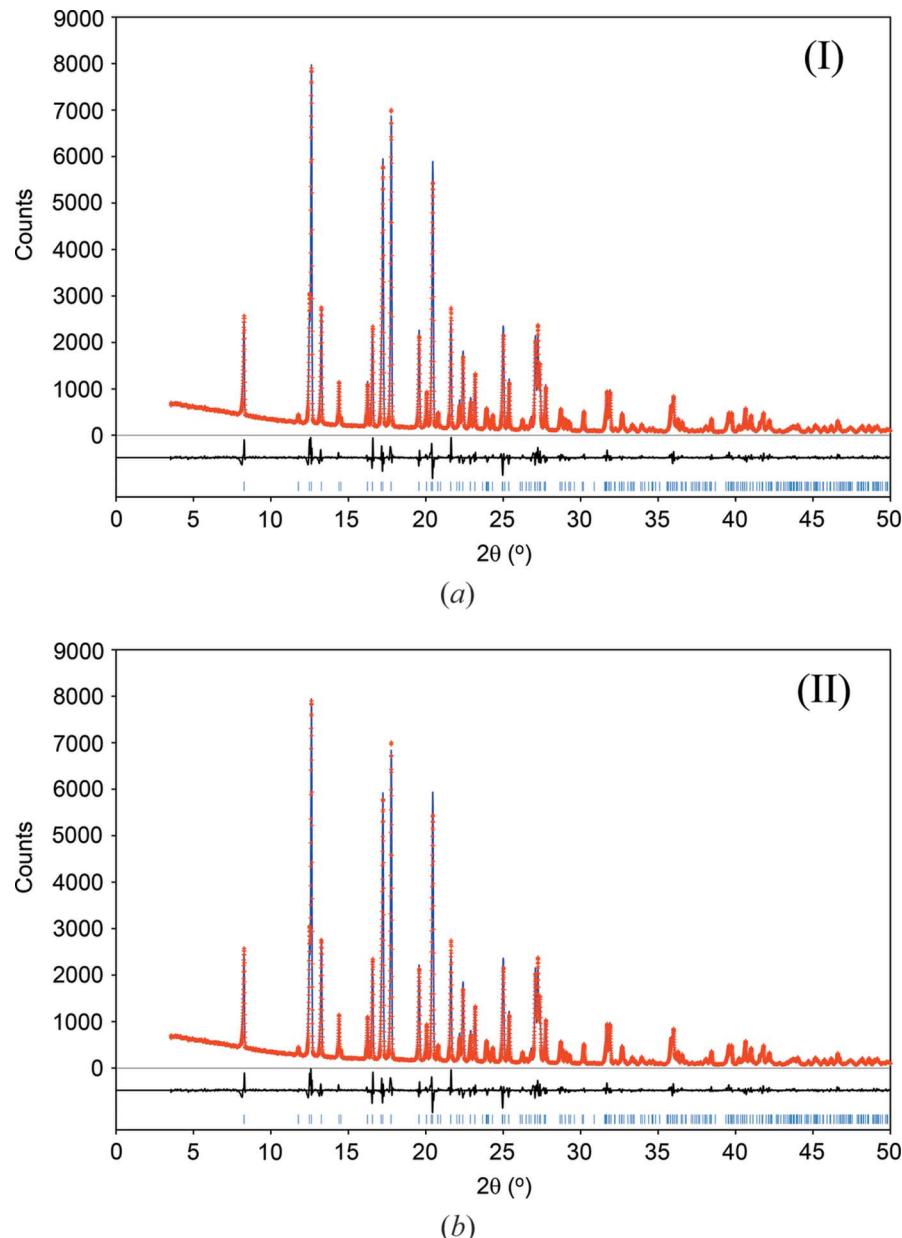


Figure 4.6: Rietveld refinement fitting against the PXRD data of UKIRAI for (a) the amine tautomer (I) and (b) the imine tautomer (II) presented in this study. The PXRD data was obtained from Hangan *et al.*⁹⁹ *Reprinted from Appendix I.*

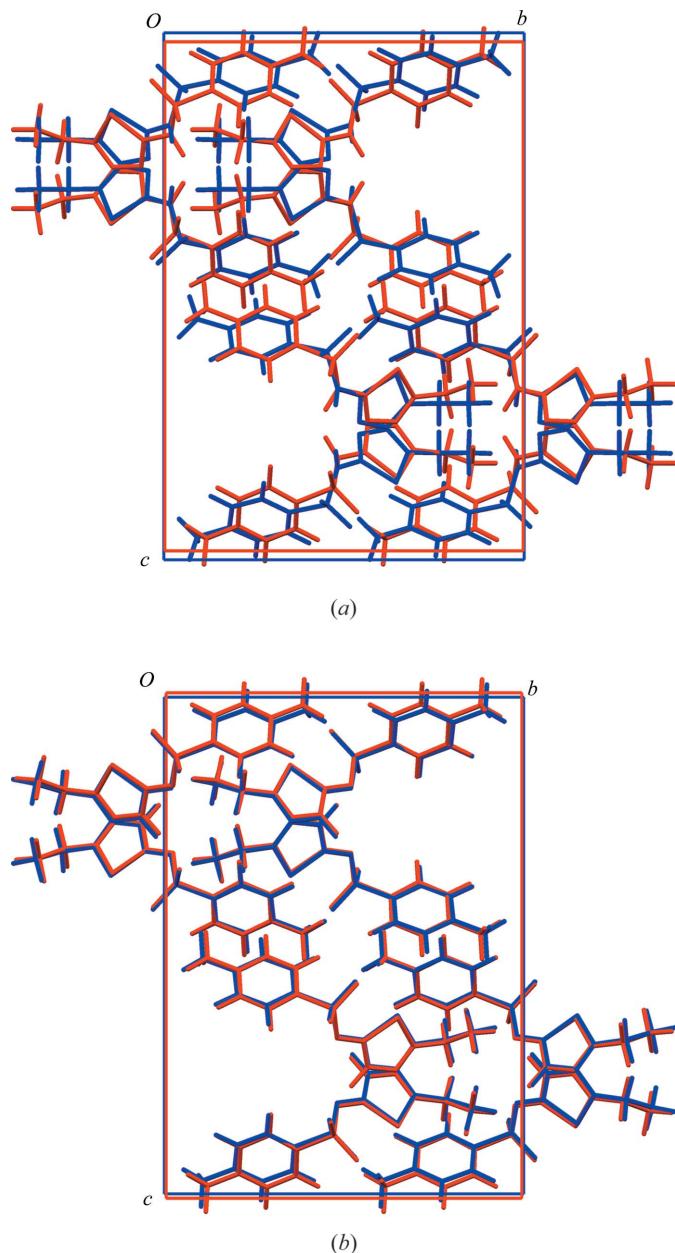


Figure 4.7: Overlays of the PBE-D2 energy minimised structure (in blue, energy cut-off at 520 eV, unit cell parameters were free to vary) and experimental structure of (in red). (a): the amine tautomer (II). (b): the imine tautomer (II). *Reprint from Appendix I.*

	Experimental	(I)	(II)	Average (I)	Average (II)
C1	14.0	11.8 (-2.2)	11.8 (-2.2)	13.2 (-0.8)	10.9 (-3.1)
C2	23.5	19.4 (-4.1)	20.9 (-2.6)	20.6 (-2.9)	22.1 (-1.4)
C3	165.7	176.4 (10.7)	168.2 (2.5)	165.4 (-0.3)	161.4 (-4.3)
C4	161.9	165.4 (3.5)	163.1 (1.2)	159.3 (-2.6)	162.6 (0.7)
C5	138.9	138.9 (0.0)	140.3 (1.4)	140.4 (1.5)	140.6 (1.7)
C6	127.6	125.5 (-2.1)	128.1 (-0.2)	126.2 (-1.4)	128.4 (0.8)
C7	130.5	128.7 (-1.8)	131.6 (1.1)	132.5 (2.0)	132.0 (1.5)
C8	145.0	147.4 (2.4)	147.1 (2.1)	148.9 (3.9)	148.2 (3.2)
C9	132.0	131.5 (-0.5)	133.3 (1.3)	136.6 (4.6)	135.0 (3.0)
C10	128.3	125.9 (-2.4)	126.7 (-0.9)	126.9 (-1.4)	128.5 (0.2)
C11	21.3	17.8 (-3.5)	17.9 (-3.4)	18.9 (-2.4)	19.3 (-2.0)
MAD		3.0	1.7	2.2	2.0
RMSD		4.0	1.9	2.5	2.3

Table 4.2: Experimental and calculated chemical shifts of UKIRAI (unit: ppm). Numbers in the brackets are the deviations (unit: ppm) between the experimental and the calculated chemical shifts. MAD: mean absolute deviation. *Reproduced from Appendix I.*

topologically equivalent atom pairs in the benzene ring of the UKIRAI molecule, namely C6/C10 and C7/C9 (Figure 4.1). Since no more experimental data can be used for a full assignment, the calculated chemical shifts were assigned so as to achieve the best fit with the experimental values.

Table 4.2 demonstrates the calculated chemical shifts of different models. Structure (I) yields an RMSD of 4.0 ppm for ^{13}C chemical shifts, which, as previously observed by Hangan *et al.*,⁹⁹ is noteworthy larger than the expectation (see Section 2.4.3). Moreover, the deviation of C3 in structure (I) is 10.7 ppm: this is significant and indicates an error in the tautomeric state. On the other hand, in structure (II), the RMSD value is 1.9 ppm and the largest deviation of an individual carbon atom is -3.4 ppm (C11): these results show much better agreement between the calculated and experimental chemical shifts.

The positions of heavy atoms of average (I) and average (II) are identical. However, as seen from the distortion of (I) and the small RMSCD of (II) after the energy minimisations, the averaged positions of heavy atoms give a better match for (I) but give a worse match for (II), as seen in Table 4.2. Their RMSDs are very close; the RMSD of average (I) is smaller than that of (I), and the RMSD of average (II) is larger than that of (II). The difference in the individual deviations of ^{13}C chemical shifts of average (I) and average (II) indicates that the ^{13}C chemical shifts are sensitive to the positions of both the hydrogen atoms and the heavy atoms. The large RMSD given by (I) can thus be ascribed to the wrong tautomeric state, which then results in the incorrect positions of heavy atoms.

4.3.2 P.Y. 138

DFT-D energy minimisations

Two methods were used for the DFT-D energy minimisations of the NH- and the OH-forms of P.Y. 138, namely PBE-D2 and BLYP-D3. The PBE functional did not consider the OH-form

	Bond length (BLYP-D3) [Å]	Bond length (neutron) [Å]	Delta [%]
NH-form	1.032	1.015	1.67
OH-form	1.093	0.993	10.17

Table 4.3: Lengths of the N–H and the O–H bond of P.Y. 138 from BLYP-D3 calculations. The neutron data were taken from the CSD.¹⁰⁸

as an energy minimum. In the case of BLYP-D3, the lattice energy given by the OH-form is 38.9 kJ/mol higher than that of the NH-form. Moreover, the O–H bond length is 10.07% larger than the neutron data, whereas the N–H bond is only 1.67% longer than neutron data (Table 4.3). These results indicate that the OH-form is not favourable.

¹³C SS-NMR calculations

The ¹³C calculated chemical shifts of the NH-, the OH- and the halfway models are given in Table 4.4, Table 4.5 and Table 4.6, respectively. The smallest RMSD is given by the halfway model (2.4 ppm), which is followed by the NH-model (2.8 ppm) and the OH-model (3.1 ppm). Based on the expectation (see Section 2.4.3), the calculations prefer the halfway model, the RMSD of the NH-model lies at the boundary of ca. 2σ confidence. The OH-model is not likely to be correct. A possible explanation for the results is that the proton is delocalised and involved in a six membered aromatic-like ring, forming a resonance-assisted hydrogen bond (RAHB).¹¹⁷ The coplanar geometry of the quinoline and indandione ring and the increased N–H bond length (Table 4.3) are strong indications of the RAHB. In 2015, Dračínský *et al.* delineated the properties of RAHBs using SS-NMR experiments and path integral MD (PIMD) simulations.¹¹⁸ The study reveals that the RAHBs can be successfully investigated if proper dynamic descriptions are introduced. The PIMD simulation is not discussed since it is out of the scope of the thesis.

The NH-form was finally determined as the correct tautomer with the assistance of SS-NMR experiments (see Appendix II).* A 2D ¹H-¹⁴N heteronuclear multiple-quantum correlation (HMQC) spectrum was recorded (Figure 4.8). The correlation peak between the ¹H and the ¹⁴C signals in the spectrum shows a ¹J-coupling between the two atoms, which is a direct evidence of the covalent N–H bond.

4.4 Conclusions

Static DFT-D energy minimisations and *ab initio* SS-NMR calculations were used to determine the tautomeric states of molecules in two molecular crystals, UKIRAI and P.Y. 138. In both cases, the correct tautomers were not able to be discriminated with PXRD data. The identification of correct tautomers was achieved by comparing the RMSDs between the calculated and experimental ¹³C chemical shifts with the established expectation, 1.9 ± 0.4 ppm.

*The structure elucidation of P.Y. 138 is a collaborative project. The SS-NMR experiments were carried out by Prof. Michele R. Chierotti and Prof. Roberto Gobetto (University of Turin, Italy). The experimental results were mentioned here for completeness.

Atom label	Exp. chemical shift [ppm]	Cal. chemical shift [ppm]	Deviation [ppm] ^a	Unassigned exp. chemical shift [ppm]
2	149.8	142.7	-7.1	149.8
3	119.5	120.0	0.5	149.8
4	141.5	141.5	0.0	149.8
5	129.9	128.4	-1.5	141.5
6	126.6	127.4	0.8	133.7
7	132.0	131.6	-0.4	133.7
8	119.9	120.4	0.4	131.9
9	130.7	128.7	-2.0	130.8
10	124.2	123.4	-0.8	130.8
11	187.6	194.6	7.0	130.8
12	99.3	102.2	2.9	124.2
13	184.4	187.0	2.6	124.2
14	130.8	130.8	0.0	124.2
15	141.5	143.0	1.5	
16	149.8	144.7	-5.1	
17	131.9	131.7	-0.2	
18	130.8	131.1	0.2	
19	130.8	130.6	-0.2	
21	161.5	164.2	2.7	
23	162.2	165.5	3.3	
24	133.7	133.8	0.1	
25	149.8	146.2	-3.6	
26	149.8	145.5	-4.3	
27	133.7	135.0	1.3	
28	124.2	123.5	-0.7	
29	124.2	124.0	-0.2	

Reference shielding constant (σ_{ref}) = 165.9 ppm
Mean absolute deviation (MAD) = 1.9 ppm
Root-mean-square deviation = 2.8 ppm

Table 4.4: Calculated and experimental ^{13}C chemical shifts with experimental assignment for the NH-form of P.Y. 138. Reproduced from Appendix II with permission.

^aDeviation = Cal. chemical shift - Exp. chemical shift.

4.4. CONCLUSIONS

Atom label	Exp. chemical shift [ppm]	Cal. chemical shift [ppm]	Deviation [ppm] ^a	Unassigned exp. chemical shift [ppm]
2	149.8	150.4	0.6	149.8
3	119.5	119.5	-0.1	149.8
4	141.5	137.4	-4.1	149.8
5	129.9	130.4	0.5	141.5
6	126.6	125.9	-0.7	133.7
7	132.0	129.9	-2.1	133.7
8	119.9	123.9	4.0	131.9
9	130.7	136.7	5.9	130.8
10	124.2	124.6	0.4	130.8
11	187.6	184.6	-3.0	130.8
12	99.3	106.1	6.8	124.2
13	184.4	187.0	2.6	124.2
14	130.8	131.3	0.4	124.2
15	141.5	143.0	1.5	
16	149.8	144.2	-5.6	
17	130.8	131.0	0.2	
18	133.7	133.6	-0.1	
19	130.8	127.4	-3.4	
21	161.5	163.9	2.4	
23	162.2	165.4	3.2	
24	133.7	133.9	0.2	
25	149.8	144.9	-4.9	
26	149.8	144.7	-5.1	
27	131.9	133.3	1.4	
28	124.2	123.5	-0.7	
29	124.2	124.0	-0.2	
Reference shielding constant (σ_{ref}) = 165.9 ppm				
Mean absolute deviation (MAD) = 2.3 ppm				
Root-mean-square deviation = 3.1 ppm				

Table 4.5: Calculated and experimental ^{13}C chemical shifts with experimental assignment for the OH-form of P.Y. 138. Reproduced from Appendix II with permission.

^aDeviation = Cal. chemical shift - Exp. chemical shift.

Atom label	Exp. chemical shift [ppm]	Cal. chemical shift [ppm]	Deviation [ppm] ^a	Unassigned exp. chemical shift [ppm]
2	149.8	150.7	0.9	149.8
3	119.5	119.8	0.3	149.8 C16/C25/C26
4	141.5	139.3	-2.2	149.8
5	129.9	128.8	-1.1	141.5 C15
6	126.6	126.7	0.1	133.7 C24/C27
7	132.0	131.0	-1.1	133.7
8	119.9	122.8	2.9	131.9 C18
9	130.7	130.6	-0.1	130.8
10	124.2	123.8	-0.4	130.8 C14/C17/C19
11	187.6	191.7	4.1	130.8
12	99.3	103.6	4.3	124.2
13	184.4	186.2	1.8	124.2 C10/C28/C29
14	130.8	131.2	0.4	124.2
15	141.5	143.3	1.8	
16	149.8	144.3	-5.5	
17	130.8	131.5	0.7	
18	131.9	133.1	1.2	
19	130.8	129.1	-1.8	
21	161.5	163.9	2.4	
23	162.2	165.4	3.2	
24	133.7	133.8	0.1	
25	149.8	145.3	-4.5	
26	149.8	144.9	-4.9	
27	133.7	134.1	0.4	
28	124.2	123.7	-0.5	
29	124.2	124.0	-0.2	

Reference shielding constant (σ_{ref}) = 165.9 ppm

Mean absolute deviation (MAD) = 1.8 ppm

Root-mean-square deviation = 2.4 ppm

Table 4.6: Calculated and experimental ^{13}C chemical shifts with experimental assignment for the halfway model of PY. 138. Reproduced from Appendix II with permission.

^aDeviation = Cal. chemical shift - Exp. chemical shift.

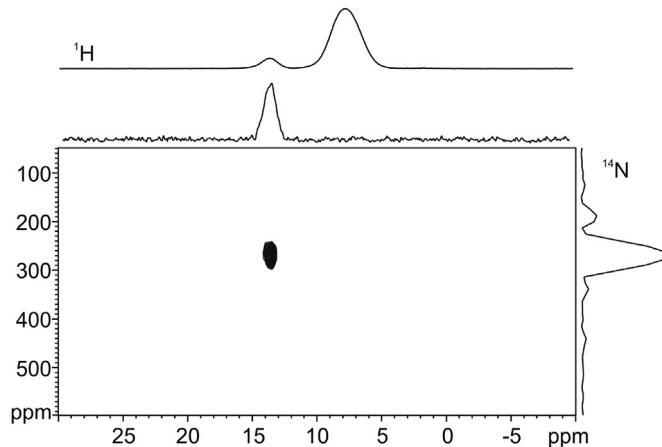


Figure 4.8: ^1H - ^{14}N HMQC spectrum of P.Y. 138. Reprinted from Appendix II with permission.

For UKIRAI, the previously reported amine tautomer turned out to be incorrect; the correct tautomer is the imine form. The RMSD of the amine form given by ^{13}C SS-NMR calculations is out of the expectation. The PBE-D2 energy minimisations prior to the SS-NMR calculations have already indicated the incorrectness of the amine tautomer; the RMSCD before and after the energy minimisations are extraordinary large.

For P.Y. 138, neither the PXRD data nor the DFT-D energy minimisations was indicative enough to tell the tautomers apart. Using ^{13}C SS-NMR calculations, the OH-model can be eliminated; nevertheless, it is still difficult to pinpoint the correct tautomer between the halfway model and the NH-model. This is primarily ascribed to the delocalisation of the proton due to the existence of RAHB. One drawback of static models is thus readily identified: they cannot represent the dynamics or quantum effects of real systems and thus result in discrepancies between the calculations and experimental results.

In general, the combination of static DFT-D energy minimisations and *ab initio* SS-NMR calculations is powerful and accessible for the routine validation of tautomeric molecular crystals. An advantage of computational methods is that the published data can be revisited without obtaining new experimental data. On the other hand, static models are not always applicable for molecular crystals showing tunnelling effects. Therefore, it is recommended to consider the quantum effects for such systems.

Chapter 5

Crystallography and dynamics of ^{13}C solid-state NMR standard compounds: DFT-D and molecular dynamics with the COMPASS force field

*This chapter is based on Appendix III “Crystallographic and dynamic aspects of solid-state NMR calibration compounds: Towards *ab initio* NMR crystallography”.*

5.1 Introduction

SS-NMR spectroscopy is widely-used in the study of solid-state pharmaceuticals, including pure APIs, salts, co-crystals, hydrates and amorphous drugs.¹¹⁹ An NMR spectrometer needs to be adjusted, calibrated and optimised with standard compounds before the measurements. Meanwhile, as discussed in Section 2.4.2, *ab initio* NMR calculations provide magnetic shielding constants, which are then converted to chemical shifts using a reference shielding. Since experimental chemical shifts are referenced to a calibration compound, it is meaningful to obtain a computational reference shielding based on calibration compounds. This requires correct reproduction of calibration compounds in computational modelling.

The study covers the crystallographic and dynamic aspects of four ^{13}C SS-NMR calibration compounds, which were elucidated by X-ray diffraction, DFT-D calculations and MD simulations. The ^{13}C isotope was selected because carbon is one of the most abundant elements in organic molecular crystals. Moreover, ^{13}C SS-NMR chemical shifts are known to be sensitive to the structure and dynamics, which enables probing of detailed information of structures.

5.1.1 Four calibration compounds

The four ^{13}C calibration compounds are adamantane, hexamethylbenzene, glycine (α - and γ -form) and 3-methylglutaric acid (Figure 5.1). At room temperature, adamantane is a plastic phase¹²⁰ and hexamethylbenzene molecules undergo fast reorientation,¹²¹ the assignment of average atomic coordinates for these two compounds is thus not chemically meaningful. For each of them, only two peaks can be identified from the ^{13}C SS-NMR spectrum (Figure 5.2): for hexamethylbenzene, one peak stands for aliphatic carbon atoms and the

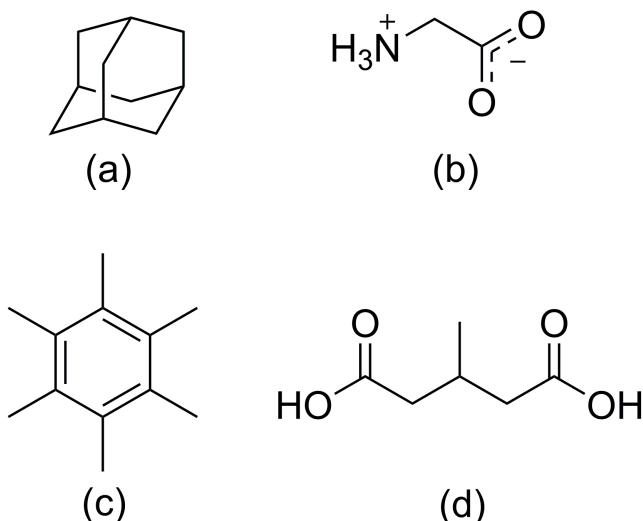


Figure 5.1: Scheme of four calibration compounds investigated in this study. (a): Adamantane. (b): Glycine (α - and γ -form). (c): Hexamethylbenzene. (d): 3-Methylglutaric acid. *Reprinted from Appendix III.*

other stands for aromatic carbon atoms; for adamantane, two peaks arise from secondary carbon atoms and tertiary carbon atoms, respectively.¹²² At ambient conditions, glycine has two stable polymorphs, α - and γ –; both of them can be found in commercial glycine samples.¹²³ A decade ago, 3-methylglutaric acid was proposed as a multifunctional compound for the calibration of NMR spectrometers,¹²⁴ however, its three-dimensional crystal structure had not been determined. The incomplete structural information prohibits us to investigate the calibration compounds computationally.

5.2 Brief summary of methods

5.2.1 Polymorph screening and crystal structure determination of 3-methylglutaric acid

Polymorph screening

Samples of 3-methylglutaric acid (Sigma-Aldrich, Munich, Germany) were dissolved in 11 different solvents and underwent crystallisation by solvent evaporation at room temperature. The resulting powder samples were subjected to PXRD to determine the forms. All powder samples turned out to be the same phase (Table 5.1).

Crystal structure determination

The single crystals of 3-methylglutaric acid were obtained from chloroform, acetonitrile and butanone. The crystals grown in chloroform were selected to undergo SCXRD. The structure was then solved by direct methods and refined by the full-matrix least-square method using SHELXS and SHELXL, respectively.¹²⁵ The experimental details and the properties of determined crystal structures of 3-methylglutaric acid at 173 K and room temperature are summarised in Table 5.2.

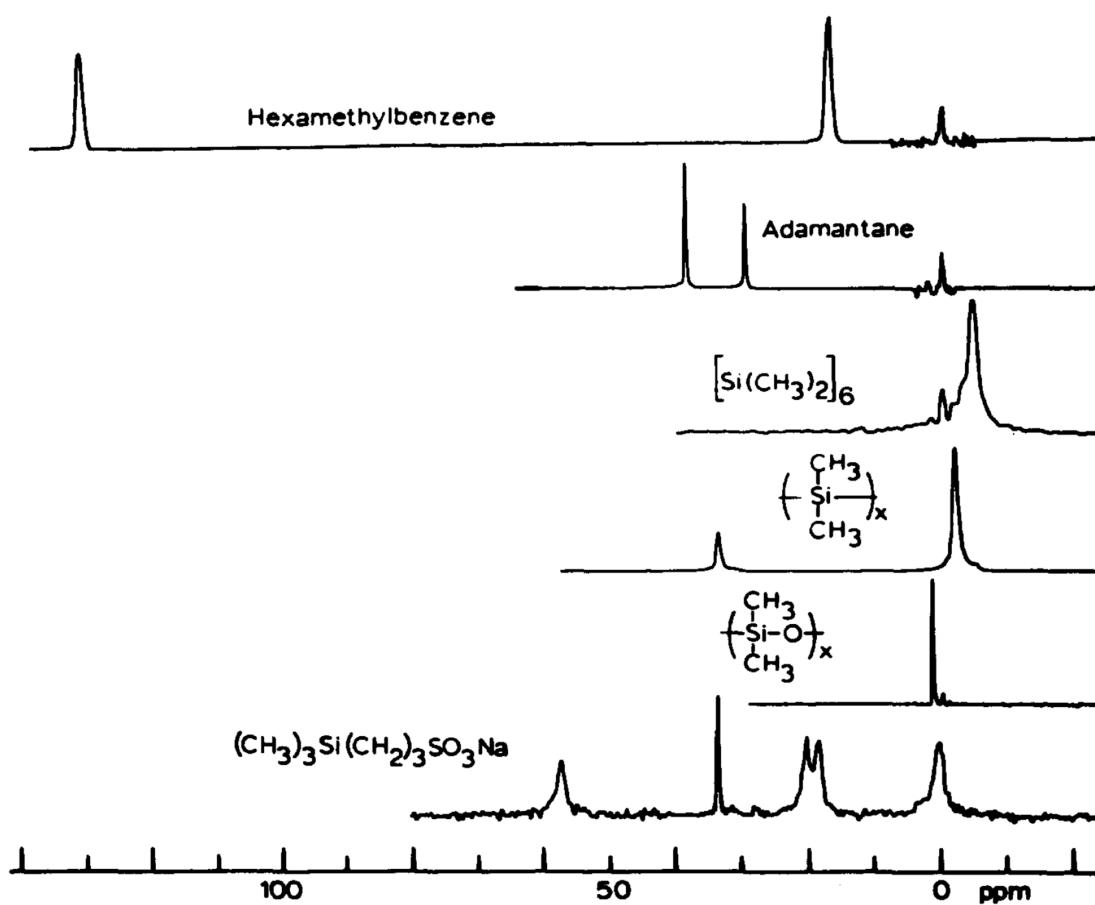


Figure 5.2: The ^{13}C SS-NMR spectra of adamantane and hexamethylbenzene. Reprinted from Earl and Vanderhart¹²² with permission.

Solvent	Concentration [$\text{mg}\cdot\text{mL}^{-1}$]	Form
Acetic acid	100	Form A
Acetone	100	Form A
Acetonitrile	100	Form A
Butanone	100	Form A
Chlorofome	100	Form A
Diisopropyl ether	20	Form A
Dimethylsulfoxide	100	Form A
Ethanol	100	Form A
Ethyl acetate	100	Form A
Tetrahydrofuran	100	Form A
Water	150	Form A

Table 5.1: The crystallisation of 3-methylglutaric acid in 11 different solvents at room temperature.

	Low-temperature	Room-temperature
Chemical formula	C ₆ H ₁₀ O ₄	C ₆ H ₁₀ O ₄
Molecular mass [M _r]	164.14	164.14
Temperature [K]	173	293(2)
Space group	P2 ₁ /c	P2 ₁ /c
Crystal system	Monoclinic	Monoclinic
	<i>a</i> = 13.849(3)	<i>a</i> = 13.9093(12)
Cell lengths [Å]	<i>b</i> = 5.3228(7)	<i>b</i> = 5.3666(6)
	<i>c</i> = 10.127(2)	<i>c</i> = 10.3070(10)
Cell angles	β = 110.284(16)°	β = 110.554(7)°
Volume [Å ³]	700.2(2)	720.40(13)
Density [g·cm ⁻³]	1.386	1.347
Z	4	4
Z'	1	1
Diffractometer	STOE-IPDS-II diffractometer	STOE-IPDS-II diffractometer
Radiation type	Mo K _α , λ = 0.71073 Å	Mo K _α , λ = 0.71073 Å
2θ _{max}	51.27°	55.304°
R-factor [%]	7.73	5.29

Table 5.2: Experimental details of the crystal structure determination of 3-methylglutaric acid.

	Adamantane	Hexamethylbenzene	α -glycine	γ -glycine
Room temperature	ADAMAN06 ¹²⁰	HMBENZ ¹²⁶	GLYCIN20 ¹²⁷	GLYCIN33 ¹²⁸
Low temperature	ADAMAN08 ¹²⁹	HMBENZ04 ¹³⁰	N/A	N/A

Table 5.3: The CSD reference codes of the crystal structures used for the modelling of adamantane, hexamethylbenzene and glycine.

5.2.2 Modelling of experimental crystal structures and DFT-D energy minimisations

Published crystal structures of adamantane and hexamethylbenzene in CSD were used to prepare room-temperature structure models. The lattice parameters and space groups of room temperature structures were combined with the atomic coordinates solved from low-temperature structures. The crystal structure of the α - and the γ -forms of glycine were obtained directly from the CSD. The CSD reference codes of the structures used in this study are listed in Table 5.3. The crystal structure of 3-methylglutaric acid was determined in this study; the low-temperature (173 K) structure was used for further DFT-D energy minimisations and SS-NMR calculations. DFT-D energy minimisations were performed for the two forms of glycine and 3-methylglutaric acid with the protocol presented in Section 4.2.2 (UKIRAI).

5.2.3 Molecular dynamics simulations and SS-NMR calculations

Materials Studio (version 6.0, Accelrys Inc., 2011) was used for MD simulations with the COMPASS force field.⁷⁸ Energy minimisations with the COMPASS force field were conducted for 3-methylglutaric acid and two polymorphs of glycine before the MD simula-

Compound	Lattice parameter						RMSCD ^a [Å]
	<i>a</i> [Å]	<i>b</i> [Å]	<i>c</i> [Å]	α [°]	β [°]	γ [°]	
α -Gly ^b , exp. ¹²⁷	5.10	11.95	5.46	90	111.78	90	—
α -Gly, DFT-D	5.06	11.78	5.47	90	112.67	90	0.089
γ -Gly, exp. ¹²⁸	7.04	7.04	5.48	90	90	120	—
γ -Gly, DFT-D	6.93	6.93	5.51	90	90	120	0.065
3-MGA ^c , exp. [173 K]	13.85	5.32	10.13	90	110.28	90	—
3-MGA, DFT-D	13.60	5.30	9.66	90	108.29	90	0.077

^a RMSCD: root-mean-square Cartesian displacement (hydrogen atoms excluded).

^b Gly: glycine. ^c 3-MGA: 3-methylglutaric acid.

Table 5.4: The cell parameters of the DFT-D energy minimised molecular crystals. *Reproduced from Appendix III.*

tions. A two-step approach which switches the size of the simulation cells was used. Firstly, for each calibration phase, the MD simulation was carried out using a large supercell not smaller than $30 \text{ \AA} \times 30 \text{ \AA} \times 30 \text{ \AA}$; the length of each cell edge ($\geq 30 \text{ \AA}$) was chosen based on two factors: the self-interactions of molecules introduced by periodic boundary conditions and the cut-off distance of the electrostatic and van der Waals interactions. Secondly, this size of supercell is not suitable for *ab initio* SS-NMR calculations. Therefore, a $1 \times 1 \times 1$ small cell ($2 \times 2 \times 2$ for hexamethylbenzene) was employed in a new MD simulation, which was built based on the supercell. Snapshots were then taken from each of the small cell MD trajectories and subjected to SS-NMR calculations using the protocol demonstrated in Section 4.2.3. Details of the MD simulations for each calibration compound, such as the size of supercells, simulation time and the temperature series were described in Appendix III.

5.3 Results and discussion

5.3.1 PBE-D2 energy minimisations

PBE-D2 energy minimisations were used to optimise the atomic positions and validate the correctness of crystal structures determined from experiments, including 3-methylglutaric acid and the α - and the γ -forms of glycine. The lattice parameters and non-hydrogen RMSCDs between the energy minimised and the experimental structures are listed in Table 5.4 for quantitative comparison. The non-hydrogen RMSCDs of all three structures are in consistency with the benchmark of crystal structure validation (Table 2.1), which can be considered as correct structures.

5.3.2 MD simulations

Adamantane and hexamethylbenzene

The spinning of adamantane molecules in the crystal was successfully reproduced by the MD simulations. In the MD simulations of hexamethylbenzene, the molecules were reorienting around six-fold axes, hopping from one minimum to another. On the other hand, the acquisition time of a ^{13}C SS-NMR experiments is 30–60 ms,¹³¹ which is six orders of mag-

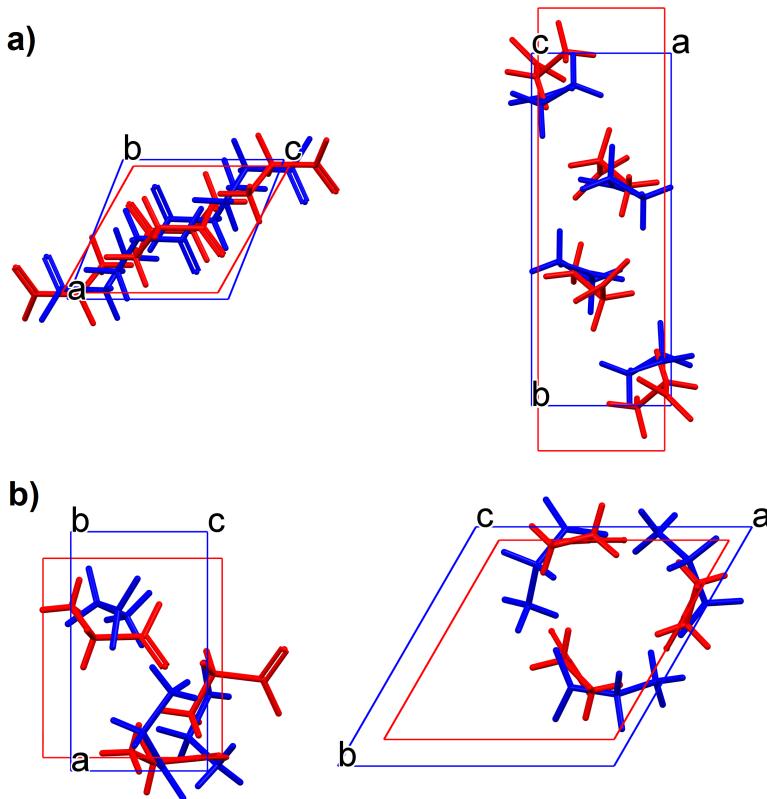


Figure 5.3: The distortion of α - and γ -glycine crystal structures upon the energy minimisations with the COMPASS force field. The experimental structures are in red and the energy minimised structures are in blue. Each overlay is viewed along the b -axis (left hand side) and along the c -axis (right hand side). *Reprinted from Appendix III.*

nitude longer than the length of the MD simulations. It is therefore reasonable to average the ^{13}C SS-NMR chemical shifts of these two compounds using MD snapshots, because the rotator phases of adamantane and hexamethylbenzene have already been captured using MD simulations with a time scale of 10 ns.

Glycine

Upon the energy minimisations with the COMPASS force field for both polymorphs of glycine, large distortions were easily identified (Figure 5.3). Therefore, for establishing the benchmark of ^{13}C SS-NMR calibration compounds, the ^{13}C chemical shifts of glycine calculated based on static DFT-D energy minimisations were used instead. MD simulations were still applied for glycine to analyse the correlation between the correctness of chemical shifts and the instantaneous atomic positions.

3-Methylglutaric acid

Although 3-methylglutaric acid did not show any notable thermal event between 173 K and 358.5 K (the melting point), we still performed MD simulations for 3-methylglutaric acid to sample different configurations, which enabled us to study the significance of motional averaging in *ab initio* SS-NMR calculations. No phase transition or noteworthy distortion was observed upon the MD simulations, which agrees with the experiments.

Site	Method	Cal. CS ^a [ppm]	Exp. CS ^b [ppm]	Deviation ^c [ppm]
α -Gly, ^d CO	DFT-D	179.3	176.5	+2.8
α -Gly, CH ₂	DFT-D	41.8	43.7	-1.9
γ -Gly, CO	DFT-D	176.4	173.4	+3.0
γ -Gly, CH ₂	DFT-D	40.1	41.4	-1.3
ADA, ^e CH	MD	30.9	29.5	+1.4
ADA, CH ₂	MD	39.4	38.5	+0.9
HMB, ^f CH (aromatic)	MD	129.4	132.0	-2.6
HMB, CH ₃ (aliphatic)	MD	16.1	17.2	-1.1
3-MGA, ^g COOH (1)	MD	179.0	180.7 ^h	-1.7
3-MGA, COOH (2)	MD	179.7	181.6 ^h	-1.9
3-MGA, CH ₂ (1)	MD	40.0	38.3 ^h	+1.7
3-MGA, CH ₂ (2)	MD	42.9	39.5 ^h	+3.4
3-MGA, CH	MD	24.1	25.2 ^h	-1.1
3-MGA, CH ₃	MD	17.3	18.8 ^h	-1.5
Reference shielding [ppm]		168.9		
MAD [ppm]		1.9		
RMSD [ppm]		2.0		

^a Cal. CS: calculated chemical shift.^b Exp. CS: experimental chemical shift.^c Deviation = Cal. CS - Exp. CS.^d Gly: glycine.^e ADA: adamantane.^f HMB: hexamethylbenzene.^g 3-MGA: 3-methylglutaric acid.^h Unpublished data, provided by Dr. Sean Delaney and Prof. Dr. Eric Munson.Table 5.5: Calculated and room-temperature experimental ¹³C chemical shifts for calibration compounds. Reproduced from Appendix III.

5.3.3 SS-NMR calculations: the importance of accurate motional averaging

The calculated ¹³C chemical shifts of four calibration compounds are listed in Table 5.5. Using a single reference shielding (σ_{ref}), 168.9 ppm, an RMSD of 2.0 ppm was obtained, which is within the expected error of DFT-based GIPAW NMR calculations (see Section 2.4.3).⁹⁵

The errors of current *ab initio* SS-NMR calculations can be attributed to two major factors. Firstly, there are intrinsic approximation errors in the current DFT scheme, which under- and over-estimate the high- and the low-field magnetic resonances, respectively.^{132,133} Secondly, SS-NMR calculations are usually performed upon static DFT energy minimised structures, which cannot represent the motional averaging over time and space in experiments. Indeed, if the chemical shifts calculated from the static DFT-D structure of 3-methylglutaric acid were used to establish the benchmark, the accuracy could be rather unsatisfactory; the trend of abovementioned errors can be clearly identified, as seen from Table 5.6. The chemical shifts of the carbon atoms on the carboxylic acid groups of 3-methylglutaric acid and on the carboxyl groups of glycine (both forms) were overestimated; in contrast, an underestimation upon the chemical shifts was found for all alkyl carbon atoms. Introducing the motional averaging by MD simulations from 3-methylglutaric acid reduced the RMSD of five phases from 2.9 ppm (Table 5.6) to 2.0 ppm (Table 5.5). This comparison brings the importance of MD simulations in decreasing the systematic intrinsic errors into sight.

Site	Method	Cal. CS [ppm]	Exp. CS [ppm]	Deviation [ppm]
α -Gly, CO	DFT-D	179.2	176.5	+2.7
α -Gly, CH ₂	DFT-D	41.7	43.7	-2.0
γ -Gly, CO	DFT-D	176.3	173.4	+2.9
γ -Gly, CH ₂	DFT-D	40.0	41.4	-1.4
ADA, CH	MD	30.8	29.5	+1.3
ADA, CH ₂	MD	39.3	38.5	+0.8
HMB, CH (aromatic)	MD	129.3	132.0	-2.7
HMB, CH ₃ (aliphatic)	MD	16.0	17.2	-1.2
3-MGA, COOH (1)	DFT-D	186.3	180.7	+5.6
3-MGA, COOH (2)	DFT-D	186.5	181.6	+4.8
3-MGA, CH ₂ (1)	DFT-D	36.7	38.3	-1.7
3-MGA, CH ₂ (2)	DFT-D	36.8	39.5	-2.7
3-MGA, CH	DFT-D	22.7	25.2	-2.6
3-MGA, CH ₃	DFT-D	14.9	18.8	-3.9
Reference shielding [ppm]		168.8		
MAD [ppm]		2.6		
RMSD [ppm]		2.9		

Table 5.6: Calculated and room-temperature experimental ¹³C chemical shifts for calibration compounds using the calculated ¹³C chemical shifts of 3-methylglutaric acid from the DFT-D static structure. *Reproduced from Appendix III.*

Site	Method	Cal. CS [ppm]	Exp. CS [ppm]	Deviation [ppm]
α -Gly, CO	MD	169.1	176.5	-7.4
α -Gly, CH ₂	MD	46.9	43.7	+3.2
γ -Gly, CO	MD	163.6	173.4	-9.8
γ -Gly, CH ₂	MD	52.5	41.4	+11.1
ADA, CH	MD	31.4	29.5	+2.0
ADA, CH ₂	MD	40.0	38.5	+1.4
HMB, CH (aromatic)	MD	130.0	132.0	-2.0
HMB, CH ₃ (aliphatic)	MD	16.7	17.2	-0.5
3-MGA, COOH (1)	MD	179.5	180.7	-1.2
3-MGA, COOH (2)	MD	180.3	181.6	-1.3
3-MGA, CH ₂ (1)	MD	40.5	38.3	2.2
3-MGA, CH ₂ (2)	MD	43.5	39.5	3.9
3-MGA, CH	MD	24.7	25.2	-0.6
3-MGA, CH ₃	MD	17.9	18.8	-1.0
Reference shielding [ppm]		169.5		
MAD [ppm]		3.4		
RMSD [ppm]		4.8		

Table 5.7: Calculated and room-temperature experimental ¹³C chemical shifts for calibration compounds using the calculated ¹³C chemical shifts of two polymorphs of glycine from the MD simulations. *Reproduced from Appendix III.*

Using the chemical shifts calculated from the MD snapshots of glycine for the benchmark of ^{13}C calibration compounds yields an RMSD of 4.8 ppm (see Table 5.7), which is much larger than the established expectation. As discussed in Section 5.3.2, a considerable distortion was found when the COMPASS force field was used for the energy minimisations and MD simulations of α - and γ -glycine. The results in Table 5.7 reveal that the improvement upon the accuracy of the SS-NMR calculations using MD simulations depends on both the correct description of atomic positions, i.e. on the energy potential used, and the motional averaging over instantaneous positions of atoms.

5.4 Conclusions

The performance of the COMPASS force field in NMR crystallography was examined by quantifying the RMSD between the calculated and experimental ^{13}C chemical shifts, yielding an RMSD of 2.0 ppm for four calibration compounds, which is in agreement with the established expectation. Applying MD simulations for rotator phases such as adamantane and hexamethylbenzene is indispensable, because they cannot be represented by a single static frame. Moreover, applying MD simulations for molecular crystals that do not exhibit thermal events at room temperature are also crucial to improve the accuracy in SS-NMR calculations. This gives a better description than that of DFT-D static structures, such as 3-methylglutaric acid in this study. On the other hand, the COMPASS force field cannot reproduce the configurations of two glycine forms correctly, which yielded errors in the prediction of chemical shifts. This reveals a typical limitation of transferable force fields: they are not applicable for a wide range of systems; it requires careful assessment to select a proper transferable force field for a certain system in order to study the motional effects.

Chapter 6

Molecular dynamics with a tailor-made force field for molecular crystals: an *ab initio* NMR crystallography study

This chapter is based on Appendix IV “The application of tailor-made force fields and molecular dynamics for NMR crystallography: a case study of free base cocaine”.

6.1 Introduction

It was discussed in Chapter 5 that the motional averaging with classical molecular dynamics can be used for NMR crystallography only if the force field reproduces the energy and the structure of a system correctly. In this Chapter, we evaluate the performance of the TMFF technique and compared it with established benchmarks, i.e. static PBE-D2 energy minimisations and motional averaging with the COMPASS force field. This is done by investigating whether a TMFF could provide further improvement on the accuracy of SS-NMR calculations.

The crystal structure of free base cocaine was used as a model compound. The 2D chemical formula is shown in Figure 6.1. The free base form of (-)-cocaine is known to exist in nature¹³⁴ and the crystal structure has been determined by Hryncuk *et al.*¹³⁵ In 2013 Baias *et al.* used a combination of CSP, ¹H and ¹³C SS-NMR experiments and DFT calculations based on static structures to identify the experimental form of cocaine and other three pharmaceutical compounds.¹³⁶ This strategy is also called “powder NMR crystallography”.¹³⁷

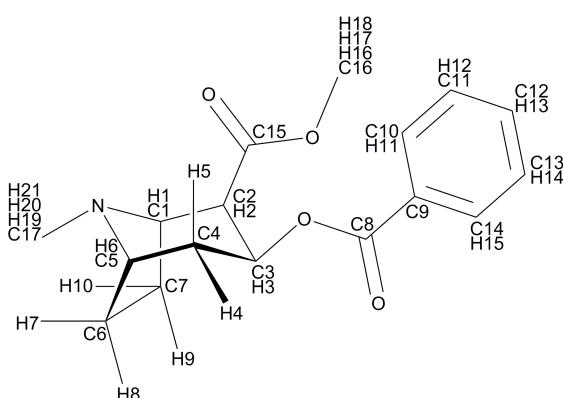


Figure 6.1: The scheme and the atomic numbers of cocaine. *Reprinted from Appendix IV.*

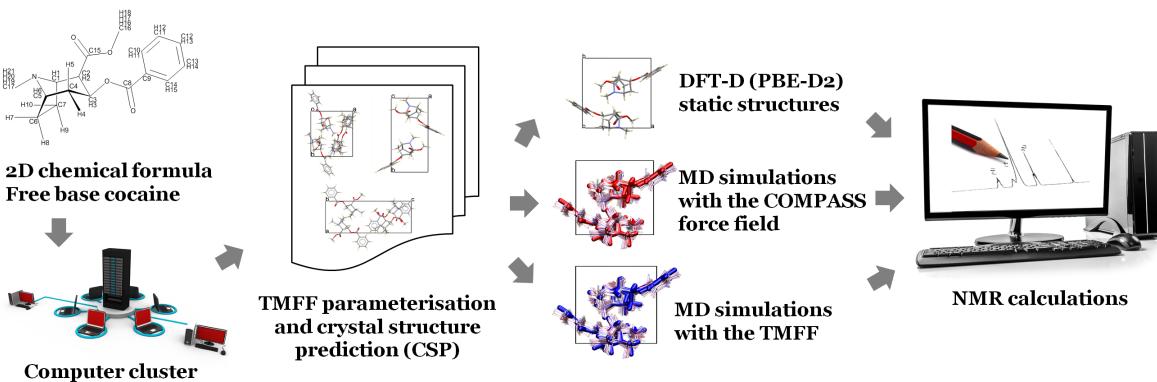


Figure 6.2: The workflow of this *ab initio* NMR crystallography study.

They concluded that for cocaine, the ^1H SS-NMR spectrum can readily discriminate the correct structure with a confidence of 1σ , whilst the ^{13}C spectrum cannot. We therefore incorporated the TMFF and investigated whether the motional averaging with a better force field has an impact on the accuracy of the calculated chemical shifts, and thus a better discrimination for the correct form.

6.2 Brief summary of methods

The workflow of this study is summarised in Figure 6.2. The 2D chemical formula of free base cocaine was used as input for parameterising the TMFF and predicting crystal structures in *GRACE* (version 2.4.87). The TMFF was parameterised in the DREIDING format.⁸⁵ The van der Waals interactions were described in the Lennard-Jones 9-6 form. All 26 predicted CSP candidates were subjected to static PBE-D2 energy minimisations and SS-NMR calculations in *CASTEP*¹⁰⁹ using the protocol described in Section 4.2.3. Those candidates with RMSDs of ^{13}C chemical shifts larger than $+3\sigma$ of the expectation (i.e. 3.1 ppm, see Section 2.4.3) were not considered in MD simulations to reduce the computational cost.

MD simulations with either the COMPASS force field or the TMFF were performed upon selected CSP candidates from PBE-D2 calculations using *Materials Studio*. A two-step MD simulation approach (see Section 5.2.3), which starts with a large cell and switches to a small cell, was adopted in this study. The SS-NMR chemical shifts were calculated for each MD trajectory of a CSP candidate by taking an average over the chemical shifts of 48 molecules from selected snapshots. The calculated chemical shifts were compared with published experimental results to evaluate the performance of different approaches.¹³⁶ The computational details of the MD simulations are reported in Appendix IV.

The accuracy of the COMPASS force field and the TMFF for CSP candidates was assessed and compared with the energies given by the final list of CSP, which used PBE-D3 for calculating the potential energies. The energies of CSP candidates given by the TMFF were calculated along the CSP procedure. For that given by the COMPASS force field, the CSP candidates were subjected to energy minimisations in *Materials Studio* before the calculation of the lattice energies. The RMSDs between the energies given by the two force fields and PBE-D3 were used as an indication of the correctness of the force fields.

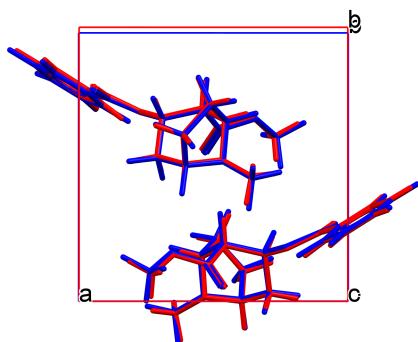


Figure 6.3: Overlay of the experimental structure (in red) and structure 1 (the most stable predicted structure) from the CSP (in blue). *Reprinted from Appendix IV.*

6.3 Results and discussion

6.3.1 Crystal energy landscape

The relative crystal energies of all 26 predicted structures in the final list are shown in Table 6.1 (PBE-D3). The lattice energy of structure 1 is more favourable than that of structure 2 by 4.43 kJ/mol. The overlay of structure 1 and the experimentally determined structure is shown in Figure 6.3 for comparison, which yields an excellent agreement.

However, in this study, only the 1D ^1H and ^{13}C SS-NMR spectra of free base cocaine with the assignment of shifts provided by 2D NMR experiments¹³⁶ were used to discriminate the correct form. The use of PXRD patterns and the crystal energy landscape were discounted for two reasons. First, it was discussed in Section 4.3.1 and Section 4.3.2 that PXRD patterns are not always able to identify the correct form. Second, as Price⁵¹ demonstrated, cases such as isocaffeine and cocaine, which have only one structure significantly more favourable, are rather rare. In most of the cases there can be several thermodynamically stable structures that have very close potential energies. Identifying the correct form from such an energy landscape requires further qualitative and quantitative analysis.

6.3.2 The accuracy of the COMPASS force field and the TMFF

Lattice energies

The accuracy of a force field in describing the CSP candidates can be quantified by the RMSD between the lattice energies given by the force field and that given by the high-quality quantum mechanical method.⁵⁶ Table 6.1 listed the relative lattice energies of CSP candidates calculated from different approaches. In general, the TMFF gives better descriptions in the energies of CSP candidates, yielding an RMSD of 3.0899 kJ/mol, which is better than the COMPASS force field by a factor of ca. 2. Both force fields rank structure 1 as the second most stable structure. However, the energy gaps (ΔE) between structure 1 and the most stable structures from these two force fields are different, $\Delta E_{\text{COMPASS}} = 2.8774 \text{ kJ/mol}$ and $\Delta E_{\text{TMFF}} = 0.0183 \text{ kJ/mol}$.

CSP candidate	Relative energy [kJ/mol]					
	PBE-D3	COMPASS	Dev. (COMPASS)	TMFF	Dev. (TMFF)	
1	-10.0577	-8.5782	1.4795	-4.9826	5.0751	
2	-5.6261	-4.0078	1.6183	-3.2749	2.3512	
3	-4.4321	-6.1491	-1.7170	-0.8151	3.6170	
4	-3.3153	-4.0911	-0.7758	-3.4565	-0.1412	
5	-3.0948	1.5974	4.6922	3.7073	6.8021	
6	-3.0407	-8.2715	-5.2308	-5.0009	-1.9602	
7	-1.8569	-6.1046	-4.2477	-0.3775	1.4794	
8	-1.0022	-0.4768	0.5254	0.0111	1.0133	
9	-0.5890	-0.7063	-0.1173	0.1477	0.7367	
10	-0.4854	-2.6460	-2.1606	-0.5732	-0.0878	
11	0.3340	11.5374	11.2034	-1.2005	-1.5345	
12	0.5082	0.0874	-0.4208	0.1889	-0.3193	
13	0.5703	1.3103	0.7400	-0.8659	-1.4362	
14	0.6362	2.8861	2.2499	-0.6412	-1.2774	
15	0.8815	-6.3781	-7.2596	-1.9341	-2.8156	
16	1.0104	-11.4556	-12.4660	1.5031	0.4927	
17	1.0772	-5.9480	-7.0252	-2.9787	-4.0559	
18	1.2421	-6.1045	-7.3466	-2.8931	-4.1352	
19	2.0844	16.1418	14.0574	7.0601	4.9757	
20	2.2139	12.5006	10.2867	2.8628	0.6489	
21	2.8203	1.1943	-1.6260	-0.9372	-3.7575	
22	3.1775	-0.4993	-3.6768	-1.5194	-4.6969	
23	3.3176	5.9433	2.6257	5.9440	2.6264	
24	3.8796	9.2684	5.3888	6.2577	2.3781	
25	4.7093	7.8532	3.1439	3.9908	-0.7185	
26	5.0377	1.0965	-3.9412	-0.2227	-5.2604	
RMSD [kJ/mol]			5.8979		3.0899	

Table 6.1: The relative lattice energies given by PBE-D3, the COMPASS force field and the TMFF. The average lattice energy given by each approach was set to zero. The deviations (Dev.) are calculated between the energies calculated by a force field (either the COMPASS force field or the TMFF) and that calculated by PBE-D3. *Reproduced from Appendix IV.*

Approach	CSP candidates selected using ^{13}C chemical shifts
PBE-D2	2, 4, 8, 10, 14, 19, 22
COMPASS	1, 2, 3, 4, 6, 9, 10, 13, 15
TMFF	4, 10

Table 6.2: CSP candidates which meet the selection threshold for ^{13}C chemical shifts.

Molecular dynamics simulations

All CSP candidates, i.e. the energy minima provided by PBE-D3, were stable in the MD simulations with the TMFF. Equilibria of the cell parameters and potential energies of the candidates were reached in a few picoseconds. No significant conformational changes or phase transitions were identified. This is expected, as one feature of the TMFF is that it is “mimicking” the DFT-D reference data (see Section 2.3.2), including some of the structures captured in the final CSP list. In a word, the dynamics of CSP candidates can be reproduced with a force field that are parameterised against them.

For structure 1, the averaged structures were calculated from MD simulations with the COMPASS force field and the TMFF and compared with the experimental structure, as shown in Figure 6.4. The averaged structure given by the TMFF shows a better agreement with the experimental data.

Most of the CSP candidates can also be reproduced by the COMPASS force field. On the other hand, phase transitions were identified in the MD simulations of structures 11, 19 and 22. For example, structure 19 underwent a significant conformational change during the MD simulation with the COMPASS force field, as seen from the overlays of the averaged structures from the MD simulation and the predicted structure in Figure 6.5. As shown in Figure 6.5 (a), the rotation of the C3–O bond and the C2–C15 bond along with the equatorial/axial conformational change on the N-methyl group of the tropane ring yield a non-H RMSD of 0.687 Å between the experimental and the averaged molecules. In contrast, the non-H RMSD given by the TMFF is 0.142 Å, which is much smaller than that given by the COMPASS force field, indicating a better agreement.

6.3.3 SS-NMR calculations

^{13}C chemical shifts

Figure 6.6 provides an overview of the RMSDs of ^{13}C chemical shifts between the calculations and the experiments. The red shaded area shows the selection threshold of structures, 1.9 ± 0.4 ppm (see Section 2.4.3). The red dashed line is positioned at $+3\sigma$ of the expectation for shortlisting structures from NMR calculations based on static PBE-D2 structures (see Section 6.2). The PBE-D2, COMPASS and TMFF approaches generate respectively seven, nine and two structures (see Table 6.2). Structure 1 is only selected through the COMPASS approach. In brief, ^{13}C chemical shifts are not capable to discriminate the correct form.

Even so, ^{13}C chemical shifts are sensitive enough to eliminate structures that have significant conformational differences from the experimental form. As shown in Figure 6.6, structures 7, 16 and 18 exhibit exceptional large RMSDs. These structures have their C17 methyl groups situated on the axial position of the six-membered ring of the tropane group, whereas all the rest CSP candidates possess the equatorial conformation. The large RMSD values stem primarily from the deviations given by the chemical shifts of C17 of these three structures, which are larger than 9.0 ppm from the experimental value. These deviations are

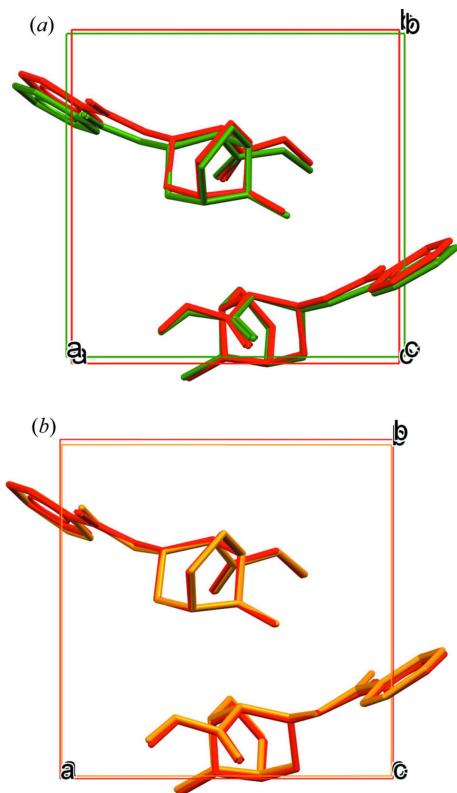


Figure 6.4: Overlays of the experimentally determined free base cocaine (in red) and (a) the average structure obtained from the MD trajectory with the COMPASS force field (in green), (b) from the MD trajectory with the TMFF (in orange). Hydrogen atoms were removed for clarity. *Reprinted from Appendix IV.*

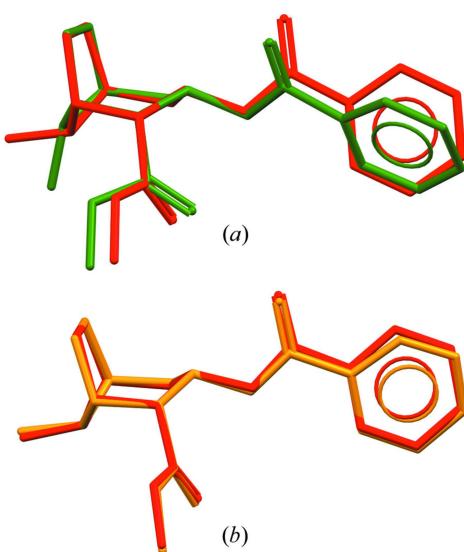


Figure 6.5: Overlays of structure 19 from the CSP (in red) and (a) the average structure obtained from the MD trajectory with the COMPASS force field, (b) from the MD trajectory with the TMFF. Hydrogen atoms were removed for clarity. *Reprinted from Appendix IV.*

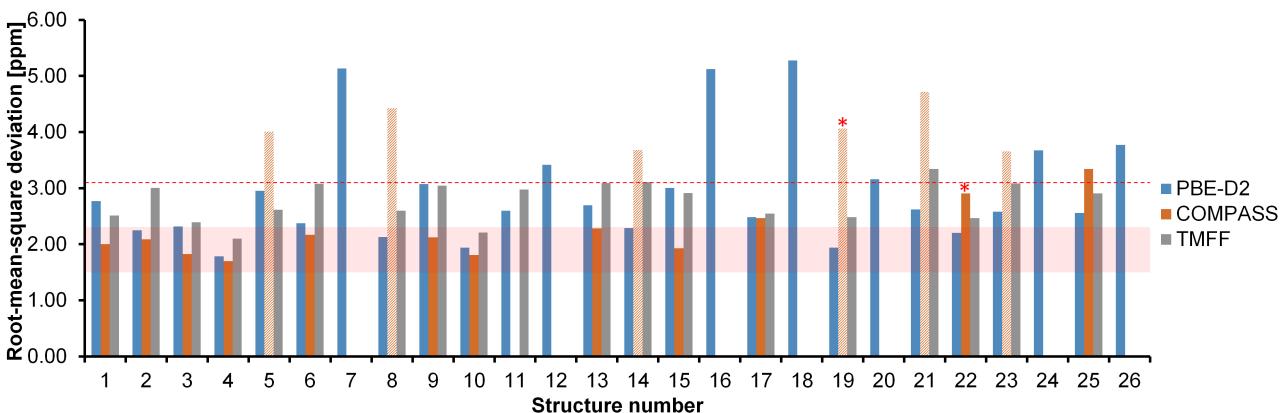


Figure 6.6: The RMSDs of ^{13}C chemical shifts between the calculations and the experiments of CSP candidates. The red dashed line is positioned at $+3\sigma$ from the mean of the expected RMSD, indicating the threshold for CSP candidates subjected to MD simulations. The red shaded zone shows the expectation of the RMSD for ^{13}C isotopes, 1.9 ± 0.4 ppm. The bars with a lighter orange colour indicate the structures underwent equatorial/axial conformational changes for the methyl group on the tropane nitrogen during the MD simulations with the COMPASS force field. The bars with asterisks on the top indicate the structures underwent phase transitions in MD simulations. *Reprinted from Appendix IV.*

in line with the solution ^{13}C NMR study of the *N*-methyl groups on tropane rings.¹³⁸

For structures 5, 8, 14, 19, 21 and 23, the *N*-methyl group on the six-membered ring of tropane changed from the equatorial position to the axial position when the COMPASS force field was applied. The ^{13}C RMSDs of these structures are indicated with a lighter shading in Figure 6.6, which are all increased in comparison with the results from PBE-D2. MD simulations with TMFFs do not show the equatorial/axial conformational change.

^1H chemical shifts

Figure 6.7 shows the RMSDs of ^1H chemical shifts for the three different approaches. The only structure which meets the expectation is structure 1, i.e. the correct structure can be immediately identified using ^1H chemical shifts, as validated by previous studies.^{71,94,136} Although the TMFF approach provides the best agreement with the experimental results with an RMSD of 0.34 ppm, it is not significantly better than the PBE-D2 approach, which gives an RMSD of 0.37 ppm.

6.3.4 Discussion

Limitations of motional averaging in the field of NMR crystallography

The use of a high-quality force field for introducing the motion does not give rise to significant improvement upon the accuracy of the calculated SS-NMR chemical shifts. This can be primarily ascribed to the sensitivity of the isotropic chemical shifts and the method for predicting the chemical shifts.

First, isotropic shieldings may not be sufficiently sensitive to probe the conformational changes of molecules, as opposed to anisotropic chemical shifts. A computational study of

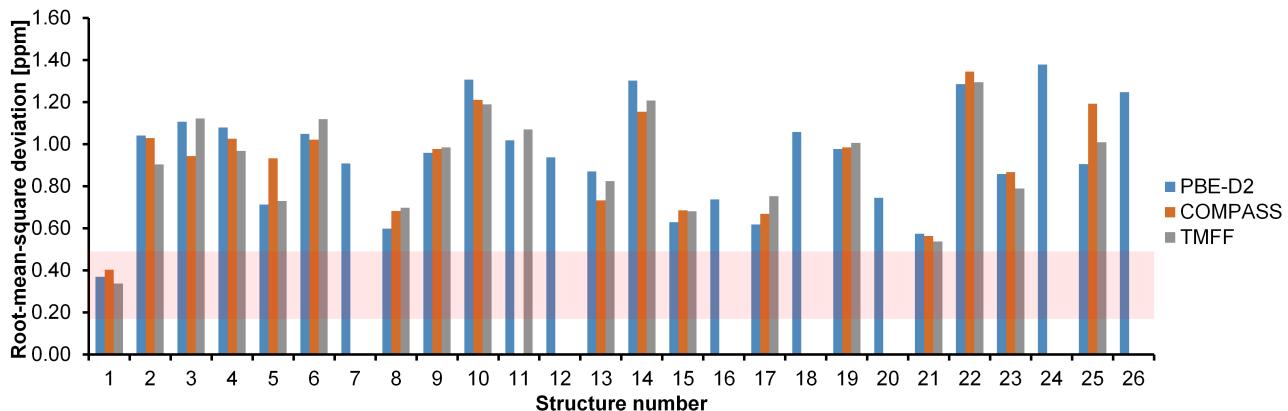


Figure 6.7: The RMSDs of ^1H chemical shifts between the calculations and the experiments of CSP candidates. The red shaded zone shows the expectation of the RMSD for ^1H chemical shifts, 0.33 ± 0.16 ppm. *Reprinted from Appendix IV.*

the ^{13}C chemical shieldings of *meso*-erythritol¹³⁹ reveals that the isotropic shieldings of C1 only varied ca. 2 ppm when then dihedral angle H[O1]-O(1)-C(1)-C(2) rotated 360° , whereas the δ_{11} and δ_{22} principal values varied more than 10 ppm at the same time. It is therefore worthwhile to study the relationship between the motion and the variation of the anisotropic shieldings, if such experimental data are accessible.

Second, the plane-wave DFT-GIPAW method is usually used with GGA-type functionals such as PBE. Hybrid functionals have shown improvements over GGA functionals for the calculations of NMR chemical shifts,^{140,141} however, the implementation of the plane-wave method prohibits the use of hybrid functionals, as discussed in Section 2.2.1. Recently, studies of several test sets of molecular crystals¹⁴² and pharmaceutical molecular crystals¹⁴³ show that fragment-based NMR calculations coupled with hybrid functionals and Gaussian-type basis sets can reduce the errors of ^{13}C chemical shifts to 1.5 ppm. In a word, the improvement brought by the motional averaging with MD simulations is smaller than the intrinsic errors of the GGA-type functional, which was used for calculating SS-NMR chemical shifts in this study.

The scopes of ^{13}C and ^1H chemical shifts

Using ^1H chemical shifts to discriminate the correct structure from a list of structure candidates was proved to be a robust method. On the other hand, the use of ^{13}C chemical shifts in studying the structural details of molecular crystals should not be ignored because either ^{13}C or ^1H chemical shifts have their own scope of applications. For proteins, it has been reported that the ^{13}C chemical shifts are strongly related to the local structure (bonded interactions).^{144,145} The ^1H chemical shifts are mostly connected to the non-local structure (non-bonded interactions).^{146,147} For molecular crystals, ^{13}C chemical shifts can be used for probing bonded interactions, such as the equatorial/axial conformational changes in this study. Meanwhile, such changes cannot be found out with ^1H chemical shifts, as shown from the ^{13}C and ^1H RMSDs of structure 21 (see Figure 6.6 and Figure 6.7). If there are no significant conformational changes during the phase transition, the ^{13}C chemical shifts

given by different polymorphs can be very similar; distinguishing the polymorphs is therefore difficult, as seen from the $\alpha \leftrightarrow \beta$ phase transition of DL-norleucine.¹⁴⁸ Combining both ^{13}C and ^1H spectra can thus be helpful for the interpretation of molecular crystals.

6.4 Conclusions

This Chapter evaluated the performance of the TMFF of free base cocaine. The TMFF outperformed the COMPASS force field in representing the energetic and conformational aspects of predicted crystal structures of cocaine. Nevertheless, introducing motional averaging with MD simulations using such a better force field for calculating the isotropic chemical shifts did not result in a significantly better agreement with the experimental results, from which the limitations of motional averaging were revealed. They were predominantly attributed to the insensitivity of the isotropic chemical shifts and the errors that underpin the methods for calculating SS-NMR chemical shifts. We therefore suggest to pay more attention to the study of anisotropic chemical shifts and to improve the accuracy of NMR computation methods.

Chapter 7

Concluding remarks

Applying computational chemistry methods to the study of molecular crystals has received widespread attention in the recent years. This thesis illustrated the usefulness of computational methods, including dispersion-corrected density functional theory, force-field based molecular dynamics and *ab initio* solid-state NMR calculations. A particular focus was placed on evaluating the performance of a tailor-made force field in the motional averaging of NMR calculations.

Combining static DFT-D energy minimisations, ^{13}C SS-NMR experiments and calculations, the tautomeric states of a potential drug compound (*Z*)-*N*-(5-ethyl-2,3-dihydro-1,3,4-thiadiazol-2-ylidene)-4-methylbenzenesulfonamide and Pigment Yellow 138 were successfully identified. This approach filled the gaps of the PXRD data, in which telling the tautomeric forms apart was not possible. The study also shows that static DFT-D is not optimal for describing systems showing quantum effects such as zero-point vibration and tunnelling.

The calculated ^{13}C chemical shifts of four commonly-used SS-NMR calibration compounds provided an excellent agreement with the experimental results by introducing motional effects for MD simulations using the COMPASS force field. The calculations were not possible without considering the dynamic effects of the plastic phases of two compounds. The study also shows that both the representation of the dynamics and the instantaneous atomic positions are crucial for improving the accuracy of SS-NMR calculations.

The hypothesis of the thesis was finally validated by the *ab initio* NMR crystallography study of free base cocaine. Both the energies and the dynamics of CSP candidates, which are DFT-D energy minima, were well-reproduced using the TMFF; the performance is better than that of the COMPASS force field. The usefulness of the TMFF was further assessed in predicting SS-NMR chemical shifts. Little improvement is identified when the TMFF, a better force field, was applied. The improvement given by the accurate motional averaging was limited by both the lack of sensitivity of isotropic chemical shifts and the insufficient accuracy of the SS-NMR prediction method.

Chapter 8

Future perspectives

Extending the application of MD simulations with TMFFs

MD simulations have been applied for investigating the dynamics of molecular solids such as crystallisation, solid-solid phase transitions,^{149,150} melting points¹⁵¹ and amorphous state.¹⁵² By parameterising more TMFFs and validating the simulations against experimental results, the capabilities of MD simulations with TMFFs can be revealed in these fields.

The use of salt formation, co-crystals and hydrates of active pharmaceutical ingredients (APIs) are very common in the drug development phase. It is valuable if their properties can be reproduced, interpreted and predicted by computational approaches such as MD simulations. Recently, it has been shown that the dehydration of ampicillin trihydrate can be studied by MD simulations.¹⁵³ However, there was an inconsistency between the energy minima given by the COMPASS force field and the DFT-D method such that the COMPASS force field could not locate the anhydrate form that was determined from experiments. Applying MD simulations with TMFFs to study the dehydration process is appealing because TMFFs provide a better representation for the energy profiles given by high-quality quantum mechanical methods.

Benchmarking the RMSD expectations for dynamical averaging in SS-NMR calculations

In this thesis, the RMSD expectations of ¹³C and ¹H chemical shifts were adopted from studies based on static DFT structures. The errors, in principle, vary based on the dynamics and the methods that are applied for molecular crystals.^{154,155} Therefore, determining expected errors for the motional averaging with force field based MD simulations in NMR calculations is desirable, which can provide a detailed overview of the performance of the force fields quantitatively. This is usually done by testing benchmarking sets of molecular crystals.

Powerful computational and experimental combinations

The combination of CSP, DFT-D calculations and MD simulations has been employed not only in the area of NMR crystallography, but also in the study of the electronic circular dichroism (ECD) spectra.¹⁵⁶ Moreover, DFT and *ab initio* MD simulations were used in corporate with terahertz time-domain spectroscopy to study the anharmonicity of crystalline purine.¹⁵⁷ Applications of these combinations can be extended to other cutting-edge experimental approaches to gain more insight into molecular solids. It is interesting to evaluate the performance of TMFFs for these combinations.

Chapter 9

References

- [1] J. Halebian and W. McCrone, "Pharmaceutical applications of polymorphism," *Journal of Pharmaceutical Sciences*, vol. 58, pp. 911–929, 1969.
- [2] U.S. Food and Drug Administration, "ANDAs: Pharmaceutical solid polymorphism. Chemistry, manufacturing, and controls information," 2007. Link: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072866.pdf>. Date accessed: 16 October 2016.
- [3] S. R. Byrn, R. R. Pfeiffer, and J. G. Stowell, *Solid-state chemistry of drugs*. West Lafayette, Indiana: SSCI, Inc., second ed., 1999.
- [4] D. J. Kempf, K. C. Marsh, J. F. Denissen, E. McDonald, S. Vasavanonda, C. A. Flentge, B. E. Green, L. Fino, C. H. Park, X.-P. Kong, N. E. Wideburg, A. Saldivar, L. Ruiz, W. M. Kati, H. L. Sham, T. Robins, K. D. Stewart, A. Hsu, J. J. Plattner, J. M. Leonard, and D. W. Norbeck, "ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, pp. 2484–2488, 1995.
- [5] S. R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, S. Spanton, W. Dziki, W. Porter, J. Quick, P. Bauer, J. Donaubauer, B. A. Narayanan, M. Soldani, D. Riley, and K. McFarland, "Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development," *Organic Process Research & Development*, vol. 4, pp. 413–417, 2000.
- [6] D.-K. Bučar, R. W. Lancaster, and J. Bernstein, "Disappearing polymorphs revisited," *Angewandte Chemie International Edition*, vol. 54, pp. 6972–6993, 2015.
- [7] A. Mullard, "2015 FDA drug approvals," *Nature Reviews Drug Discovery*, vol. 15, pp. 73–76, 2016.
- [8] U.S. Food and Drug Administration, "Novel drug approvals for 2016," 2016. Link: <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/ucm483775.htm>. Date accessed: 19 December 2016.
- [9] L. Z. Benet, C.-Y. Wu, and J. M. Custodio, "Predicting drug absorption and the effects of food on oral bioavailability," in *Bulletin Technique Gattefossé 2006 - 40th anniversary: Influence of lipid excipients on oral drug absorption*, pp. 9–16, 2006.

- [10] G. L. Amidon, H. Lennernäs, V. P. Shah, and J. R. Crison, "A theoretical basis for a bio-pharmaceutic drug classification: the correlation of *in vitro* drug product dissolution and *in vivo* bioavailability," *Pharmaceutical Research*, vol. 12, pp. 413–420, 1995.
- [11] R. G. Hill and H. P. Rang, "Drug development: introduction," in *Drug Discovery and Development: Technology in Transition* (R. G. Hill and H. P. Rang, eds.), ch. 14, pp. 203–210, Churchill Livingstone, second ed., 2013.
- [12] J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris, "Ritonavir: An extraordinary example of conformational polymorphism," *Pharmaceutical Research*, vol. 18, pp. 859–866, 2001.
- [13] Y. A. Abramov, "Computational pharmaceutical solid-state chemistry: an introduction," in *Computational pharmaceutical solid-state chemistry* (Y. A. Abramov, ed.), ch. 1, pp. 1–13, John Wiley & Sons, Inc., 2016.
- [14] J. Rantanen and J. Khinast, "The future of pharmaceutical manufacturing sciences," *Journal of Pharmaceutical Sciences*, vol. 104, pp. 3612–3638, 2015.
- [15] F. Jensen, "Optimization techniques," in *Introduction to Computational Chemistry*, ch. 12, pp. 380–420, Chichester: John Wiley & Sons Ltd., second ed., 2007.
- [16] A. D. Becke, "Perspective: Fifty years of density-functional theory in chemical physics," *The Journal of Chemical Physics*, vol. 140, p. 18A301, 2014.
- [17] Nobel Media AB, "The Nobel prize in chemistry 1998," 1998. Link: https://www.nobelprize.org/nobel_prizes/chemistry/laureates/1998/. Date accessed: 17 October 2016.
- [18] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical Review A*, vol. 140, pp. 1133–1138, 1965.
- [19] A. D. Becke, "Completely numerical calculations on diatomic molecules in the local-density approximation," *Physical Review A*, vol. 33, pp. 2786–2788, 1986.
- [20] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Physical Review A*, vol. 38, pp. 3098–3100, 1988.
- [21] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B*, vol. 37, pp. 785–789, 1988.
- [22] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, "Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation," *Physical Review B*, vol. 46, pp. 6671–6687, 1992.

CHAPTER 9. REFERENCES

- [23] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Physical Review Letters*, vol. 77, pp. 3865–3868, 1996.
- [24] A. D. Becke, "A new mixing of Hartree-Fock and local density-functional theories," *The Journal of Chemical Physics*, vol. 98, pp. 1372–1377, 1993.
- [25] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *The Journal of Chemical Physics*, vol. 98, pp. 5648–5652, 1993.
- [26] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields," *The Journal of Physical Chemistry*, vol. 98, pp. 11623–11627, 1994.
- [27] K. Kim and K. D. Jordan, "Comparison of density functional and MP2 calculations on the water monomer and dimer," *The Journal of Physical Chemistry*, vol. 98, pp. 10089–10094, 1994.
- [28] P. J. Hasnip, K. Refson, M. I. J. Probert, J. R. Yates, S. J. Clark, and C. J. Pickard, "Density functional theory in the solid state," *Philosophical transactions. Series A, Mathematical, Physical, and Engineering Sciences*, vol. 372, p. 20130270, 2014.
- [29] D. E. Williams, "Crystal packing of molecules," *Science*, vol. 147, pp. 605–606, 1965.
- [30] S. Kristyán and P. Pulay, "Can (semi)local density functional theory account for the London dispersion forces?," *Chemical Physics Letters*, vol. 229, pp. 175–180, 1994.
- [31] J. M. Pérez-Jordá, E. San-Fabián, and A. J. Pérez-Jiménez, "Density-functional study of van der Waals forces on rare-gas diatomics: Hartree-Fock exchange," *The Journal of Chemical Physics*, vol. 110, pp. 1916–1920, 1999.
- [32] D. C. Patton and M. R. Pederson, "Application of the generalized-gradient approximation to rare gas dimers," *Physical Review A*, vol. 56, pp. R2495–R2498, 1997.
- [33] E. F. C. Byrd, G. E. Scuseria, and C. F. Chabalowski, "An ab initio study of solid nitromethane, HMX, RDX, and CL20: Successes and failures of DFT," *Journal of Physical Chemistry B*, vol. 108, pp. 13100–13106, 2004.
- [34] M. A. Neumann and M.-A. Perrin, "Energy ranking of molecular crystals using density functional theory calculations and an empirical van der waals correction," *The Journal of Physical Chemistry B*, vol. 109, pp. 15531–15541, 2005.
- [35] Q. Wu and W. Yang, "Empirical correction to density functional theory for van der Waals interactions," *Journal of Chemical Physics*, vol. 116, pp. 515–524, 2002.
- [36] S. Grimme, "Density functional theory with London dispersion corrections," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, pp. 211–228, 2011.

- [37] S. Grimme, “Semiempirical GGA-type density functional constructed with a long-range dispersion correction,” *Journal of Computational Chemistry*, vol. 27, pp. 1787–1799, 2006.
- [38] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, “A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu,” *The Journal of Chemical Physics*, vol. 132, p. 154104, 2010.
- [39] A. Otero-de-la Roza and E. R. Johnson, “A benchmark for non-covalent interactions in solids,” *The Journal of Chemical Physics*, vol. 137, p. 54103, 2012.
- [40] J. Moellmann and S. Grimme, “DFT-D3 study of some molecular crystals,” *The Journal of Physical Chemistry C*, vol. 118, pp. 7615–7621, 2014.
- [41] N. W. Ashcroft and N. D. Mermin, “Electron levels in a periodic potential: General properties,” in *Solid State Physics*, ch. 8, pp. 131–150, Cengage Learning, 1976.
- [42] H. J. Monkhorst and J. D. Pack, “Special points for Brillouin-zone integrations,” *Physical Review B*, vol. 13, pp. 5188–5192, 1976.
- [43] D. Vanderbilt, “Soft self-consistent pseudopotentials in a generalized eigenvalue formalism,” *Physical Review B*, vol. 41, pp. 7892–7895, 1990.
- [44] J. S. Lin, A. Qteish, M. C. Payne, and V. Heine, “Optimized and transferable nonlocal separable ab initio pseudopotentials,” *Physical Review B*, vol. 47, pp. 4174–4180, 1993.
- [45] M.-H. Lee, *Advanced pseudopotentials for large scale electronic structure calculations*. PhD thesis, University of Cambridge, 1996.
- [46] J. R. Yates, C. J. Pickard, and F. Mauri, “Calculation of NMR chemical shifts for extended systems using ultrasoft pseudopotentials,” *Physical Review B*, vol. 76, p. 024401, 2007.
- [47] K. Umemoto, R. M. Wentzcovitch, and P. B. Allen, “Dissociation of MgSiO₃ in the cores of gas giants and terrestrial exoplanets,” *Science*, vol. 311, pp. 983–986, 2006.
- [48] J. Li, O. Sode, G. A. Voth, and S. Hirata, “A solid-solid phase transition in carbon dioxide at high pressures and intermediate temperatures,” *Nature Communications*, vol. 4, p. 2647, 2013.
- [49] F. Zhang, M. Hayashi, H.-W. Wang, K. Tominaga, O. Kambara, J. Nishizawa, and T. Sasaki, “Terahertz spectroscopy and solid-state density functional theory calculation of anthracene: Effect of dispersion force on the vibrational modes,” *The Journal of Chemical Physics*, vol. 140, p. 174509, 2014.

CHAPTER 9. REFERENCES

- [50] O. Sode, M. Keçeli, S. Hirata, and K. Yagi, "Coupled-cluster and many-body perturbation study of energies, structures, and phonon dispersions of solid hydrogen fluoride," *International Journal of Quantum Chemistry*, vol. 109, pp. 1928–1939, 2009.
- [51] S. L. Price, "Predicting crystal structures of organic compounds," *Chemical Society Reviews*, vol. 43, pp. 2098–2111, 2014.
- [52] G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B. P. van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. A. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta, P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, J. van de Streek, A. K. Wolf, and B. Schweizer, "Significant progress in predicting the crystal structures of small organic molecules—a report on the fourth blind test," *Acta Crystallographica Section B*, vol. 65, pp. 107–125, 2009.
- [53] M. A. Neumann, F. J. J. Leusen, and J. Kendrick, "A major advance in crystal structure prediction," *Angewandte Chemie International Edition*, vol. 47, pp. 2427–2430, 2008.
- [54] D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Desiraju, B. P. van Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta, S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. van de Streek, T. S. Thakur, S. Tiwari, E. Venuti, and I. K. Zhitkov, "Towards crystal structure prediction of complex organic compounds - A report on the fifth blind test," *Acta Crystallographica Section B*, vol. 67, pp. 535–551, 2011.
- [55] A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, D. H. Case, R. Chadha, J. C. Cole, K. Cosburn, H. M. Cuppen, F. Curtis, G. M. Day, R. A. DiStasio Jr, A. Dzyabchenko, B. P. van Eijck, D. M. Elking, J. A. van den Ende, J. C. Facelli, M. B. Ferraro, L. Fusti-Molnar, C.-A. Gatsiou, T. S. Gee, R. de Gelder, L. M. Ghiringhelli, H. Goto, S. Grimme, R. Guo, D. W. Hofmann, J. Hoja, R. K. Hylton, L. Iuzzolino, W. Jankiewicz, D. T. de Jong, J. Kendrick, N. J. de Klerk, H.-Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. J. Leusen, A. M. Lund, J. Lv, Y. Ma, N. Marom, A. E. Masunov, P. McCabe, D. P. McMahon, H. Meekes, M. P. Metz, A. J. Misquitta, S. Mohamed, B. Monserrat, R. J. Needs, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J. Pickard, R. Podeszwa, L. S. Price, S. L. Price, A. Pulido, M. G. Read, K. Reuter, E. Schneider, C. Schober, G. P. Shields, P. Singh, I. J. Sugden, K. Szaleqicz, C. R. Taylor, A. Tkatchenko, M. E. Tuckerman, F. Vacarro, M. Vasileiadis,

- A. Vazquez-Mayagoitia, L. Vogt, Y. Wang, R. E. Watson, G. A. de Wijs, J. Yang, Q. Zhu, and C. R. Groom, "Report on the sixth blind test of organic crystal-structure prediction methods," *Acta Crystallographica Section B*, vol. 72, pp. 439–459, 2016.
- [56] J. Kendrick, F. J. J. Leusen, M. A. Neumann, and J. van de Streek, "Progress in crystal structure prediction," *Chemistry - A European Journal*, vol. 17, pp. 10736–10744, 2011.
- [57] P. G. Karamertzanis and C. C. Pantelides, "Ab initio crystal structure prediction. II. Flexible molecules," *Molecular Physics*, vol. 105, pp. 273–291, 2007.
- [58] G. M. Day, W. D. S. Motherwell, and W. Jones, "A strategy for predicting the crystal structures of flexible molecules: the polymorphism of phenobarbital," *Physical Chemistry Chemical Physics*, vol. 9, pp. 1693–1704, 2007.
- [59] T. G. Cooper, K. E. Hejczyk, W. Jones, and G. M. Day, "Molecular polarization effects on the relative energies of the real and putative crystal structures of valine," *Journal of Chemical Theory and Computation*, vol. 4, pp. 1795–1805, 2008.
- [60] S. L. Price, M. Leslie, G. W. A. Welch, M. Habgood, L. S. Price, P. G. Karamertzanis, and G. M. Day, "Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials," *Physical Chemistry Chemical Physics*, vol. 12, pp. 8478–8490, 2010.
- [61] A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, and C. C. Pantelides, "Efficient handling of molecular flexibility in lattice energy minimization of organic crystals," *Journal of Chemical Theory and Computation*, vol. 7, pp. 1998–2016, 2011.
- [62] J. Nyman and G. M. Day, "Static and lattice vibrational energy differences between polymorphs," *CrystEngComm*, vol. 17, pp. 5154–5165, 2015.
- [63] D. H. Case, J. E. Campbell, P. J. Bygrave, and G. M. Day, "Convergence properties of crystal structure prediction by quasi-random sampling," *Journal of Chemical Theory and Computation*, vol. 12, pp. 910–924, 2016.
- [64] A. V. Kazantsev, P. G. Karamertzanis, C. C. Pantelides, and C. S. Adjiman, "CrystalOptimizer: An efficient algorithm for lattice energy minimization of organic crystals using isolated-molecule quantum mechanical calculations," in *Process Systems Engineering*, vol. 6-7, pp. 1–42, Wiley-VCH Verlag GmbH & Co. KGaA, 2011.
- [65] M. Habgood, I. J. Sugden, A. V. Kazantsev, C. S. Adjiman, and C. C. Pantelides, "Efficient handling of molecular flexibility in ab initio generation of crystal structures," *Journal of Chemical Theory and Computation*, vol. 11, pp. 1957–1969, 2015.
- [66] A. V. Kazantsev, P. G. Karamertzanis, C. C. Pantelides, and C. S. Adjiman, "Ab initio crystal structure prediction for flexible molecules," in *Computer Aided Chemical Engineering*, vol. 28, pp. 817–822, Elsevier B.V., 2010.

CHAPTER 9. REFERENCES

- [67] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, “Ab initio molecular simulations with numeric atom-centered orbitals,” *Computer Physics Communications*, vol. 180, pp. 2175–2196, 2009.
- [68] A. Tkatchenko and M. Scheffler, “Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data,” *Physical Review Letters*, vol. 102, p. 073005, 2009.
- [69] A. Tkatchenko, R. A. DiStasio, R. Car, and M. Scheffler, “Accurate and efficient method for many-body van der Waals interactions,” *Physical Review Letters*, vol. 108, no. 23, p. 236402, 2012.
- [70] A. Ambrosetti, A. M. Reilly, R. A. DiStasio, and A. Tkatchenko, “Long-range correlation energy calculated from coupled atomic response functions,” *The Journal of Chemical Physics*, vol. 140, p. 18A508, 2014.
- [71] M. Baias, J.-N. Dumez, P. H. Svensson, S. Schantz, G. M. Day, and L. Emsley, “De novo determination of the crystal structure of a large drug molecule by crystal structure prediction-based powder NMR crystallography,” *Journal of the American Chemical Society*, vol. 135, pp. 17501–17507, 2013.
- [72] M. A. Neumann, J. van de Streek, F. P. A. Fabbiani, P. Hidber, and O. Grassmann, “Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening,” *Nature Communications*, vol. 6, p. 7793, 2015.
- [73] J. van de Streek and M. A. Neumann, “Validation of experimental molecular crystal structures with dispersion-corrected density functional theory calculations,” *Acta crystallographica Section B*, vol. 66, pp. 544–558, 2010.
- [74] J. van de Streek and M. A. Neumann, “Validation of molecular crystal structures from powder diffraction data with dispersion-corrected density functional theory (DFT-D),” *Acta Crystallographica Section B*, vol. 70, pp. 1020–1032, 2014.
- [75] L. Brooks, M. Brunelli, P. Pattison, G. R. Jones, and A. Fitch, “Crystal structures of eight mono-methyl alkanes ($C_{26}C_{32}$) via single-crystal and powder diffraction and DFT-D optimization,” *IUCrJ*, vol. 2, pp. 490–497, 2015.
- [76] J. Brüning and M. U. Schmidt, “The determination of crystal structures of active pharmaceutical ingredients from X-ray powder diffraction data: A brief, practical introduction, with fexofenadine hydrochloride as example,” *Journal of Pharmacy and Pharmacology*, vol. 67, pp. 773–781, 2015.
- [77] D. Frenkel and B. Smit, “Molecular dynamics simulations,” in *Understanding Molecular Simulation*, ch. 4, pp. 63–107, Academic Press, second ed., 2002.

- [78] H. Sun, "COMPASS: An ab initio force-Field optimized for condensed-phase applications - Overview with details on alkane and benzene compounds," *The Journal of Physical Chemistry B*, vol. 102, pp. 7338–7364, 1998.
- [79] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *Journal of the American Chemical Society*, vol. 117, pp. 5179–5197, 1995.
- [80] W. L. Jorgensen and J. Tirado-Rives, "The OPLS potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin," *Journal of the American Chemical Society*, vol. 110, pp. 1657–1666, 1988.
- [81] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-Atom empirical potential for molecular modeling and dynamics studies of proteins," *The Journal of Physical Chemistry B*, vol. 102, pp. 3586–3616, 1998.
- [82] A. Nemkevich, H.-B. Bürgi, M. A. Spackman, and B. Corry, "Molecular dynamics simulations of structure and dynamics of organic molecular crystals," *Physical Chemistry Chemical Physics*, vol. 12, pp. 14916–14929, 2010.
- [83] J. Nyman, O. S. Pundyke, and G. M. Day, "Accurate force fields and methods for modelling organic molecular crystals at finite temperatures," *Physical Chemistry Chemical Physics*, vol. 18, pp. 15828–15837, 2016.
- [84] M. A. Neumann, "Tailor-made force fields for crystal-structure prediction," *The Journal of Physical Chemistry B*, vol. 112, pp. 9810–9829, 2008.
- [85] S. L. Mayo, B. D. Olafson, and W. A. Goddard III, "DREIDING: A generic force field for molecular simulations," *Journal of Physical Chemistry*, vol. 101, pp. 8897–8909, 1990.
- [86] D. Marx and J. Hutter, "Getting started: Unifying molecular dynamics and electronic structure," in *Ab initio molecular dynamics. Basic theory and advanced methods*, ch. 2, pp. 11–84, Cambridge University Press, 2009.
- [87] R. K. Harris, R. E. Wasylissen, and M. J. Duer, *NMR crystallography*. Chichester, England: John Wiley & Sons, 2009.
- [88] K. E. Dempah, D. H. Barich, A. M. Kaushal, Z. Zong, S. D. Desai, R. Suryanarayanan, L. Kirsch, and E. J. Munson, "Investigating gabapentin polymorphism using solid-state NMR spectroscopy," *AAPS PharmSciTech*, vol. 14, pp. 19–28, 2013.

CHAPTER 9. REFERENCES

- [89] K. D. M. Harris, C. E. Hughes, and P. A. Williams, "Monitoring the evolution of crystallization processes by in-situ solid-state NMR spectroscopy," *Solid State Nuclear Magnetic Resonance*, vol. 65, pp. 107–113, 2015.
- [90] S. K. Lee, "Effect of pressure on structure of oxide glasses at high pressure: Insights from solid-state NMR of quadrupolar nuclides," *Solid State Nuclear Magnetic Resonance*, vol. 38, pp. 45–57, 2010.
- [91] C. Bonhomme, C. Gervais, F. Babonneau, C. Coelho, F. Pourpoint, T. Azaïs, S. E. Ashbrook, J. M. Griffin, J. R. Yates, F. Mauri, and C. J. Pickard, "First-principles calculation of NMR parameters using the gauge including projector augmented wave method: a chemist's point of view," *Chemical Reviews*, vol. 112, pp. 5733–5779, 2012.
- [92] C. G. van de Walle and P. E. Blöchl, "First-principles calculations of hyperfine parameters," *Physical Review B*, vol. 47, pp. 4244–4255, 1993.
- [93] C. J. Pickard and F. Mauri, "All-electron magnetic response with pseudopotentials: NMR chemical shifts," *Physical Review B*, vol. 63, p. 245101, 2001.
- [94] E. Salager, G. M. Day, R. S. Stein, C. J. Pickard, B. Elena, and L. Emsley, "Powder crystallography by combined crystal structure prediction and high-resolution ^1H solid-state NMR spectroscopy," *Journal of the American Chemical Society*, vol. 132, pp. 2564–2566, 2010.
- [95] C. M. Widdifield, H. Robson, and P. Hodgkinson, "Furosemide's one little hydrogen atom: NMR crystallography structure verification of powdered molecular organics," *Chemical Communications*, vol. 52, pp. 6685–6688, 2016.
- [96] J. Elguero, "Tautomerism: A historical perspective," in *Tautomerism*, pp. 1–10, Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2016.
- [97] T. Sugawara and I. Takasu, "Tautomerism in the solid state," in *Advances in Physical Organic Chemistry*, vol. 32, pp. 219–265, Elsevier B.V., 1999.
- [98] B. Kuhn, P. Mohr, and M. Stahl, "Intramolecular hydrogen bonding in medicinal chemistry," *Journal of Medicinal Chemistry*, vol. 53, pp. 2601–2611, 2010.
- [99] A. Hangan, G. Borodi, X. Filip, C. Tripon, C. Morari, L. Oprean, and C. Filip, "Structure of N-(5-ethyl-[1,3,4]-thiadiazole-2-yl)toluenesulfonamide by combined X-ray powder diffraction, ^{13}C solid-state NMR and molecular modelling," *Acta Crystallographica Section B*, vol. 66, pp. 615–621, 2010.
- [100] G. Alzuet, J. Borrás, F. Estevan, M. Liu-González, and F. Sanz-Ruiz, "Zinc complexation to *N*-substituted sulfonamide ligands. Preparation, properties and crystal structure of copper(II) doped $\{[\text{Zn}(\text{sulfamethizolate})_2(\text{py})]\cdot\text{H}_2\text{O}\}_{\infty}$," *Inorganica Chimica Acta*, vol. 343, pp. 56–60, 2003.

- [101] W. Herbst, K. Hunger, G. Wilker, H. Ohleier, and R. Winter, "Polycyclic Pigments," in *Industrial Organic Pigments*, pp. 421–566, Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2004.
- [102] H. Sakai, H. Ando, H. Murata, S. Aoki, and I. Oshiumi, "Organic pigments (II) yellow organic pigments," *Journal of the Japan Society of Colour Material*, vol. 55, no. 9, pp. 677–689, 1982.
- [103] J. Dehnert, "Gelbe Farbstoffe der Phthalimidochinophthalonreihe," 1968. Patent DE 000001770960 A, BASF AG, Germany.
- [104] J. Dehnert, "Gelbe Farbstoffe der Phthalimidochinophthalonreihe, deren Herstellung und deren Verwendung," 1968. Patent DE 000001770960 B2, BASF AG, Germany.
- [105] J. Dehnert, "Gelbe Farbstoffe der Phthalimidochinophthalonreihe, deren Herstellung und deren Verwendung," 1968. Patent DE 000001770960 C3, BASF AG, Germany.
- [106] G. Kuth and B. Ort, "Verfahren zur Herstellung von Chinophthalonen," 1990. Patent DE 000004020423 A1, BASF AG, Germany.
- [107] T. Sakagami, A. Hashimoto, and Y. Teraji, "Multicolor-development laser marking sheet for card, and laser marking method," 2013. Patent JP WO2013031399 A1, Japan Coloring Co. Ltd., Japan.
- [108] C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward, "The Cambridge structural database," *Acta Crystallographica Section B*, vol. 72, pp. 171–179, 2016.
- [109] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, "First principles methods using CASTEP," *Zeitschrift für Kristallographie*, vol. 220, pp. 567–570, 2005.
- [110] W. I. F. David, K. Shankland, J. van de Streek, E. Pidcock, W. D. S. Motherwell, and J. C. Cole, "DASH: A program for crystal structure determination from powder diffraction data," *Journal of Applied Crystallography*, vol. 39, pp. 910–915, 2006.
- [111] X. Li, A. D. Bond, K. E. Johansson, and J. van de Streek, "Distinguishing tautomerism in the crystal structure of (Z)-N-(5-ethyl-2,3-dihydro-1,3,4-thiadiazol-2-ylidene)-4-methylbenzenesulfonamide using DFT-D calculations and ^{13}C solid-state NMR," *Acta Crystallographica Section C*, vol. 70, pp. 784–789, 2014.
- [112] B. Santra, J. Klimeš, D. Alfè, A. Tkatchenko, B. Slater, A. Michaelides, R. Car, and M. Scheffler, "Hydrogen bonds and van der Waals forces in ice at ambient and high pressures," *Physical Review Letters*, vol. 107, p. 185701, 2011.
- [113] G. Kresse and J. Hafner, "Ab initio molecular dynamics for liquid metals," *Physical Review B*, vol. 47, pp. 558–561, 1993.

CHAPTER 9. REFERENCES

- [114] G. Kresse and J. Furthmüller, "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set," *Computational Materials Science*, vol. 6, pp. 15–50, 1996.
- [115] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Physical Review B*, vol. 54, pp. 11169–11186, 1996.
- [116] G. Kresse, "From ultrasoft pseudopotentials to the projector augmented-wave method," *Physical Review B*, vol. 59, pp. 1758–1775, 1999.
- [117] G. Gilli, F. Bellucci, V. Ferretti, and V. Bertolasi, "Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the β -diketone fragment," *Journal of the American Chemical Society*, vol. 111, pp. 1023–1028, 1989.
- [118] M. Dračínský, L. Čechová, P. Hodgkinson, E. Procházková, and Z. Janeba, "Resonance-assisted stabilisation of hydrogen bonds probed by NMR spectroscopy and path integral molecular dynamics," *Chemical Communications*, vol. 51, no. 73, pp. 13986–13989, 2015.
- [119] F. G. Vogt, "Characterization of pharmaceutical compounds by solid-state NMR," *eMagRes*, vol. 4, pp. 255–268, 2015.
- [120] J. P. Amoureaux, M. Bee, and J. C. Damien, "Structure of adamantane, $C_{10}H_{16}$, in the disordered phase," *Acta Crystallographica Section B*, vol. 36, pp. 2633–2636, 1980.
- [121] J. Krawczyk, J. Mayer, I. Natkaniec, M. Nowina Konopka, O. Steinsvoll, and J. A. Janik, "Quasi-elastic (QENS) and inelastic neutron scattering (INS) on hexamethylbenzene," *Physica B: Condensed Matter*, vol. 362, pp. 271–277, 2005.
- [122] W. L. Earl and D. L. Vanderhart, "Measurement of ^{13}C chemical shifts in solids," *Journal of Magnetic Resonance*, vol. 48, pp. 35–54, 1982.
- [123] E. V. Boldyreva, V. A. Drebushchak, T. N. Drebushchak, I. E. Paukov, Y. A. Kovalevskaya, and E. S. Shutova, "Polymorphism of glycine, Part I," *Journal of Thermal Analysis and Calorimetry*, vol. 73, pp. 409–418, 2003.
- [124] D. H. Barich, E. M. Gorman, M. T. Zell, and E. J. Munson, "3-Methylglutaric acid as a ^{13}C solid-state NMR standard," *Solid State Nuclear Magnetic Resonance*, vol. 30, pp. 125–129, 2006.
- [125] G. M. Sheldrick, "A short history of *SHELX*," *Acta Crystallographica Section A*, vol. 64, pp. 112–122, 2008.
- [126] L. O. Brockway and J. M. Robertson, "The crystal structure of hexamethylbenzene and the length of the methyl group bond to aromatic carbon atoms," *Journal of the Chemical Society*, pp. 1324–1332, 1939.

- [127] P. Langan, S. A. Mason, D. Myles, and B. P. Schoenborn, "Structural characterization of crystals of α -glycine during anomalous electrical behaviour," *Acta Crystallographica Section B*, vol. 58, pp. 728–733, 2002.
- [128] E. V. Boldyreva, T. N. Drebushchak, and E. S. Shutova, "Structural distortion of the α , β , and γ polymorphs of glycine on cooling," *Zeitschrift für Kristallographie*, vol. 218, pp. 366–376, 2003.
- [129] J. P. Amoureaux and M. Foulon, "Comparison between structural analyses of plastic and brittle crystals," *Acta Crystallographica Section B*, vol. 43, pp. 470–479, 1987.
- [130] P. L. Maguérès, S. V. Lindeman, and J. K. Kochi, "Electron redistribution of aromatic ligands in (Arene)Cr(CO)₃ complexes. Structural (bond-length) changes as quantitative measures," *Organometallics*, vol. 20, pp. 115–125, 2001.
- [131] R. E. Taylor, "Setting up ¹³C CP/MAS experiments," *Concepts in Magnetic Resonance*, vol. 22A, pp. 37–49, 2004.
- [132] R. K. Harris, P. Hodgkinson, C. J. Pickard, J. R. Yates, and V. Zorin, "Chemical shift computations on a crystallographic basis: some reflections and comments," *Magnetic Resonance in Chemistry*, vol. 45, pp. S174–S186, 2007.
- [133] J. C. Facelli, "Chemical shift tensors: theory and application to molecular structural problems," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 58, pp. 176–201, 2011.
- [134] J. F. Casale, "A practical total synthesis of cocaine's enantiomers," *Forensic Science International*, vol. 33, pp. 275–298, 1987.
- [135] R. J. Hryncuk, R. J. Barton, and B. E. Robertson, "The crystal structure of free base cocaine, C₁₇H₂₁NO₄," *Canadian Journal of Chemistry*, vol. 61, pp. 481–487, 1983.
- [136] M. Baias, C. M. Widdifield, J.-N. Dumez, H. P. G. Thompson, T. G. Cooper, E. Salager, S. Bassil, R. S. Stein, A. Lesage, G. M. Day, and L. Emsley, "Powder crystallography of pharmaceutical materials by combined crystal structure prediction and solid-state ¹H NMR spectroscopy," *Physical Chemistry Chemical Physics*, vol. 15, pp. 8069–8080, 2013.
- [137] J. Brus, J. Czernek, L. Kobera, M. Urbanova, S. Abbrent, and M. Husak, "Predicting the crystal structure of decitabine by powder NMR crystallography: influence of long-range molecular packing symmetry on NMR parameters," *Crystal Growth & Design*, vol. 16, pp. 7102–7111, 2016.
- [138] H.-J. Schneider and L. Sturm, "Conformations and N-inversion barriers in tropane compounds," *Angewandte Chemie International Edition*, vol. 15, pp. 545–546, 1976.

CHAPTER 9. REFERENCES

- [139] F. Liu, C. G. Phung, D. W. Alderman, and D. M. Grant, "Carbon-13 chemical shift tensors in *meso*-erythritol, measuring OH dihedral angles," *Journal of the American Chemical Society*, vol. 117, pp. 9323–9328, 1995.
- [140] J. D. Hartman and G. J. O. Beran, "Fragment-Based electronic structure approach for computing nuclear magnetic resonance chemical shifts in molecular crystals," *Journal of Chemical Theory and Computation*, vol. 10, pp. 4862–4872, 2014.
- [141] S. T. Holmes, R. J. Iuliucci, K. T. Mueller, and C. Dybowski, "Critical analysis of cluster models and exchange-correlation functionals for calculating magnetic shielding in molecular solids," *Journal of Chemical Theory and Computation*, vol. 11, pp. 5229–5241, 2015.
- [142] J. D. Hartman, R. A. Kudla, G. M. Day, L. J. Mueller, and G. J. O. Beran, "Benchmark fragment-based ^1H , ^{13}C , ^{15}N and ^{17}O chemical shift predictions in molecular crystals," *Physical Chemistry Chemical Physics*, vol. 18, pp. 21686–21709, 2016.
- [143] J. D. Hartman, G. M. Day, and G. J. O. Beran, "Enhanced NMR discrimination of pharmaceutically relevant molecular crystal forms through fragment-based ab initio chemical shift predictions," *Crystal Growth & Design*, vol. 16, pp. 6479–6493, 2016.
- [144] R. E. London, B. D. Wingad, and G. A. Mueller, "Dependence of amino acid side chain ^{13}C shifts on dihedral angle: Application to conformational analysis," *Journal of the American Chemical Society*, vol. 130, pp. 11097–11105, 2008.
- [145] F. A. A. Mulder, "Leucine side-chain conformation and dynamics in proteins from ^{13}C NMR chemical shifts," *ChemBioChem*, vol. 10, pp. 1477–1479, 2009.
- [146] A. B. Sahakyan, W. F. Vranken, A. Cavalli, and M. Vendruscolo, "Structure-based prediction of methyl chemical shifts in proteins," *Journal of Biomolecular NMR*, vol. 50, pp. 331–346, 2011.
- [147] B. Han, Y. Liu, S. W. Ginzinger, and D. S. Wishart, "SHIFTX2: Significantly improved protein chemical shift prediction," *Journal of Biomolecular NMR*, vol. 50, pp. 43–57, 2011.
- [148] M. M. H. Smets, S. J. T. Brugman, E. R. H. van Eck, J. A. van den Ende, H. Meekes, and H. M. Cuppen, "Understanding the solid-state phase transitions of DL-Norleucine: An in situ DSC, microscopy, and solid-state NMR study," *Crystal Growth & Design*, vol. 15, pp. 5157–5167, 2015.
- [149] J. A. van den Ende and H. M. Cuppen, "Dynamics of the α and β polymorphs of DL-norleucine at different temperatures: sliding to a partial phase transition," *Crystal Growth & Design*, vol. 14, pp. 3343–3351, 2014.

- [150] J. Kawano, A. Miyake, N. Shimobayashi, and M. Kitamura, "Molecular dynamics simulation of the phase transition between calcite and CaCO₃-II," *Journal of Physics: Condensed Matter*, vol. 21, p. 275403, 2009.
- [151] S. W. Watt, J. A. Chisholm, W. Jones, and S. Motherwell, "A molecular dynamics simulation of the melting points and glass transition temperatures of myo- and neo-inositol," *Journal of Chemical Physics*, vol. 121, pp. 9565–9573, 2004.
- [152] T. X. Xiang and B. D. Anderson, "Molecular dynamics simulation of amorphous indomethacin," *Molecular Pharmaceutics*, vol. 10, pp. 102–114, 2013.
- [153] A. S. Larsen, J. Rantanen, and K. E. Johansson, "Computational dehydration of crystalline hydrates using molecular dynamics simulations," *Journal of Pharmaceutical Sciences*, vol. 106, pp. 348–355, 2017.
- [154] J. C. Johnston, R. J. Iuliucci, J. C. Facelli, G. Fitzgerald, and K. T. Mueller, "Intermolecular shielding contributions studied by modeling the ¹³C chemical-shift tensors of organic single crystals with plane waves," *The Journal of Chemical Physics*, vol. 131, p. 144503, 2009.
- [155] J.-N. Dumez and C. J. Pickard, "Calculation of NMR chemical shifts in organic solids: accounting for motional effects," *The Journal of Chemical Physics*, vol. 130, p. 104701, 2009.
- [156] C. Bannwarth, J. Seibert, and S. Grimme, "Electronic circular dichroism of [16]helicene with simplified TD-DFT: Beyond the single structure approach," *Chirality*, vol. 28, pp. 365–369, 2016.
- [157] M. T. Ruggiero and J. A. Zeitler, "Resolving the origins of crystalline anharmonicity using terahertz time-domain spectroscopy and ab initio simulations," *The Journal of Physical Chemistry B*, vol. 120, pp. 11733–11739, 2016.