

**A COMPUTATIONAL ANALYSIS OF THE  
VIBRATIONAL ABSORPTION OF MOLECULAR  
SOLIDS IN THE TERAHERTZ RANGE**

A Thesis Submitted to the University of Cambridge for the  
Degree of Doctor of Philosophy

by  
**DANIELE TOMERINI**



Queens' college



April 2012



## DECLARATION

---

The work described in this thesis was performed in the Department of Chemistry at the University of Cambridge under the supervision of Dr Graeme Day.

This thesis is the result of my own work except where specific reference is made.

This dissertation is not the same as any other dissertation that I have submitted for a degree, diploma or any other qualifications at any other university. No part of this thesis has been, or is concurrently being submitted for any such degree, diploma or qualification.

In accordance with the regulation of the Degree Committee of the School of Physical Sciences, this thesis does not exceed the specified word limit of 60000 words in length.

*Cambridge, April 2012*

---

Daniele Tomerini



## SUMMARY

---

In this thesis, we deal with the application of transmission terahertz spectroscopy as an analysis tool for the study of molecular solids, in particular organic crystals of pharmaceutical interest. Most of the work has been performed using two computational packages aimed at the interpretation of the spectra, one based on molecular forcefields (DMACRYS), the other on solid state density functional theory (CASTEP).

We compare low temperature determinations of several molecular organic crystals to calculated spectra, and attempt to assign calculated modes of vibrations to absorption peaks, based on the similarity in frequency between the measured and calculated peaks.

One of the main aims of this work is to establish the limits of our forcefield approach, which is based on the approximation that the intramolecular degrees of freedom can be neglected. We analyse the normal modes of vibration calculated with CASTEP, evaluating the amount of rigid molecule rotational and translational contribution to each eigenvector as a function of frequency, in order to validate our forcefield approach. We also compare the two sets of eigenvectors from the DMACRYS and CASTEP calculations to assess the similarity between the two approaches.

We perform the same eigenvectors analysis on several hydrate systems in order to understand the role of water in the lattice dynamics of crystalline hydrates. We attempt a classification of the eigenvectors based on the strength of the forces involved in the molecular vibrations and based on the amount of the water contribution to each normal mode.

A set of isostructural crystals is analysed in order to understand the effect that small variations (in the molecular formula and in the unit cell arrangement) have on the measured and calculated absorption spectra of a crystal.

Finally, we discuss the use and development of computational methods that allow us to have a more realistic description of the molecular electrostatic in DMACRYS.



## ACKNOWLEDGEMENTS

---

*The financial contribution of AstraZeneca is gratefully acknowledged.*

*Many thanks to my supervisor, Dr Greame Day, for introducing me to this fascinating field of research, and for his guidance during the project. Thanks to Greame also for being always available for discussion and for very valuable insight on problems; more importantly, thanks for the corrections of work that I often handled in at the very last minute!*

*Thanks to Dr Axel Zeitler for teaching me how to use the terahertz spectrometer, and for his help in the experimental determinations. Also, thanks for answering my doubts on the experimental methods, and for valuable comments on my work and for his suggestions.*

*Thank you to all of the present and past members of the Material Chemistry Group in Cambridge I had the pleasure to meet, for providing a very pleasant and intellectually stimulating working environment. So long, and thanks for all the fish!*

*Thanks to all my family: my mother, my father, my sister Tiziana, and Daniele, for their continuous support, while I was in England, and when I was in Italy. Grazie di tutto il vostro appoggio!*

*Finally, a special, big thank to Martina, for her continuous encouragement and for making my time in Cambridge so special.*



# CONTENTS

---

DECLARATION III

1 INTRODUCTION 1

- 1.1 Crystal forms: definitions 1
- 1.2 Polymorphism in the pharmaceutical industry 2
- 1.3 Methods of characterisation 6
- 1.4 Peculiarities of terahertz spectroscopy 11
- 1.5 Thesis outline 12

2 TERAHERTZ PULSED SPECTROSCOPY 13

- 2.1 Femtosecond laser sources 14
- 2.2 Electro-optical rectification 16
- 2.3 Experimental setup 18
- 2.4 Temperature effects on the spectra 23

3 VIBRATIONS IN A PERIODIC SYSTEM 27

- 3.1 Single molecule calculations 27
- 3.2 Phonons in a crystal 30
- 3.3 Simple parameterisation for ionic systems 35
- 3.4 Force field methods 38
- 3.5 Solid state electronic structure calculations 44

4 ANALYSIS METHODS 53

- 4.1 The atom-atom calculations 53
- 4.2 The solid state DFT calculations 57
- 4.3 Comparison with the experimental data 60
- 4.4 Rigid-molecule analysis 60

5 MODELLING POLARISATION BY THE CRYSTAL ENVIRONMENT 65

- 5.1 Point charge method 65
- 5.2 The SANTHA script 69

6 SYSTEMS STUDIED 77

6.1	Paracetamol	78
6.2	Theophylline	81
6.3	Nitrofurantoin	82
6.4	Carbamazepine	84
6.5	Benzoic acid	85
6.6	Substituted acetanilides	87
7	EXPERIMENTAL AND CALCULATED TERAHERTZ SPECTRA	91
7.1	Benzoic acid	91
7.2	Anhydrous Theophylline	97
7.3	Theophylline monohydrate	99
7.4	Paracetamol form I	106
7.5	N-phenyl-acetamide derivatives	107
7.6	Nitrofurantoin form $\beta$	107
7.7	Nitrofurantoin hydrate, form II	112
7.8	Carbamazepine form I	114
7.9	Carbamazepine form III	116
7.10	Selected computed spectra	119
7.11	Analysis of the DMACRYS eigenvectors	129
7.12	Conclusions	129
8	EFFECT OF THE RIGIDITY OF THE SYSTEM ON THE SPECTRUM	135
8.1	Flexibility	135
8.2	Systems considered and their rigidity ranking	136
8.3	CASTEP simulations	139
8.4	Eigenvector agreement between CASTEP and DMACRYS	158
8.5	Conclusions	168
9	THE ROLE OF WATER IN THE LATTICE DYNAMICS OF CRYSTALLINE HYDRATES	171
9.1	CASTEP calculations	173
9.2	DMACRYS calculations and analysis	178
9.3	Eigenvector agreement between DMACRYS and CASTEP	197
9.4	Conclusions	200
10	THE EFFECT OF POLARISATION	203
10.1	Dipole moment variations in the crystal	204
10.2	Effects of the full minimisation on the phonon modes	210
10.3	Intensities recalculation	224
10.4	Conclusions	229

11	ISOSTRUCTURAL CRYSTALS	231
11.1	Experimental and calculated spectra	231
11.2	Spectra similarities between the systems	245
11.3	Polarisation effect	246
11.4	Rigidity analysis	249
11.5	Conclusions	254
12	CONCLUSIONS AND OUTLOOK	255
A	VARIABLE TEMPERATURE SPECTRA	261



## INTRODUCTION

---

**T**HIS THESIS DEALS with a recently emerging analysis method, terahertz spectroscopy, and its use as a tool applied to the study of polymorphic molecular solids, in particular organic crystals of pharmaceutical interest.

In this chapter we introduce the concept of polymorphism and its importance for pharmaceutical science, with a practical example. We will therefore briefly consider the techniques and tools used for polymorph analysis and characterisation.

### 1.1 CRYSTAL FORMS: DEFINITIONS

A crystal is a solid material where atoms are arranged in a regular pattern translationally repeated in all three spatial dimensions. The packing arrangement of molecules is the result of the balancing of the intermolecular forces to form an aggregate in equilibrium. Crystal polymorphism (from the greek πολύς = *many* and μορφή = *forms*) is defined as the ability of a compound to exist in more than one crystal species.

Polymorphism is a widespread feature in nature. Allotropism, elemental polymorphism, is encountered in as many as 55 elements [1] and the difference in behaviour between two forms can be as striking as for two of the allotropes of carbon, graphite and diamond. For a small molecule like water (ice in the crystal form) 15 different polymorphs have been discovered up to now, stable under different conditions of temperature and pressure [2].

Solvates, or solvatomorphic forms are defined as crystal systems of a molecule in which the solvent of crystallisation is stoichiometrically included in the crystal lattice. In the common case where water is the solvent, solvates are called hydrates.

Solvates are not equivalent to polymorphs, as their elemental composition is not the same as the corresponding neat forms. The presence of

the solvent inside a crystal can greatly affect the physical properties of a crystal: this is discussed in the later sections.

The existence of more than one crystal form of a molecule is by no means a rare occurrence in organic chemistry: depending on the study [3, 4], it is estimated that 30 % to 50 % of small organic molecules display polymorphism. These numbers are even higher considering solvato-morphic forms, and they range from 56 % to 87 %.

In the following sections we will briefly discuss polymorphism in small organic molecular crystals of pharmaceutical interest: about 80 % of pharmaceutical molecules [5–8] have one or more additional crystal form.

## 1.2 POLYMORPHISM IN THE PHARMACEUTICAL INDUSTRY

Most of the drug products are available as oral dosage forms, under the form of a tablet. Even when the active pharmaceutical ingredient (API) in the final formulation is not a solid, the active molecule itself is usually obtained and isolated as a solid by crystallisation during its preparation[9]. The properties of the solid form of the drug (and of the other ingredients in the formulation, e.g. the excipient) are therefore very important at all stages of life of a drug: its synthesis, its manufacturing, its shelf life.

As we already mentioned, there may be significant variations in the physical properties between solid forms of a molecule (see Table 1.1); these differences are very important, as they can have profound implications in the phases of processing [10, 11], formulation [12], and in the final properties (such as bioavailability) of the marketed drug [13].

It is necessary for the product to be obtained consistently with the same properties, and the manufacturer must be sure that the interaction with the potential solvating agents are known at all stages of preparation. It is also crucial for the consumer that a drug does not convert into another form during its shelf time, as doing so might change its properties: an example is the formation of a hydrate form of a drug conserved in a humid atmosphere (such as a bathroom cabinet).

The regulation for the marketing of a drug is therefore very strict regarding polymorphism [14], and there is a lot of time and resources spent by pharmaceutical companies researching the possibility of polymorphism and the stability and interconversion of each solid form of a

Packing properties	Molar volume and density Refractive index Conductivity: electrical and thermal Hygroscopicity
Thermodynamic properties	Melting and sublimation temperatures Internal or structural energy Enthalpy Heat capacity Entropy Free Energy and Chemical Potential Thermodynamic Activity Vapour Pressure Solubility
Spectroscopic properties	Electronic state transitions Vibrational state transitions Nuclear spin state transitions
Kinetic properties	Dissolution rate Rates of solid-state reactions Stability
Surface properties	Surface free energy Interfacial tensions
Mechanical properties	Crystal habit Hardness Tensile strength Compactibility, tableting Handling, flow, and blending

Table 1.1. Some of the physical properties that may display a significant variation in different crystal forms of a polymorphic system. From [8]

new drug [14, 15].

From the manufacturers' point of view it would be better if the marketed form of the drug were the most stable crystal form, as this would bring the risks of problems during preparation and storage down to a minimum [16]. Unfortunately, the most stable polymorph will also be the one to display the lowest solubility, since its chemical potential is the lowest; the same will also be true for the dissolution rate, another important parameter. Therefore, if the medical properties require it, the formulator may be forced to choose a form that enhances the drug efficiency or (as in the case of Ritonavir discussed in the next section) to opt for another formulation: for example, an amorphous form, a solution or an aerosol.

In some cases other properties of a polymorph may be used to obtain a

more desirable behaviour in tablets. An example is paracetamol: forms I and II have a very similar bioavailability [17], but different mechanical properties. The commercially used, stable form I has poor compression properties [18] that requires the use of binders, while the metastable form II would not be affected by this problem.

### 1.2.1 When things go wrong: the case of Ritonavir

One of the most famous quotes regarding polymorphism is an observation by Walter McCrone [19]:

Those who study polymorphism are rapidly reaching the conclusion that all compounds, organic and inorganic, can crystallize in different forms and polymorphs. In fact, the more diligently any system is studied the larger the number of polymorphs discovered.

A peculiar point of view on McCrone's opinion has been – perhaps provocatively – expressed by Angelo Gavezzotti [20], reversing the argument:

...if no effort and no money are spent, no polymorphs will be found, meaning that most organic substances *do* oblige and crystallize nearly always in the same crystal form, when not bothered by the use of exotic solvents or unusual temperature, pressure, and crystallization conditions.

An answer to this objection can be found in the literature, looking at a very infamous example where polymorphism played a major role.

Ritonavir (see Figure 1.1) is an antiretroviral drug, (an inhibitor of HIV protease), discovered in 1992 [21, 22]. After clinical trials Abbott laboratories filed a new drug application (molecule ABT-538) in 1995, and it was marketed in 1996 as Norvir®. Ritonavir at that time was not found to be polymorphic: only one form was obtained during the preparation tests. Furthermore, no particular screening was supposed to be necessary, since Ritonavir is not bioavailable from the solid state: the two marketed formulations of the drug were in capsules with ethanol/water solutions.

In 1998, however, some batches of Ritonavir capsules failed the dissolution test requirements, and the analysis showed that a new polymorph (form II) was precipitating.

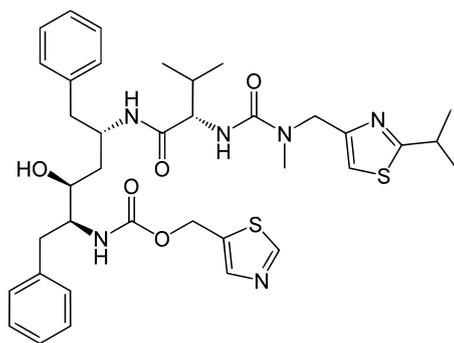


Figure 1.1. The molecule of Ritonavir: IUPAC [5S - (5R\*,8R\*,10R\*,11R\*)]10 - (Hydroxy - 2 - methyl - 5 - (1 - methylethyl) - 1 - [2 - (1 - methyl-ethyl) - 4 - thiazolyl] - 3,6 - di oxo-8,11-bis(phenylmethyl)-2,4,7,12-tetraazatridecan-13-oic acid,5-thiazolylmethyl ester

Form II has a far reduced solubility in ethanol/water solutions compared to form I at the suggested storage temperature of 278 K (see Table 1.2). The supersaturation with respect to form II was therefore the reason behind the failure in the tests.

Furthermore, form II, despite the difficulties of its crystallisation, was found to be the more stable form: within weeks from the first discovery, form II was produced in all the production lines, and the product had to be withdrawn from the market, with obvious problems for the patients (sudden unavailability of a life saving drug [24]) and for the company (withdrawing of a best selling product, and need to restart the development process [25]). Eventually, a new formulation was marketed, after its approval by the Food and Drug Administration [26].

Additional studies on the molecule proved that the difficulties in observing the more stable form II were due to kinetic reasons [23]: the formation of form II in solution requires the molecule to be in a higher energy conformational state. Although form II is thermodynamically favoured, a high energy barrier made the necessary state hard to reach in solution. The precipitation of form II in 1998 was possibly due to seeding agents and/or slightly different and unaccounted variation in the preparation conditions [25].

Ethanol:water ratio	99:1	95:5	90:10	85:15	80:20	75:25
Form I (mg ml <sup>-1</sup> )	90	188	234	294	236	170
Form II (mg ml <sup>-1</sup> )	19	41	60	61	45	30

Table 1.2. Solubility profile of Ritonavir form I and II at 278 K in ethanol:water mixtures. From [23]

As a consequence of the unexpected implications of polymorphism in Ritonavir formulation, pharmaceutical practice and regulations were tightened, to guarantee a better control of the drug phase diagram, inter-conversion mechanisms and stability of every new drug product [27, 28]. In the following years, more thorough polymorph screenings led to the discovery of an additional metastable polymorph (form IV) and two solvate forms of Ritonavir [29].

### 1.3 METHODS OF CHARACTERISATION

There is a wide range of methods available for the characterisation of polymorphism: in this section we will briefly mention some of the most used laboratory techniques.

#### 1.3.1 Crystallographic techniques

X-Ray Diffraction (XRD) is one of the fundamental methods of characterisation of solid state materials, and it is often described as the “gold standard” of polymorph analysis. The diffraction of the high energy X-rays by the electrons in the structure gives unique information on the periodic arrangement of atoms, as the diffraction of the high energy photons depends on the positions of the atoms in the lattice [9].

The determination of the atomic positions in the unit cell of a crystal is possible by the use of Single Crystal XRD [30]. The requirements for the diffraction of the radiation by the crystalline environment is that the distance between crystal planes  $d$ , the wavelength of the X-ray light source  $\lambda$  and the incidence angle of the light with respect to the planes satisfy Bragg's law,

$$n\lambda = 2d \sin \theta, \quad (1.1)$$

where  $n$  is integer [31]. During the measurement the crystal is rotated, so that every possible plane direction is exposed to the source. From the diffraction pattern of a sufficiently large and pure single crystal it is possible to work out the unit cell parameters and the atomic positions in the crystal.

There is an ever increasing number of crystal structure determina-

tions: for easier access and analysis of data, most are stored and catalogued in databases. The larger collections are the Cambridge Structural Database for small organic and metallorganic molecules, the Protein Data Bank (PDB) for polypeptides and polysaccharides, and the Inorganic Crystal Structure Database (ICSD) for inorganic molecules.

Powder XRD (PXRD) is another important technique. The analysis is performed on a powder composed of a multitude of randomly orientated microcrystals, rather than a single crystal. This has the advantage of a simpler preparation, since a large single crystal might be very difficult to obtain for molecular solids. However, the Bragg refraction condition is not satisfied by a single direction determined by one crystal plane, but whenever one of the microcrystals is in the right orientation: as a result, radiation is diffracted on uniform cones at the angle  $\theta$ , so that there is loss of information.

Despite this problem, PXRD is very useful: different polymorphic forms of a molecule have very different diffraction patterns, that can be used as unique fingerprints. Since the technique is not limited to a single crystal it can be used to precisely measure the composition of a mixed powder (a pattern from a mixture of known components is the sum of the the patterns of the pure phases). Crystal structure determination from powder XRD data, although much more difficult than with single crystal, is not impossible and it is a growing research field [32, 33].

#### *The Cambridge Structural Database*

The Cambridge Structural Database (CSD) [34] is the main repository for the crystal structures of small organic and organometallic molecules, including solvates and salts. The database started in 1965 by collecting published information of the crystal structures of small molecules and it is maintained by the Cambridge Crystallographic Data Centre (CCDC). As of January 2012 it contains almost 580 000 structures [35].

Each entry in the database is identified by a unique 6 letter identifier, eventually accompanied by a two digit number to specify experimental determinations of a compound (different author/journal/polymorph/-experimental condition). The CSD is therefore an invaluable tool to access information on up-to-date determination of polymorphs. Solvates are not stored with the same identifier as the corresponding neat crystal form, as their elemental composition is different: for example, parace-

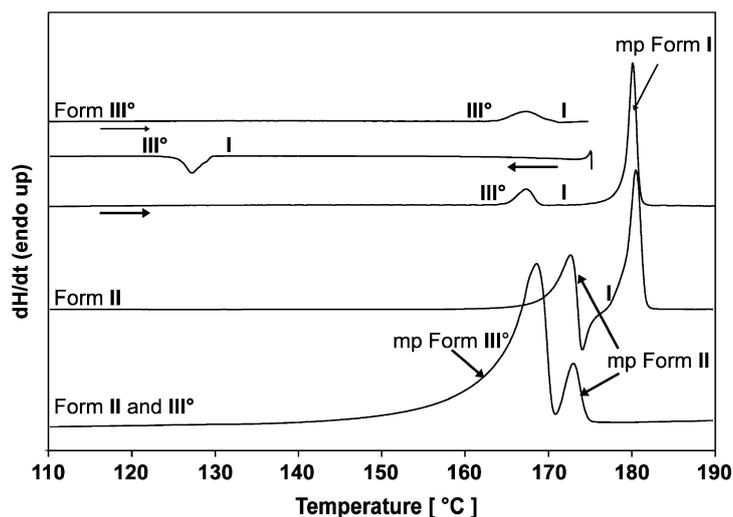


Figure 1.2. DSC thermograms of the possible interconversion of the three polymorphic forms of R-Cinacalcet Hydrochloride. From top to bottom: form III to I transition (heating and cooling); heating until melting (mp); heating of form II with recrystallisation event; heating of a mixture of II and III. From [37]

tamol is identified as HXACAN (with 31 different determinations), while the monohydrate and trihydrate are respectively HUMJEE and XOMWOL.

The database is accompanied by a set of software: among the most useful for our study were CONQUEST, allowing access to the information in the database and MERCURY, a powerful visualisation and analysis tool [36]. With these programs it is possible to generate computed data (such as PXRD pattern to compare with measurement), and retrieve useful statistical information (e.g. number of occurrence of specific intermolecular interactions or length/directionality of hydrogen bonds).

### 1.3.2 Thermoanalytical techniques

Thermal analysis techniques study the relation of properties as a function of the temperature change [38, 39]. These methods can be very useful in the study of phase changes (melting events, crystallisation, glass transitions, polymorph transition, chemical degradation) as variation of properties will be particularly noticeable.

Differential scanning calorimetry (DSC) measures the difference of energy required to vary the temperature of the analyte and a reference, plotted against temperature variation. An example of this method applied to polymorph analysis is seen in Figure 1.2 for R-Cinacalcet Hy-

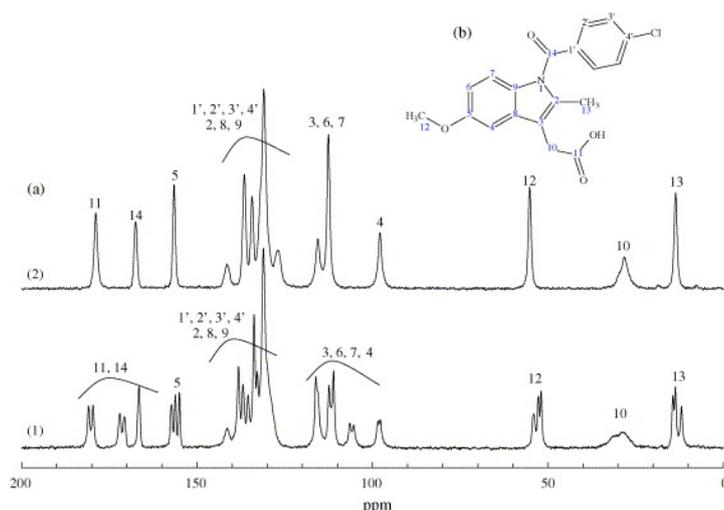


Figure 1.4. (a) Solid-state  $^{13}\text{C}$  ssNMR spectra of crystal forms of indomethacin. 1, bottom graph,  $\alpha$ -form; 2, top graph  $\gamma$ -form. (b) Structure of indomethacin with atoms numbered. From [45]

drochloride (Figure 1.3, only the API shown), a medication used in the treatment of hyperparathyroidism [40]. The DSC analysis shows thermal events corresponding to polymorph interconversions, melting and a recrystallisation (fourth graph, recognisable as an exothermic process).

Thermogravimetric Analysis (TGA) is an analysis technique where the weight change of a sample is measured as a function of temperature or time. It can be useful in the analysis of solvatomorphic systems [41–43], as the analysis can help distinguish between residual solvent and solvent included in the crystal lattice.

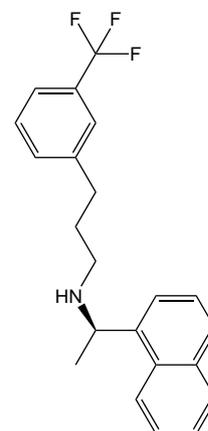


Figure 1.3.

### 1.3.3 Magnetic resonance

Nuclear magnetic resonance (NMR) is an analysis technique based on properties of spin  $1/2$  nuclei (most frequently isotopes  $^1\text{H}$  and  $^{13}\text{C}$ ) in a magnetic field. The molecular environment of the nuclei influences the frequency of resonance of the nuclei (Larmor frequency), and the resulting shifts therefore give information about the sample [44].

While NMR is most frequently used for measurements of samples in solution, solid state NMR (ssNMR) is increasingly used as a methodology for the study of polymorphism in the pharmaceutical industry [46]. ssNMR tends to be very specific in distinguishing polymorphic forms, and it is used as complementary method of analysis. As an example, Figure 1.4 shows the ssNMR spectrum of two forms of indometacin,  $\alpha$  and  $\gamma$ .

The peaks of the two polymorphs exhibit different shifts depending on the chemical environment in their respective unit cells (see the labels on the carbon on the molecular diagram and on the spectrum in Figure 1.4 [45]).

### 1.3.4 Vibrational spectroscopy

Vibrational spectroscopy studies the absorption of electromagnetic radiation by probing the vibrational modes of a molecule or of one of its fragments [47, 48].

Throughout this work we will use a spectroscopic unit, the wavenumber ( $\text{cm}^{-1}$ ), defined as the inverse of the wavelength of light; the advantage is that the energy is directly proportional to the wavenumber.

The frequencies for mid-infrared spectroscopy (mid-IR) range from about  $4000 \text{ cm}^{-1}$  to  $400 \text{ cm}^{-1}$ . The highest frequency region comprises of frequencies that are easily assignable to specific intramolecular bonds: for example stretching of N–H bonds (around  $3400 \text{ cm}^{-1}$ ) or C=O ( $1600 \text{ cm}^{-1}$  to  $1800 \text{ cm}^{-1}$  depending on the group). A lower region of the spectrum is called the “fingerprint region”, as the absorptions are peculiar of each molecule: the corresponding modes of vibration are typically too complex for a 1:1 assignment to normal modes of vibration.

The main strength of mid-IR is undoubtedly in the recognition of molecular features, as the subtle changes in the spectrum of different polymorphic forms of a drug are usually very small, and can be also due to differences in the preparation and in the samples. A number of studies have confirmed that advanced methods characterised by increased sensitivity, such as Fourier Transform IR and Attenuated Total Reflection IR have the necessary accuracy and reproducibility to probe polymorphism [49–51].

Near infrared spectroscopy (NIR, frequency range  $4000 \text{ cm}^{-1}$  to  $14\,000 \text{ cm}^{-1}$ ) is associated with multiple absorption of photons from the material (overtones or combination bands) [47]. It has been used for detection and characterisation of polymorphs and solvates [52, 53], but it is preferentially being used for the determination of sample composition together with multivariate methods of analysis [54].

The lowest energy part of the infrared spectrum (far IR, or terahertz) is characterised by a spectral range from  $2 \text{ cm}^{-1}$  to  $400 \text{ cm}^{-1}$ . In this re-

gion the interactions probed belong to the so-called “lattice modes” of the crystal, where the radiation interacts with the intermolecular bonds rather than intramolecular vibrations: the absorption is therefore dependent on the packing, and spectra of different polymorphs display easily distinguishable features.

While it is not possible to determine crystallographic data or the elemental composition without comparison with a database, terahertz spectroscopy has a number of advantages that will be listed in the next section.

#### 1.4 PECULIARITIES OF TERAHERTZ SPECTROSCOPY

Instrumentation and methods of operation of terahertz pulsed spectroscopy (which is more commonly referred to as terahertz time-domain spectroscopy – THz-TDS), will be described in more detailed in chapter 2. Here we will just mention some of the important points of this technique [9, 55]:

- Speed of analysis: absorption spectra can be obtained in less than a second, making it a technology with potential to be used as a fast screening tool. It is also possible to measure while continuously varying the conditions [56] and monitoring interconversion between polymorphic crystal forms *in situ* [57];
- No need of lengthy preparation of the sample, and non destructive technology;
- Non ionising radiation: the photon energy in the terahertz region is low, and there is almost no risk of photochemical reaction or degradation of the sample;
- Terahertz spectra can be sensitive enough in cases when NIR alone would not be sufficient [58].

Despite the usefulness of terahertz analysis of polymorphs, as of now it is not possible to assign individual absorption features to molecular vibrations without a computation (differently to what happens with mid-IR), due to the complexity of the vibrational modes in the crystal at these frequencies.

Several techniques have been used to approach the problem of interpreting the spectra of molecular crystalline systems (as one finds in pharmaceutical chemistry): isolated molecule calculations of the normal modes of vibration [59] (although conceptually incorrect, see Table 3.2, page 30), rigid molecule approximations of lattice dynamics [60], molecular mechanics/molecular dynamics with empirical force fields [61], and solid-state density functional theory [62].

## 1.5 THESIS OUTLINE

The main aim of this thesis is the development and application of computational methods to the study of small organic molecules of pharmaceutical interest. In chapter 2 we review the theory of terahertz pulsed spectroscopy, and we describe in detail our experimental setup.

Chapter 3 describes the theory of molecular vibrations, in particular the calculation of normal modes of vibration in periodic systems and its implementation in force field based methods and density functional theory. Chapter 4 deals in particular with the use of our two programs of choice, *DMACRYS* and *CASTEP*.

Chapter 5 deals with computational methods we employed to deal with the effect of local polarisation in the crystal using *DMACRYS*, attempting to improve the accuracy of the description of intermolecular interactions.

The molecular systems we select to study are listed in chapter 6, with mention of their crystal structure, presence of polymorphic and solvate forms.

The last part of the thesis is devoted to the comparison of the experimental spectra with their computational counterparts and in the analysis of the results. In particular, in chapter 7 and 10 we report the spectra computed respectively with and without polarisation correction. In chapter 8 we analyse the effect of the rigid body approximation used in *DMACRYS*, and the comparison of the vibration eigenvectors with *CASTEP*.

Chapter 9 deals with the effect of water inclusion in solvatomorphic system, and its effect on the calculated eigenvectors. Finally, in chapter 11 we consider the terahertz spectra of structurally similar molecular systems.

## TERAHERTZ PULSED SPECTROSCOPY

---

**T**ERAHERTZ SPECTROSCOPY is the branch of spectroscopy dealing with the interaction of samples with the far-infrared region of the electromagnetic spectrum, usually defined as the range of 0.1 THz to 4 THz (equivalent to 3 to  $133\text{ cm}^{-1}$  in spectroscopic units) [63, 64]. This part of the spectrum bridges the infrared region and the microwave region: its importance lies in the fact that it can probe low energy vibrational and rotational motions within solids without inducing structural transformation or ionisation, due to the low energy of the radiation used.

Historically there has been considerable difficulty in using terahertz radiation compared to infrared and microwave radiation. Due to the intrinsic limits of the commonly employed techniques of generation of radiation there was a lack of sources in the region, which was therefore also referred to as “the terahertz gap” [65]. This occurs because the frequencies in this region are too low to be generated with standard optical techniques (involving electronic transitions in atoms and molecules), while being at the same time too high to use electronic sources. Lasers exist in the terahertz range [66] (for example molecular gas sources such as the  $\text{CO}_2$  laser [67]), but they are not tunable: semiconductor-based lasers are partially tunable, but non continuously and they are not broad enough to be useful for spectroscopic measurements [68]. A recent overview of electronic terahertz sources can be found in the review article by Inguscio and Moruzzi [69].

However, not only was the generation of terahertz radiation a challenge, but also its detection. Standard optical detection techniques suffer from a range of limitations, because the band gap in photodiodes cannot be excited by terahertz radiation. On the lower energy side of the spectrum, the upper frequency limit of microwave sources of radiation such as cavity magnetrons and klystrons are limited to around  $10\text{ cm}^{-1}$  to  $20\text{ cm}^{-1}$ , at the low end of our spectroscopic range of interest [70].

Using thermal black body source was the most common method to perform spectroscopic experiments in the far infrared. However, the detection sensitivity was rather poor due to the fact that cryogenically cooled bolometers had to be used for detection. Using such a setup the signal-to-noise that can be achieved is limited by the weak signal that has to be discerned from the strong thermal background radiation.

## 2.1 FEMTOSECOND LASER SOURCES

The availability of easily generated broadband terahertz radiation changed during the 1990s [71–74]. The development of femtosecond laser sources, together with the advances of semiconductor physics, provided a much easier way to generate and detect a broadband signal.

Nowadays the two most widespread techniques to generate and detect terahertz radiation make use of photoconductive antennas and non-linear optical crystals. They will be described in detail in the next sections.

### 2.1.1 Photoconductive antennas

Photoconduction is defined as the change in conductivity in a material upon irradiation. The radiation process is the result of the acceleration of photogenerated charge carriers inside a device (antenna) excited by a subpicosecond laser pulse.

An antenna comprises of a semiconductor substrate with high carrier mobility (usually GaAs), surmounted by two metallic gates separated by a gap ranging from a few microns to a few hundreds of microns (Figure 2.1). A DC voltage is applied between the two metallic gates, and in the absence of a field the substrate acts as an insulator, i.e. there is no current.

When photons with higher energy than the band gap of the semiconductor hit the antenna, electron-hole pairs are generated, which are subsequently accelerated across the device towards the gates. The carrier density  $N(t)$  and their velocity  $v(t)$  depends on the intensity and on the time shape of the incoming laser pulse (its shape is important for such a short pulse); the associated current  $J(t)$  creates a polarisation field  $P(t)$  inside the semiconductor, which itself affects the motion of the charge

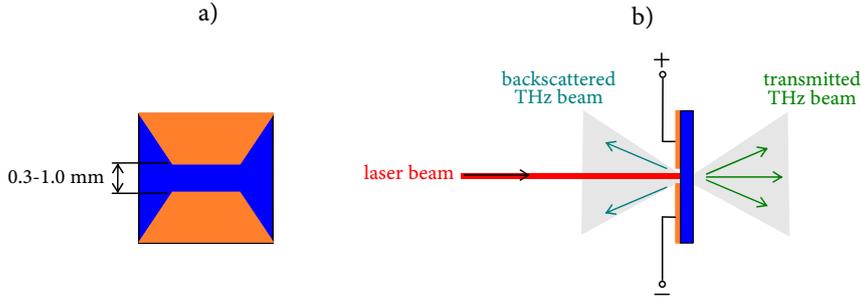


Figure 2.1. Schematic view of a photoconductive antenna used for generation of the THz pulse. a) Frontal view. b) Sketched pulse generation. Note that there are two waves, one transmitted, the other backscattered

carriers.

The generation of the terahertz pulse, defined from its electric and magnetic components ( $\mathcal{E}_{\text{THz}}$  and  $\mathcal{H}_{\text{THz}}$  respectively), follows from the transient current, according to Maxwell's equations:

$$\begin{aligned}\nabla \times \mathcal{E}_{\text{THz}}(t) &= -\mu \frac{\partial \mathcal{H}_{\text{THz}}(t)}{\partial t} \\ \nabla \times \mathcal{H}_{\text{THz}}(t) &= \mathbf{J}(t) + \frac{\partial}{\partial t}(\epsilon(t)\mathcal{E}_{\text{THz}}(t))\end{aligned}\quad (2.1)$$

where  $\epsilon$  and  $\mu$  are the dielectric constants of the semiconductor and  $\mathbf{J}$  is the induced photocurrent, which is directly related to the number and the velocity of the charge carriers. The polarisation induced by the accumulation of the charge carriers affects itself the electromagnetic field. The acceleration of the charges produces an electric field perpendicular to the device surface, its frequency depending on the profile of the incident laser pulse, as well as on the characteristic parameters of the carriers [75], that need to possess a very fast response (in the femtoseconds regime).

Analytic and numerical solutions of Equations 2.1 exist in the dipole approximation [76], where the generated electric field is directly proportional to the change of  $\mathbf{J}$  with time, and is therefore strictly connected to the characteristics of the incident laser pulse. The spectral bandwidth of the generated terahertz field can be approximated from the uncertainty relationship on the field: since  $\Delta t \Delta \omega \geq 1/2$ , a 100 fs incoming signal should create a bandwidth of at least 3 THz, with a shorter pulse generating a wider wavelength range.

In summary, the laser pulse acts as a switch for the generation of the terahertz field: in between two pulses the high resistance inside the semiconductor does not allow for a current to flow.

The relaxation processes of the polarisation field in the semiconductor is not on a femtosecond scale: the recovery of the initial configuration, with the charge carriers recombining in the semiconductor, takes hundreds of picoseconds or longer, which leads to the terahertz pulse being effectively much longer than the laser pulse.

Due to this difference in time scale a similar structure can be used for the detection of the terahertz radiation. Using an optical delay the focused terahertz beam and the laser pulse can be configured to reach a detection device at the same time; however, for the detection process there is no need for a DC bias. When the laser pulse hits the antenna the semiconductor surface becomes conductive: the incident terahertz beam (which is an electric field itself) generates a drift of the carriers, proportional to the amplitude of the electric field, and the associated current can be measured. The antenna is only conductive when the semiconductor charge carriers are present, and the recombination within the semiconductor is much faster (due to the absence of the depolarisation effects). It is therefore possible to perform a mapping of the much longer terahertz pulse by controlling the time delay of the laser, allowing a reconstruction of the terahertz profile [64].

## 2.2 ELECTRO-OPTICAL RECTIFICATION

An alternative, widely used method to generate and measure terahertz pulses uses the electro-optical effect. This phenomenon, observed in some crystals, is based on the non-linear response to an electric field of the polarisation in a medium [77].

The scalar polarisation response  $P(t)$  to an electric field  $\mathcal{E}(t)$  inside a material can be expressed as a power expansion:

$$P(t) = \chi_1 \mathcal{E}(t) + \chi_2 \mathcal{E}^2(t) + \chi_3 \mathcal{E}^3(t) + \dots \quad (2.2)$$

with the nonlinear terms  $\chi_2, \chi_3, \dots$  usually much smaller than  $\chi_1$ .

If the second term of the expansion  $\chi_2$  is not negligible it is possible to observe second-order effects in this material: for example, by trigonometric relationships, for a field of the form  $\mathcal{E}(t) = \mathcal{E}_0 \cos \omega t$  there is a

DC polarisation and second harmonic generation,  $P_2$ :

$$P_2(t) = \chi_2 \mathcal{E}^2(t) = \frac{1}{2} \chi_2 \mathcal{E}_0^2 (1 + \cos 2\omega t) \quad (2.3)$$

Similarly, when two electric fields are present,  $\mathcal{E}_1(t) = \mathcal{E}_0 \cos \omega_1 t$  and  $\mathcal{E}_2(t) = \mathcal{E}_0 \cos \omega_2 t$ :

$$P_2(t) = \chi_2 \mathcal{E}_1 \mathcal{E}_2 = \frac{1}{2} \chi_2 \mathcal{E}_0^2 (\cos(\omega_1 - \omega_2)t + \cos(\omega_1 + \omega_2)t) \quad (2.4)$$

The generation of terahertz radiation is due to the profile of the femto-second laser impulse: the field  $E(t)$  of a Gaussian pulse governed by a temporal width  $\Gamma$  and centred at  $\omega_0$  is

$$E(t) = E_0 e^{i\omega_0 t - \Gamma t^2} \quad (2.5)$$

and its frequency distribution is its Fourier transform:

$$E(\omega) = \exp\left(-\frac{(\omega - \omega_0)^2}{4\Gamma}\right), \quad (2.6)$$

meaning that a pulse in the femtosecond range has significant frequency contributions that span over a few terahertz: the difference mixing of the frequencies generates a pulse in the terahertz range.

The advantage of this method in the generation of terahertz radiation is that it depends only on the length of the pulse, and is not limited by upper limits in the mobility of carriers as in the case of antennas: very short terahertz pulses (displaying bandwidths up to 100 THz) have been generated in this way.

The detection of the signal can be performed with a non-linear medium, using the Pockels effect [78], which is closely related to the electro-optical generation: it is possible to induce birefringence in an optically active medium by a constant or varying electric field, and the birefringence is proportional to the electric field.

The detection method is based on a non-linear optical crystal (often ZnTe), followed by a quarter-wave plate, a Wollaston prism and a pair of balanced photodiodes. The working principle is outlined in Figure 2.2.

A linearly polarised probe beam (our femtosecond pulse) propagates through the crystal unchanged in the absence of a field. After transmitting the non linear crystal its polarisation is changed from linear to

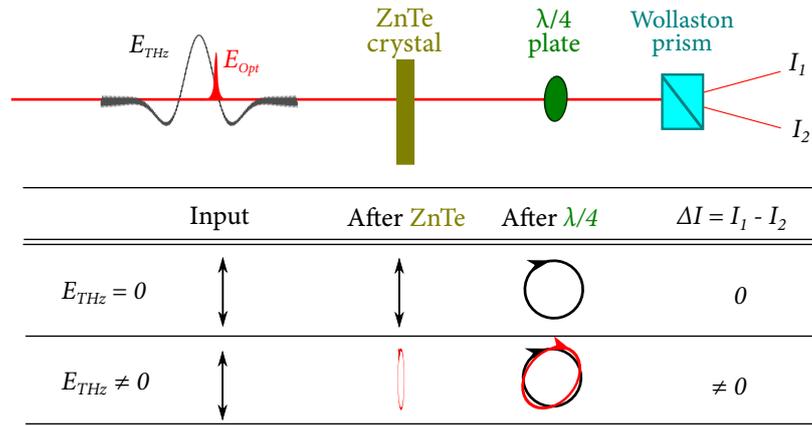


Figure 2.2. Detection based on electro-optical rectification. Top graph: terahertz wave ( $E_{THz}$ ), the optical pulse ( $E_{Opt}$ , in red), and the optical devices. Top table: variation steps of the terahertz wave polarisation. Bottom table: variation steps of the terahertz wave polarisation influenced by the optical pulse in the ZnTe crystal

circular using the quarter-wave plate. The Wollaston prism is used to split the signal into two orthogonal components, measured by two photodiodes. In the case of circular polarisation there is no difference in intensity between the two components.

If the pulse reaches the crystal at the same time as the terahertz pulse, the induced birefringence changes the polarisation from linear to elliptical: the quarter-wave plate keeps the elliptical polarisation and the two components separated by the Wollaston prism are now different: the measured intensity  $\Delta I$  is directly proportional to the terahertz field.

As in the case of detecting antennas, the shortness of the femtosecond pulse compared to the terahertz pulse allows sampling of the terahertz signal by time delayed measurement.

## 2.3 EXPERIMENTAL SETUP

A photograph of the spectroscopic equipment we use for all of our measurements is reported in Figure 2.3. The terahertz pulses are generated by a photoconductive GaAs bowtie antenna, with a 0.5 eV band gap. The femtosecond source is a Ti:Sapphire laser (Femtolasers), producing 12 fs laser pulses centred at 800 nm wavelength. Detection is performed with electro-optical sampling using a ZnTe crystal.

In Figure 2.4a we highlight the path of the terahertz beam, split in

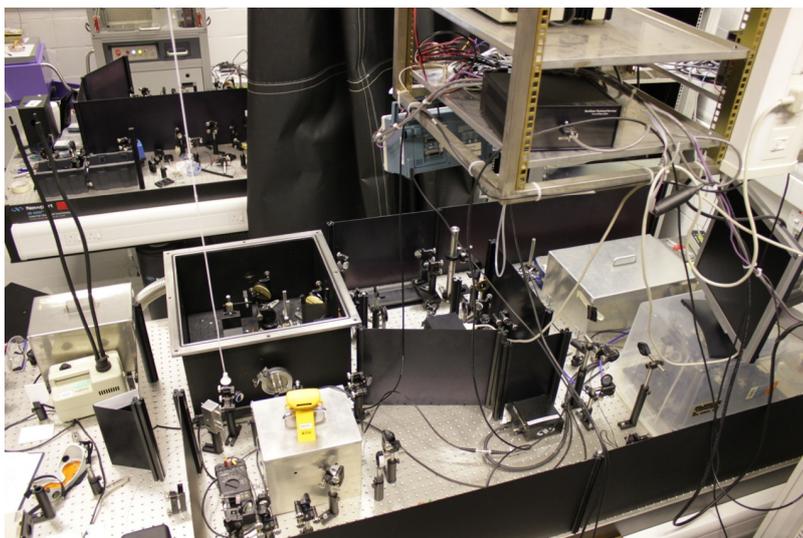


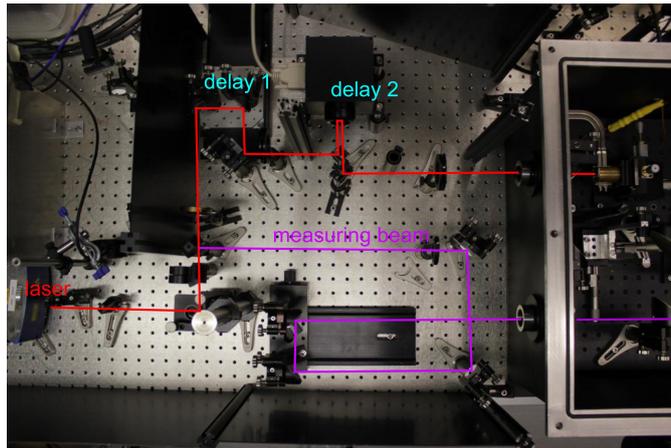
Figure 2.3. Experimental setup used for all the experiments in this work

two parts: one used for the generation of the terahertz pulse, the other (drawn in purple) for the detection mechanism. The isolated box (Figure 2.4b) contains the antenna, the sample holder where the terahertz is focussed by parabolic lenses. The detecting mechanism, with the quarter-wave plate, the Wollaston prism and the diodes is shown in Figure 2.4c.

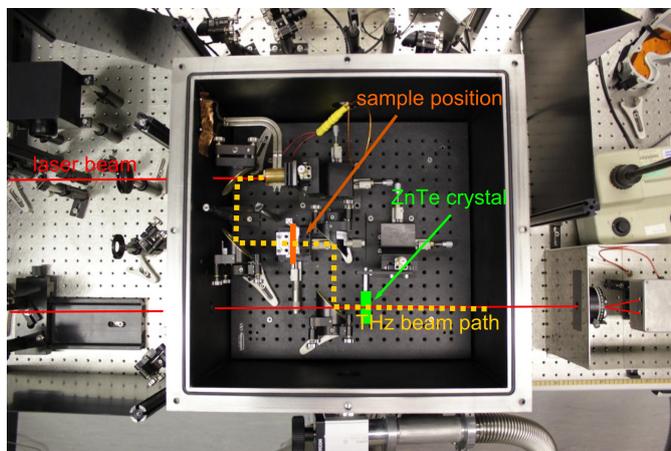
During the acquisition of the terahertz spectra the sample compartment is either purged with dry nitrogen gas or evacuated to remove any atmospheric water vapour, as water molecules exhibit very intense rotational absorption features in the measurement frequency window (see Figure 2.5).

### 2.3.1 Preparation of the samples

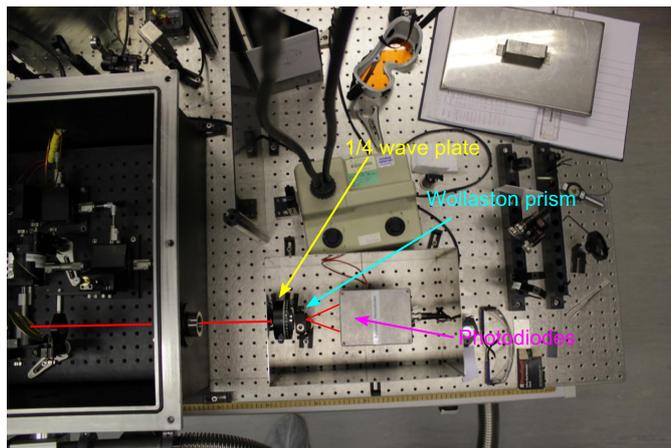
In order to measure the transmission spectrum of a sample material a pellet of sufficient concentration and mechanical strength has to be prepared. For a standard size sample pellet (13 mm in diameter) between 10 and 20 mg of sample material are thoroughly mixed with 360 mg of polyethylene powder (PE, which is almost transparent at terahertz frequencies). Mixing is performed using an agate mortar and the method of geometric mixing. The mixture is then compressed into a pellet using a hydraulic press at 2 ton load for about 2 minutes. The resulting pellets are between 2 and 3 mm thickness. For each sample three pellets were prepared to ensure reproducibility of the measurement.



(a) Laser path and optical delays prior to the generation of the terahertz beam



(b) Inside the isolated box: generation of the terahertz beam (path highlighted), with the sample handler and the non-linear optical crystal



(c) Detection mechanism, with the quarter-wave plate, the Wollaston prism and the balanced photodiodes

Figure 2.4. Photographs of the spectrometer we used for our measurements. Courtesy of Axel Zeitler

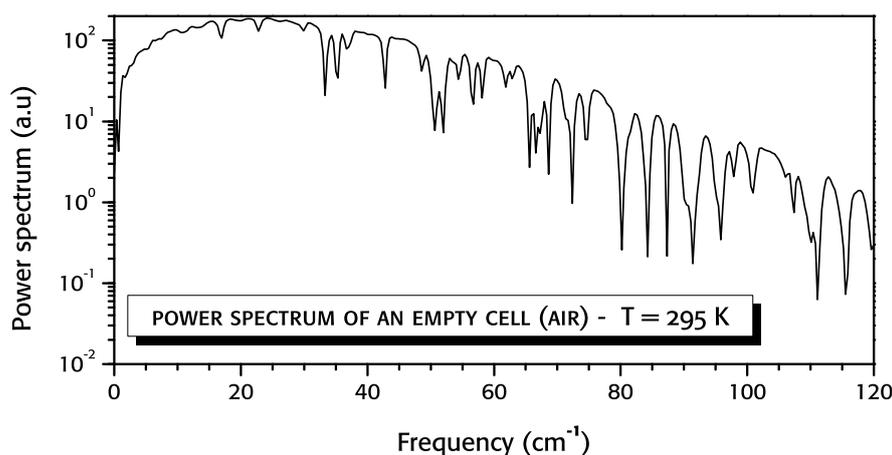


Figure 2.5. Power spectrum of an unpurged empty cell at 295 K, showing the absorption features of water vapour. Absorption is expressed in arbitrary units (a.u.) as throughout the rest of this thesis

### 2.3.2 Measurement of the absorption spectra

The output of the photobalanced diodes is the difference in intensity between the two components of the polarisation of the laser pulse, which is proportional to the electric field amplitude of the terahertz pulse. The pulse is sampled by the use of the optical delays to obtain its profile after the absorption by the pellet (Figure 2.6). The pulse shape is then Fourier-transformed to reconstruct its frequency components, to recover the power spectrum (Figure 2.7, black trace).

The power spectrum of a pellet containing the same amount of polyethylene powder as the measured sample (360 mg powder, compressed at 2 ton load for about 2 minutes) is collected in the same way and at the same temperature, and it is used to remove the baseline from the power spectrum of the drug, to finally obtain the absorption spectrum (Figure 2.8).

In order to improve the signal-to-noise we co-averaged over 200 time-domain waveforms prior to Fourier transformation. The dynamic range of the instrument is not constant with frequency (see Figure 2.7, logarithmic scale): there is much more power available in the spectral region  $5\text{ cm}^{-1}$  to  $80\text{ cm}^{-1}$ . The analysis is therefore much more reliable in this region, while over  $110\text{ cm}^{-1}$  is almost completely noise.

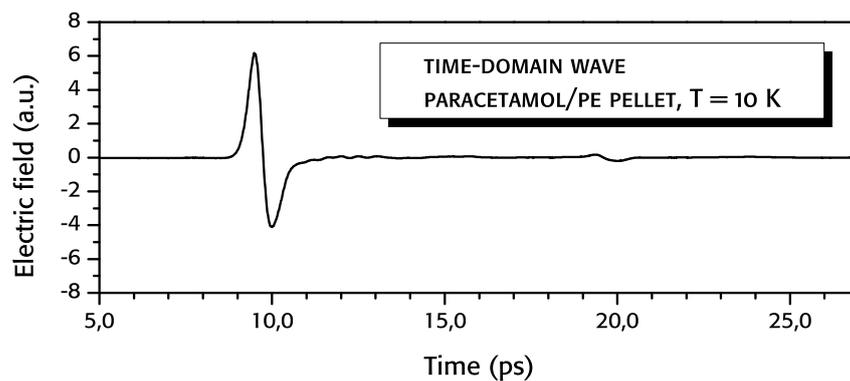


Figure 2.6. Time-domain shape of the terahertz pulse after absorption from a pellet with 360 mg PE and 15 mg paracetamol at 10 K

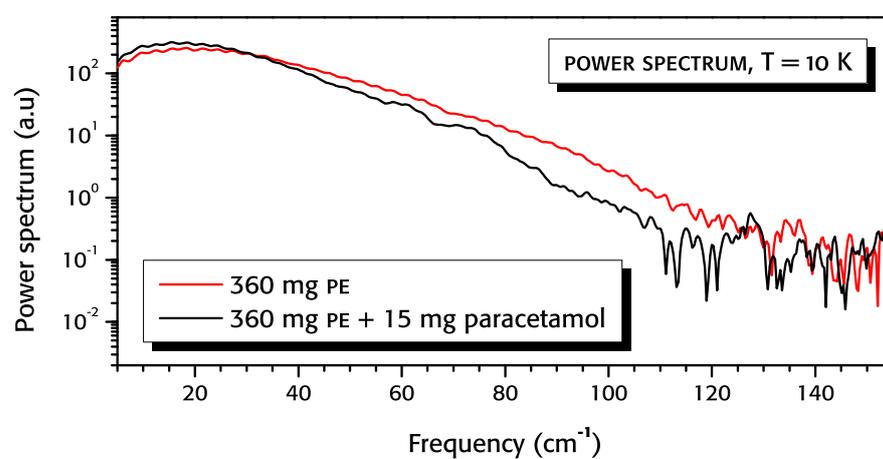


Figure 2.7. Power spectrum of the terahertz time-domain waveform after absorption from a paracetamol/PE pellet (containing 15 mg/360 mg - black trace) and a tablet containing 360 mg PE only (red trace)

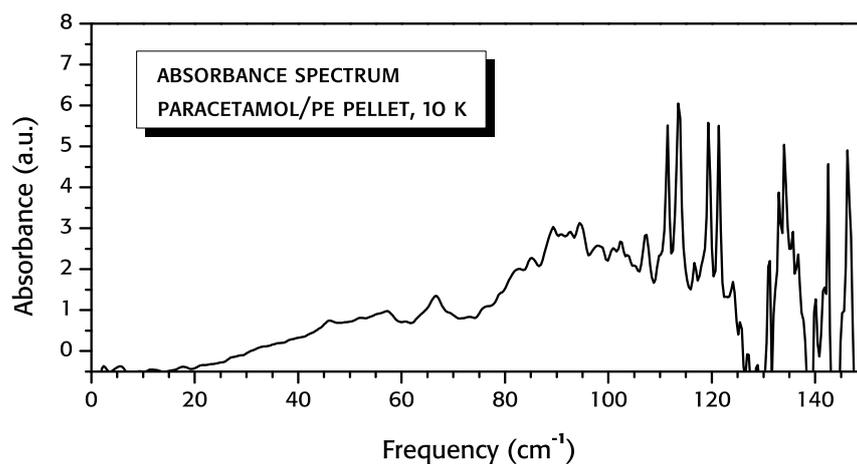


Figure 2.8. Absorption spectrum of the paracetamol/PE pellet after removal of the PE baseline

## 2.4 TEMPERATURE EFFECTS ON THE SPECTRA

The characteristic absorption peaks of room temperature terahertz spectra of different crystal forms usually exhibit sufficiently characteristic absorption peaks to easily distinguish one from the other. However, since our goal is the comparison of the absorption spectra with computed data, we need a clear spectrum as possible, and this is better achieved with measurement at low temperature, for thermodynamic reasons.

Within a quantum mechanical description, the absorption of a photon of frequency  $\nu$  by a vibrational mode induces a transition from an energy level  $n$  to the higher  $n + 1$  level. The intensity of absorption  $\mathcal{I}$  is proportional to the relative population of the vibrational states, according to the Boltzmann distribution:

$$\mathcal{I} \propto \exp\left(-\frac{E_n}{k_B T}\right) - \exp\left(-\frac{E_{n+1}}{k_B T}\right), \quad (2.7)$$

where  $E_n$  is the energy of the  $n$ th energy level,  $T$  the temperature and  $k_B$  the Boltzmann constant. For a harmonic oscillator (first approximation of a vibrational mode) the energy is  $E_n = h\nu(n + 1/2)$ :

$$\mathcal{I} \propto \left[1 - \exp\left(-\frac{h\nu}{k_B T}\right)\right] \exp\left(-\frac{h\nu(n + 1/2)}{k_B T}\right) \quad (2.8)$$

The population of energy levels higher than the ground state is not negligible at room temperature, since the factor  $h\nu/k_B T$  is close to unity

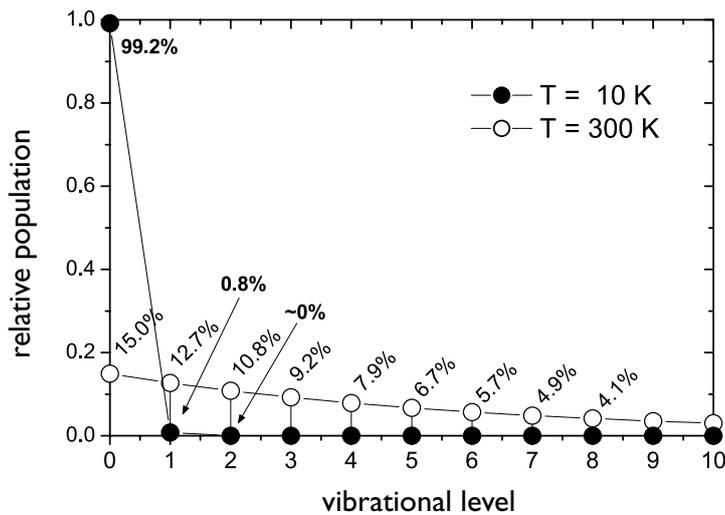


Figure 2.9. Population of the first 10 levels of a harmonic oscillator with frequency  $\nu = 1$  THz, at different temperatures. From [79]

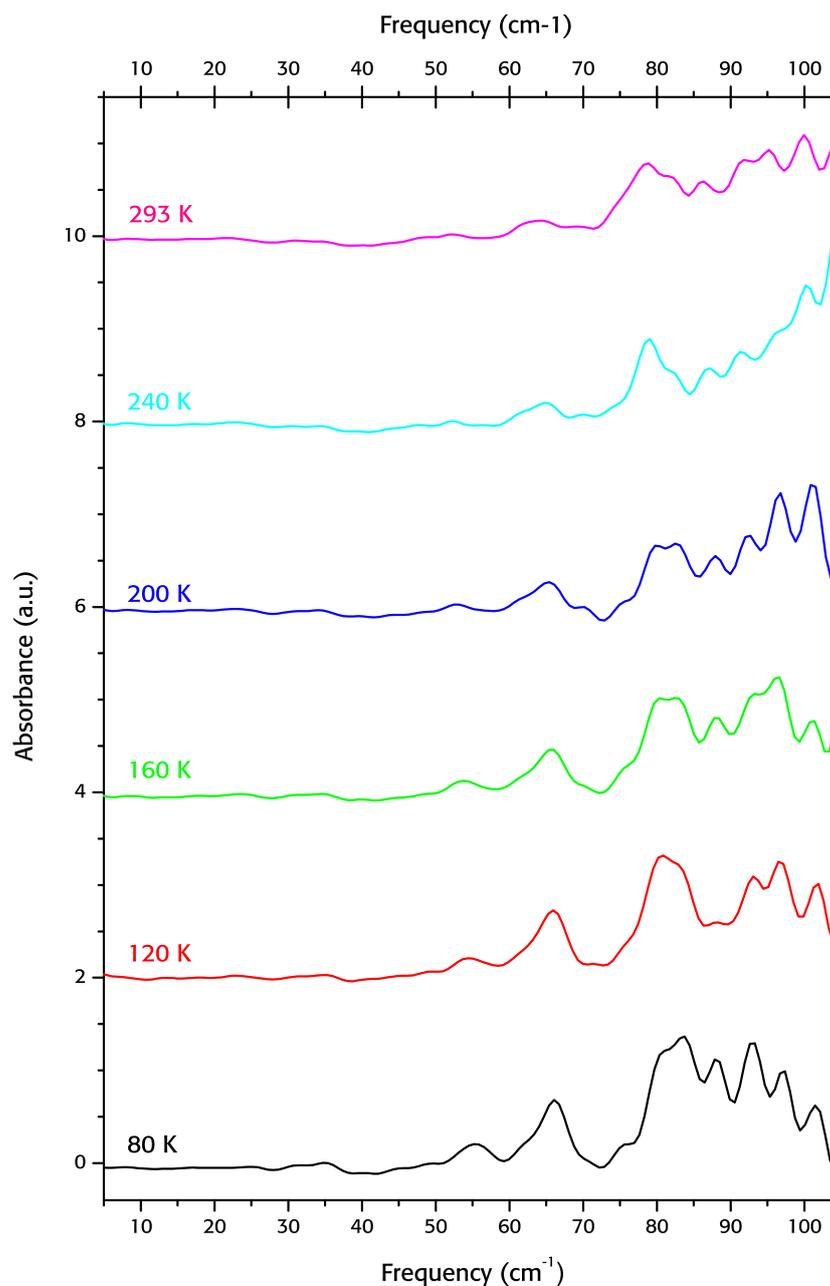


Figure 2.10. A series of measurements of a tablet containing 360 mg PE and 15 mg N-(4-Bromo-phenyl) acetamide. Each spectrum is offset by 2 a.u. for clarity

within the measurement range (it is 1 for  $\nu \approx 6.15$  THz at 298 K). As a consequence, absorptions are much stronger at low temperature, where the population of the fundamental state is enhanced (see for example Figure 2.9).

The presence of “hot states” ( $n > 0$ ) has another effect: due to anharmonicity the energy levels are not equally spaced, and absorption

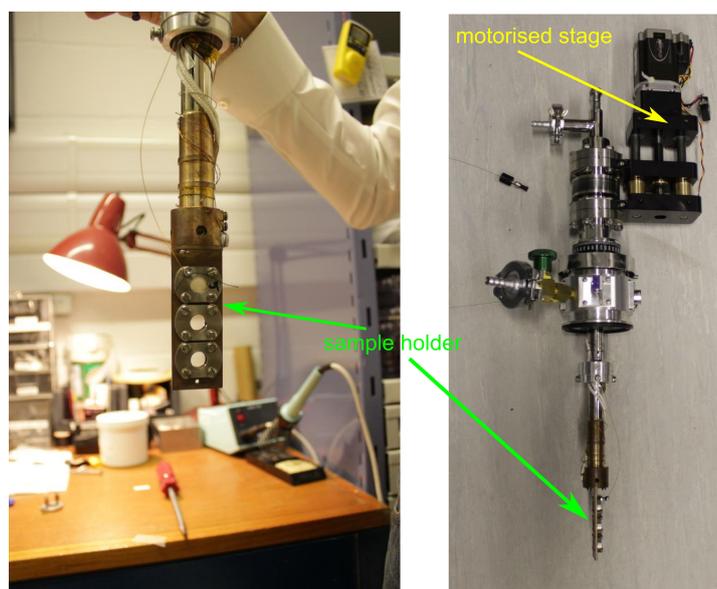


Figure 2.11. Two views of the cryostat we used for the low temperature measurements

lines can be red-shifted. The same effect is also caused by the thermal expansion of the crystallographic unit cell of the measured sample. Lattice dynamics methods of simulation (see next chapter) assume that the temperature is 0 K: therefore the comparison of experimental with computed spectra is more meaningful for low temperature measurements. An example of how temperature affects the measured spectra is reported in Figure 2.10.

#### 2.4.1 Low temperature setup

Low temperature experiments were performed using a continuous flow cryostat (modified ST-100-FTIR, Janis, Wilmington, MA USA, see Figure 2.11). The cryogen was either liquid nitrogen or liquid helium. The sample chamber was evacuated to a pressure of  $< 50$  mbar, when using liquid nitrogen and  $< 1.0 \times 10^{-6}$  mbar when operating with liquid helium. Temperature control was achieved by adjusting the flow of cryogen using a needle valve in the transfer tube and by a heater in the cold finger which was adjusted using a temperature controller (331S, Lake Shore, Westerville, OH USA). The temperature stability was better than 0.1 K and measurements were acquired after an equilibration time of 1 minute at each temperature. A motorised line stage (Figure 2.11, right) was fitted to the cold finger to switch between the sample and reference

pellet at each temperature, thus avoiding any potential instrument drift between the two measurements. each temperature.

For paracetamol form I and nitrofurantoin form  $\beta$  (see chapter 6) we measured spectra ranging from 10 K to 75 K using liquid helium as cryogen. The use of nitrogen as a cooling gas, however, is both cheaper and simpler from the experimental point of view; we therefore decided to use nitrogen for all the other measurements, since we judged the quality of the spectra to be good enough at 80 K. For all the other systems we also measured the spectra from 80 K to room temperature, with an interval of about 40 K. All of the variable temperature measurements are reported in appendix A.

## VIBRATIONS IN A PERIODIC SYSTEM

---

**I**N CLASSICAL INFRARED spectroscopy it is often possible to directly connect an observed absorption to the distortion (e.g. stretch or bend) of a bond in a molecule, usually without the need of calculation. This is because the vibrational normal modes are often fairly localised and because their associated vibrational frequencies are only slightly shifted by their chemical environment. Therefore, certain vibrational frequency ranges are characteristic of known distortions of covalent bonds [48].

However, the vibrational normal modes become increasingly complex at lower frequencies; in the terahertz region, it is not generally possible to associate an absorption feature to a simple feature such as the stretching of a bond. Instead, absorption features generally result from collective motions of all the atoms in the material. To correctly assign the nature of the molecular motions associated with a particular absorption, it is therefore necessary to rely on simulations.

### 3.1 SINGLE MOLECULE CALCULATIONS

If a simulation is theoretically well-founded and can reproduce the positions and intensities of features in an observed spectrum, it can also be trusted to provide a faithful description of the vibrational motions that give rise to these features.

A tempting approach to calculating the absorption spectrum of a molecular crystal is to calculate the normal modes of vibration of an isolated molecule, as single-molecule calculations are generally much cheaper (in computing time and required computing resources) than calculations that include the entire periodic crystal structure. However, the vibrational features observed in the gas phase and in a solid can be radically different: in the gas phase, the interactions between molecules can

usually be neglected, while in solids this is not true, especially when considering low frequency vibrations. Furthermore, pure rotational modes of molecules are allowed in the gas phase [80, 81] (as can be seen in the absorption spectrum of water vapour in Figure 2.5), while molecular rotations are hindered in the solid phase by molecular close-packing and by the interactions between molecules.

Consider the case of naphthalene (Figure 3.1), whose lowest frequency infrared active vibrational frequencies are listed in Table 3.1.

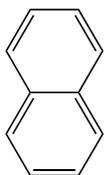


Figure 3.1.

The terahertz spectrum of the gas phase and the calculation of the isolated molecular vibrational frequencies [82] agree in finding no absorption below  $175\text{ cm}^{-1}$ , while the experimental spectrum of crystalline naphthalene [83] shows several features below  $100\text{ cm}^{-1}$ . These additional features only appear for the crystalline sample, so they must be related to the intermolecular interactions within the crystal. Indeed, simulations that include the periodic structure of the crystal structure [84, 85] reproduce the observed frequencies and indicate that the corresponding vibrations relate to whole-molecule motions about the equilibrium crystallographic positions.

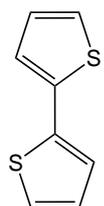


Figure 3.2.

Another example is the terahertz spectrum of 2,2'-bithiophene (Figure 3.2), shown in Figure 3.3. Again, there are clear differences between the vibrational modes of the isolated molecule and that of the crystal: in this case, the isolated molecule does have low energy vibrational modes near and below  $100\text{ cm}^{-1}$ . However, this region of the spectrum becomes much more detailed in the crystal and the number and position of these features can only be reproduced in a calculation that includes the entire crystal structure [86, 87].

A trickier example is represented by the calculations of the drug molecule 3,4-methylenedioxymethamphetamine (MDMA or ecstasy, Figure

Gas phase		Crystal		
Experimental	Calculated	Experimental	Calculated	
[82]	[82]	[82]	[84]	[85]
166	179	53	45	54
359	376	66	62	74
473	492	98	91	97

Table 3.1. Comparison of the three lowest energy infrared/terahertz active absorption frequencies ( $\text{cm}^{-1}$ ) for naphthalene, in crystal and gas phase

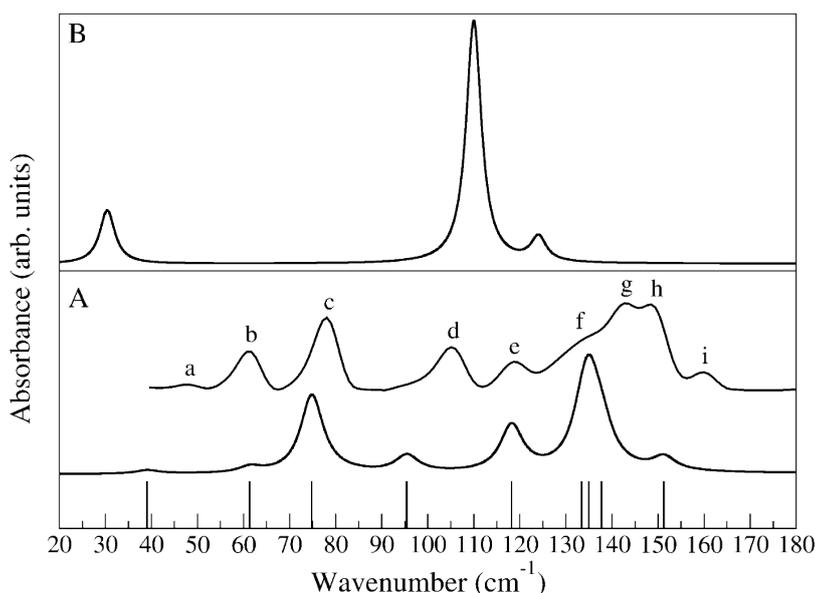


Figure 3.3. Terahertz spectrum of a 2,2'-bithiophene (Subfigure A, upper spectrum) compared with a solid state calculation (A, lower spectrum) and an isolated molecule calculation (B). The strongest experimental features are indicated with letters. From [86]

3.4). In this case, calculations on the isolated molecule [88] seem to provide excellent agreement with the experimental spectrum (Table 3.2), so that features of the terahertz spectrum were assigned to intramolecular vibrations. However, a series of calculations [89, 90] based on the dynamics of the known crystal structure, showed that the *intramolecular* vibrations are shifted out of the terahertz region by coupling to the crystal environment (see section 3.2.3). Moreover, the calculations produce a series of *intermolecular* vibrational modes at the frequencies seen in the experimental spectrum. Here, it seems that the agreement of the frequencies found in the isolated molecule calculation with the experimental spectrum was a result of fortuitous coincidence.

In conclusion, to correctly calculate the terahertz spectrum and the corresponding vibrational modes (phonons) of a crystalline material it is necessary to employ methods that consider the molecular arrangement in the periodic structure and the associated intermolecular interactions present in a crystal.

In the next sections we briefly describe the classical theory of lattice vibrations in a periodic systems (see for example [91]): assuming small displacements allows us to simplify the theoretical treatment to only consider harmonic oscillations.

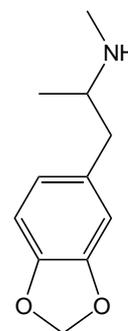


Figure 3.4.

Experimental	Isolated molecule		Crystal	
	Freq.	Assignment	Freq.	Assignment
37.0	36.6	Internal C–C bond torsion	38.0	Rigid molecule rotations
59.3	58.8	Internal C–C bond bending	61.0	Rigid molecule translations
86.6	85.8	Internal CH <sub>2</sub> group rocking	94.4	Internal C–C bond bending + rigid translations

Table 3.2. Experimental [88] and calculated [88, 89] frequencies of the terahertz absorption of MDMA ( $\text{cm}^{-1}$ ), from isolated molecule and from lattice dynamics of the crystal, and relative assignment to vibrational features

### 3.2 PHONONS IN A CRYSTAL

The vibrational properties of a periodic system in equilibrium are evaluated by taking into account the motion of the atoms around their equilibrium positions within the crystal. The potential energy of the crystal,  $\varphi$ , is a function of the atomic coordinates; small displacements of an atom are counteracted by a restoring force, pulling the system back to its equilibrium position.

The potential energy  $\varphi$  can be expressed in a Taylor series up to second order in the displacement  $u_\alpha^{nl}$ , the subscript  $\alpha$  representing one of the Cartesian coordinates ( $x, y, z$ ) of atom  $n$  within the unit cell  $l$ :

$$\varphi = \varphi_0 + \frac{1}{2} \sum_{\alpha n l} \sum_{\alpha' n' l'} \left. \frac{\partial^2 \varphi}{\partial u_\alpha^{nl} \partial u_{\alpha'}^{n' l'}} \right|_0 u_\alpha^{nl} u_{\alpha'}^{n' l'}. \quad (3.1)$$

In the equilibrium configuration there are no net forces acting on the atoms, hence there is no linear term;  $\varphi_0$  is the equilibrium potential energy.

From here, we can write the equations of motion:

$$\begin{aligned} m_n \frac{d^2 u_\alpha^{nl}}{dt^2} &= - \sum_{\alpha' n' l'} \left. \frac{\partial^2 \varphi}{\partial u_\alpha^{nl} \partial u_{\alpha'}^{n' l'}} \right|_0 u_{\alpha'}^{n' l'} \\ &= - \sum_{\alpha' n' l'} \varphi_{\alpha\alpha'}^{nn'(l-l')} u_{\alpha'}^{n' l'} \end{aligned} \quad (3.2)$$

Due to the translational invariance within a crystal, the force  $\varphi_{\alpha\alpha'}^{nn'(l-l')}$

depends only on the distance between the unit cells  $l$  and  $l'$ , so we can drop the  $l$  dependence and consider  $l'$  as the distance to a reference cell.

Because of the periodicity of the system, we look for a solution in the form of plane waves of the form

$$u_{\alpha}^{n'l'} = \frac{A_{\alpha}^n(\mathbf{k})}{\sqrt{m_n}} \exp\left(\mathbf{i}\mathbf{k} \cdot \mathbf{r}^{l'} - i\omega(\mathbf{k})t\right), \quad (3.3)$$

where the frequency depends on  $\mathbf{k}$ , which is now a vector (called the wavevector). Substitution into Equation 3.2 gives

$$\begin{aligned} -\omega^2(\mathbf{k})A_{\alpha}^n(\mathbf{k}) &= \sum_{\alpha'n'l'} \frac{\varphi_{\alpha\alpha'}^{nn'l'}}{\sqrt{m_n m_{n'}}} A_{\alpha'}^{n'}(\mathbf{k}) e^{-i\mathbf{k} \cdot \mathbf{r}^{l'}} \\ &= \sum_{\alpha'n'l'} D_{\alpha\alpha'}^{nn'l'}(\mathbf{k}) A_{\alpha'}^{n'}(\mathbf{k}) \end{aligned} \quad (3.4)$$

that are now a set of algebraic equations.  $D(\mathbf{k})$  is called the dynamical matrix, which is the mass weighted Fourier transform of the force matrix.

The wavevector  $\mathbf{k}$  is an important parameter in a crystal and is associated with a crystal momentum (or quasi-momentum)  $\hbar\mathbf{k}$ . Quantum transitions within the crystal require the quasi-momentum to be conserved: this conservation law, together with the energy conservation in the absorption process, dictates that the interaction of a phonon with light is possible only if  $\mathbf{k} \approx 0$ , due to the low momentum of the photon.

We can therefore drop the exponential dependence in 3.4, leading to the simplified equation

$$\omega^2(0)A_{\alpha}^n(0) = \sum_{\alpha'n'l'} D_{\alpha\alpha'}^{nn'l'}(0)A_{\alpha'}^{n'}(0) \quad (3.5)$$

This is an eigenvalue problem, composed of  $3n_b$  equations, where  $n_b$  is the number of atoms in the unit cell. It is possible to obtain both the eigenvectors  $A_{\alpha}^n(\mathbf{k})$  and their associated frequencies  $\omega(\mathbf{k})$  by standard linear algebra techniques. Three (acoustic) phonons have a frequency  $\omega = 0$  at  $\mathbf{k} = 0$ : these correspond to bulk translation of all the atoms in the unit cell. The remaining  $3n_b - 3$  modes are called optical phonons: they have non-zero frequencies and they can interact with an electric field.

From the resulting eigenvectors  $A_{\alpha}^n(0)$  it is possible to return to the

atomic displacements  $u_\alpha^{nl}$  using the relation 3.3.

Terahertz intensities  $I_N$  are directly correlated with the change in electric dipole moment  $\mu$  of the system with respect to all of the atomic motions under excitation of a phonon mode  $Q_N$  [48]:

$$I_N \propto \left| \frac{\partial \mu}{\partial Q_N} \right|^2 \quad (3.6)$$

The calculation of the cell dipole varies between different methods. If the electronic part of the system is not treated quantum mechanically, as it happens in force field calculations (see chapter 3.4), the dipole variation can be simply related to the rigid displacement of the atomic charges.

In quantum mechanical methods (chapter 3.5) the dipole has to be related to the perturbation of the electronic distribution by the normal mode displacement; to do so, the contribution of each of the  $n$  atoms to the cell dipole variation is expressed using the Born effective charge tensor  $Z_n$ , which is defined as the first derivative of the polarisation  $\mathbf{P}$  per unit cell with respect to the atomic displacements  $\mathbf{u}^{nl}$ :

$$Z_{n,\alpha\beta} \propto \frac{dP_\alpha}{du_\beta^{nl}} \quad (3.7)$$

where the  $\alpha$  and  $\beta$  are used to indicate the Cartesian coordinates.

For comparison with measured terahertz spectra, the results of lattice dynamics calculations are commonly graphically presented as a sum of Lorentzian functions

$$f(x) = \sum_N \frac{I_N}{\pi} \left( \frac{\gamma}{(x - x_N)^2 + \gamma^2} \right), \quad (3.8)$$

centred at the absorption frequencies  $x_0$ , where the relative intensity  $I_N$  calculated from Equation 3.8 equals the area under the curve for each vibrational mode. The parameter  $\gamma$  simulates the signal broadening of the experimental spectrum. In the proceeding chapters we will commonly use this approach for visual comparison of our calculations with the experimental results.

### 3.2.1 Quantisation of the vibrational model

The transformation from the classical to quantum mechanical treatment of the vibrations is straightforward: we can substitute the momentum  $p_i$  and the displacement  $x_i$  for each atom with the equivalent quantum operators  $\hat{p}_i$  and  $\hat{x}_i$  [92]. The coordinate transformation of Equations 3.3–3.5 keep the canonical commutation rules for the normal modes of vibration: as a result we can construct the Hamiltonian of the system, which is a system of independent harmonic oscillators, each characterised by a frequency  $\omega_i$ .

The energy spectrum for each harmonic oscillator is formed by equally spaced levels, separated by an energy  $\hbar\omega_i$ . For each quantum level  $\nu$ :

$$E_\nu = \hbar\omega_i \left( \nu + \frac{1}{2} \right) \quad (3.9)$$

The selection rules for the absorptions dictates that one-photon induced transition in the dipole approximation are only possible if  $\Delta\nu = \pm 1$ .

Without the harmonic approximation we lose the simplicity of the solution: the energy levels are no longer equally spaced, and multi-photon transitions are possible.

### 3.2.2 Temperature effects

The calculation of the normal modes of vibration in the crystal assumes the equilibrium position as a starting point: this corresponds to the hypothetical and motionless state at 0 K. Temperature effects, dependent on the atom motions, are therefore not taken into account by the normal modes calculation.

The modelling of the forces (section 3.3) necessary for the evaluation of the vibrational modes, can contain temperature effects: this is true whenever parameterisation of the forces relies on experimental data (as in the case of empirical force fields).

One simple way to deal with the temperature dependence is by the use of the quasi-harmonic approximation [93, 94], by assuming that the dependence of the frequencies with temperature is contained within the thermal expansion of the crystal. We therefore fix the crystal lattice parameters to the experimental values at the desired temperature, and

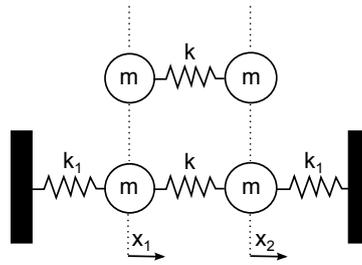


Figure 3.5. Simple system of a diatomic molecule. In the first picture, two equal masses connected with a spring of elastic constant  $k$ . In the second the same molecule interacts with fixed walls through springs of elastic constant  $k_1$ . The displacements  $x_1$  and  $x_2$  are indicated

calculate the vibrational properties for this system from the resulting harmonic frequencies.

### 3.2.3 Normal modes in vacuum and in condensed state

When it comes to molecular solids, it is useful to consider the coupling between the strong bonding within a molecule and the relatively weak intermolecular interactions. As an illustration, consider a model of a homonuclear diatomic molecule: two atoms of mass  $m$  connected by a spring of elastic constant  $k$  (Figure 3.5). In a simple model of this molecule in the condensed phase, the atoms also interact with their surroundings (here, a pair of fixed walls) via additional bonds of spring constant  $k_1$ . This second system is equivalent to the isolated molecule when  $k_1 = 0$ .

The normal modes can be calculated by considering the system of equations

$$\begin{aligned} m \frac{d^2 x_1}{dt^2} &= k_1 x_1 + k(x_2 - x_1) \\ m \frac{d^2 x_2}{dt^2} &= k_1 x_2 + k(x_1 - x_2) \end{aligned} \quad (3.10)$$

and looking for an oscillatory solution for the displacements:

$$\begin{aligned} x_1 &= A_1 e^{i\omega t} \\ x_2 &= A_2 e^{i\omega t} \end{aligned} \quad (3.11)$$

Substitution into Equation 3.10, in matrix representation:

$$\begin{pmatrix} k + k_1 - \omega^2 m & k \\ k & k + k_1 - \omega^2 m \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = 0, \quad (3.12)$$

which has non-trivial solution only if the matrix is singular. The solutions are:

$$\begin{aligned} \omega_1^2 &= k_1/m \\ \omega_2^2 &= (2k + k_1)/m \end{aligned} \quad (3.13)$$

In the case of a molecule in vacuum ( $k_1 = 0$ ) there is only one mode of vibration: stretching of the bond with frequency  $\omega_2^2 = 2k/m$ . For the constrained molecule interacting with the two walls, there is the same stretching mode, but at increased frequency  $\omega_2^2 = (2k + k_1)/m$ , and a second vibration at frequency  $\omega_1^2 = k_1/m$  corresponding to translation of the whole molecule.

The absorption frequencies of intramolecular vibrations are always shifted to higher frequencies, depending on the strength of the intermolecular interactions. The new vibrational modes that are introduced as a consequence of interactions of a molecule with its surroundings correspond to whole-molecule motions. In two- or three-dimensional systems, these can involve molecular rotation as well as the translation seen in this simple one-dimensional example.

### 3.3 SIMPLE PARAMETERISATION FOR IONIC SYSTEMS

It should be clear from the previous sections that, in order to perform a phonon calculation, we need a model for the forces acting between atoms within the system of interest. Various computational methods, with different degrees of approximation, have been applied to evaluate the required forces to model the vibrations in crystals.

One of the earliest calculations of the phonon spectrum of an ionic crystal was attempted by Kellermann [95] in 1940 for the simple cubic salt structure NaCl. This is a simple cubic crystal: each atom is located at the corner of a cube, and each  $\text{Na}^+$  ion is surrounded by 6  $\text{Cl}^-$  neighbours (and vice versa). The system was treated as a field of interacting

classical point charges held apart by a repulsive potential,  $v_0$ , between nearest neighbour atoms, which is necessary to keep the atoms at their equilibrium positions. It is not necessary to know the exact form of this potential, but only its behaviour near to the equilibrium position: the first derivative was calculated from the equilibration of forces within the crystal, knowing that the forces on all atoms in the crystal must vanish at the equilibrium structure. The second derivative of the interatomic potential was determined from experimental data on compressibility of the crystal.

From this model for the interatomic forces, a 6 dimensional (3 dimensions for each atom in the unit cell) dynamical matrix was then constructed and diagonalised. One of the two measurable frequencies of NaCl (two coinciding longitudinal optical phonons, LO) is in remarkable agreement with experimental data (3.3), especially considering the simplicity of the model potential; the other (the transverse optical phonon, TO) is more dependent on the effects of charge displacements (polarisation), which were not included in this first model.

This method is simple enough to be performed on a simple system without the need for a computer (Kellermann produced a phonon spectrum for  $\mathbf{k} \neq 0$  as well!), and has been repeated for a number of structurally equivalent alkali halides and oxides [96, 97].

Unfortunately the model for the interactions between atoms here is too simplified to be generalisable to more complex solids. More sophisticated models have been addressed in a number of publications, considering models to account for polarisation of atomic charges [99, 100] and extended interactions beyond nearest neighbours [101]. Many more adjustable parameters must be introduced to the model of interatomic forces to account for these interactions, and the simplicity of the model is lost. Interesting systems usually consist of more than 2 atoms in the unit cell, so that many different atom-atom interactions must be considered.

Phonon type	Kellermann	Lattice dynamics	Experimental
LO	4.55	4.92	$5.17 \pm 0.03$
TO	9.58	7.84	$7.92 \pm 0.07$

Table 3.3. Phonon frequencies (THz) of the optical phonons of NaCl for  $\mathbf{k} = 0$  by Kellermann compared with experimental values [98] and a calculation that includes charge displacement and polarisation [99]

Furthermore, for crystals of electrically neutral molecules, the interactions between atomic charges are usually not the dominant attractive intermolecular interactions. Instead, a more realistic model of the van der Waals forces is necessary. Overall, it is generally necessary to calculate the phonon spectrum with the aid of a computer, by numerically evaluating the matrix of second derivatives from a detailed description of the forces acting within a crystal.

The quality of a molecular simulation depends of how well the interactions between atoms in the system are represented. Two types of approach have been applied to calculate the necessary forces:

**ATOM-ATOM METHODS** are a generalisation of the approach used in the NaCl example given above. A functional form is assumed for the important interatomic interactions and these functions are parameterised to provide a description of the energy and forces within the crystal. The combination of functional form and parameters is often referred to as a force field. Electrostatic interactions in the atom-atom method are treated classically, usually by atomic partial charges, and sometimes with higher order atomic or molecular multipole expansions.

**QUANTUM MECHANICAL METHODS** consider all of the forces as ultimately arising from the electrostatic interactions between the electrons and nuclei. The electronic problem can be solved by separating the nuclear and electronic wavefunctions (the Born-Oppenheimer approximation) so that the energy and forces acting within a crystal can be calculated for any configuration of its constituent atoms. The most commonly applied electronic structure method in recent years is density functional theory (DFT) for which several commercial packages (see for example [102, 103]) are available to perform this type of calculation.

Electronic structure calculations on periodic systems are orders of magnitude more computationally expensive than those based on the atom-atom force field approach. The techniques necessary for DFT based phonon calculations (density functional perturbation theory [104]), along with the necessary computational power, have only been available since the 1990s. While large-scale computing is nowadays common in research facilities, the cost of such calculations still limits their application to the crystals of fairly small molecules. For large systems, such

as the crystal structures of pharmaceutical molecules with hundreds of atoms within the unit cell, the atom-atom approach is still often the only practical solution. In the next two sections we describe these two methodologies in more detail.

### 3.4 FORCE FIELD METHODS

The atom-atom potential method has been very successfully applied to modelling a wide range of materials, and some of the early development and applications are described by Pertsin and Kitaigorodskii [94].

The main approximation behind a force field method is that the electrons are not treated explicitly, leaving only a classical description of the degrees of freedom and of the interactions. Given the wide range of types of interactions present in a crystal, some other approximations are necessary.

The energy of a system can be expressed as a sum of all the possible interactions over all the atoms:

$$U = \frac{1}{2!} \sum_{i,j=1}^N U_{ij} + \frac{1}{3!} \sum_{i,j,k=1}^N U_{ijk} + \frac{1}{4!} \sum_{i,j,k,l=1}^N U_{ijkl} + \dots \quad (3.14)$$

where the terms represents, respectively, two-body interactions, three-body interactions, and so on. Interactions between more than two atoms are too expensive to consider and are usually left out, except for internal degrees of freedom between first neighbours inside molecules.

**INTRAMOLECULAR POTENTIALS** are used to model covalent bonding interactions, and are necessary to correctly describe the conformation of a molecule. At its most basic, an intramolecular force field consists (see Figure 3.6) of a two-body bond stretching functions depending on the distance  $R$  and the equilibrium distance  $R_0$ , usually modelled either as harmonic in the bond length or using a more realistic Morse potential form:

$$\begin{aligned} U_{ij}^{\text{Harmonic}} &= C_2^{ij} (R - R_0)^2 \\ U_{ij}^{\text{Morse}} &= D_2^{ij} \left( \left( 1 - e^{-A_{ij}(R-R_0)} \right)^2 - 1 \right) \end{aligned} \quad (3.15)$$

The 3-atom angle terms penalise distortions from the preferred

bond angle  $\Phi_0$  (for example  $120^\circ$  in an aromatic ring):

$$U_{ijk} = C_3^{ijk} (\Phi - \Phi_0)^2 \quad (3.16)$$

Four bond interactions put an energy cost on changing the preferred torsion angle  $\theta_0$ : it is the angle between the plane containing the atoms  $i, j, k$  and the atoms  $j, k, m$  (see Figure 3.6). The number of minima in the torsional potential is governed by the parameter  $N$  in the following equation:

$$U_{ijkm} = C_4^{ijkm} (1 + \cos(N\theta - \theta_0)) \quad (3.17)$$

Higher many-body interactions are usually not needed, but their effect might be important in particular cases: Gale notices [105] for example that 6-body interactions may be used to impose planarity conditions on groups of resonant atoms.

Two-body non-bonded interactions can be generally categorised according to the physical interaction they represent, and have different dependence on the distance  $R$  between the two atoms. The most common type of interactions are listed below.

EXCHANGE-REPULSION INTERACTIONS are a sum of two distinct phenomena. Classically, they arise from an unfavourable overlap of electron densities as the internuclear separation between non-bonded atoms is decreased. From the quantum mechanical point of view there is also an attractive contribution related to the sym-

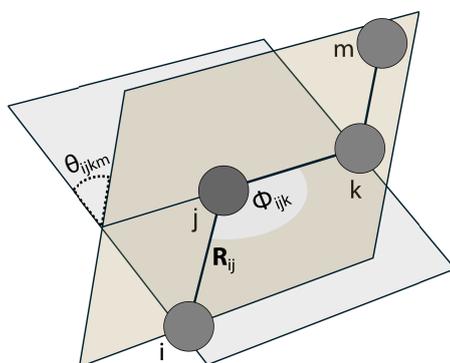


Figure 3.6. Schematic representation of intramolecular parameters between 4 atom  $i, j, k, m$ : the interatomic distance  $R_{ij}$ , the bond angle  $\varphi_{ijk}$  and the torsion angle  $\theta_{ijkm}$

metry exchange between equivalent electrons, and a repulsion effect when two electrons try to occupy the same region of space. The two commonly used forms of the repulsion term are an  $AR^{-n}$  term where  $n$  is usually in the range from 9 to 14, or a theoretically better founded (but less efficient computationally) exponential,  $Ae^{-BR}$ .

DISPERSION INTERACTIONS arise from correlated fluctuations in the electron charge distribution around the atoms, the leading term of which corresponds to fluctuating dipole-dipole interactions and has an  $R^{-6}$  dependence.  $R^{-8}$  and higher terms arise from higher order correlated electron density fluctuations, but are less important and usually omitted.

ELECTROSTATIC INTERACTIONS arise because electronic charge is not uniformly spread within a molecule: its distribution can most simply be modelled by assigning fractional point charges to each atom in the molecule. The electrostatic interaction between point charges has a long range  $R^{-1}$  dependence on interatomic separation and does not depend on the mutual orientation of the interacting atoms. However, some features of the electrostatic potential around molecules cannot be adequately modelled using such a simple spherical atom model. For example, localised lone pairs and aromatic  $\pi$ -electron density introduce important anisotropy in the electron charge distribution around atoms. Therefore, some atom-atom models include higher order multipoles (dipole, quadrupole, etc.) on each atom. See chapter 3.4.1 for the distributed multipole analysis treatment used in this work.

INDUCTION INTERACTIONS are corrections to the electrostatic distribution of a molecule due to the proximity with other molecules and the resulting induced redistribution of electron density within the molecule, and it depends on the molecular polarisabilities. It is usually ignored in simple treatments, because its non-additivity makes it difficult to handle correctly [105]; it can be calculated iteratively.

In most atom-atom calculations, there are three terms in the force field used to describe the interactions between atoms [94]: repulsion-dispersion, electrostatics, and intramolecular terms.

Two common forms of atom-atom repulsion-dispersion terms are:

$$\begin{aligned} U_{ik}^{\text{Lennard-Jones}} &= A^{i\kappa} R_{ik}^{-12} - C^{i\kappa} R_{ik}^{-6} \\ U_{ik}^{\text{Buckingham}} &= A^{i\kappa} \exp(-B^{i\kappa} R_{ik}) - C^{i\kappa} R_{ik}^{-6} \end{aligned} \quad (3.18)$$

where  $R_{ik}$  is the separation between atoms  $i$  and  $k$ . These interactions are determined by the parameters  $A$ ,  $B$  and  $C$ , whose values depend on the types ( $i$  and  $\kappa$ ) of atoms involved.

Force fields differ in which of the above terms are included, their exact functional form, and how the parameters in each term are determined. These parameters depend on the types of atoms that are interacting and are often developed to be transferable between systems with similar chemical functionality: for example, the parameters describing repulsion-dispersion interactions between carbon atoms might be fitted to model any carbon atoms in organic molecules. More elaborate parameter sets might include separate sets of parameters for carbon atoms in different chemical environments, such as different parameters for aromatic and aliphatic carbon atoms. The advantage of such transferable parameter sets is that there is no need to develop a new force field for each new system that is to be studied.

Examples of this approach are the repulsion-dispersion parameters developed by Williams for hydrocarbons [106], oxygen [107], nitrogen [108] and fluorine containing [109] hydrocarbons. The parameters in such transferable force fields were fitted to reproduce structural parameters and heats of sublimation of a large set of molecular organic crystal structures and can therefore be used to describe a wide class of organic molecules.

An alternative approach to these transferable parameters is to develop and optimise specific force fields for a single molecule or small family of molecules [85, 110, 111], sometimes without the need to fit to experimental data. While such molecule-specific models involve much more work, the advantage is that the functional form and parameters can be fine tuned to very accurately describe a particular molecule.

### 3.4.1 Rigid molecule approach

The range of frequencies from the calculation of the vibrational frequencies of a molecular crystal varies from the terahertz range accessible by

our experimental system (4 to  $110\text{ cm}^{-1}$ ) to the infrared range (400 to  $4000\text{ cm}^{-1}$ ). This variability suggests that there are differences between the type of vibrations.

A different absorption frequency means a different energy and force probed by the vibration. In molecular solids the highest absorption frequencies perturb the high energy intramolecular bonds, while at lower frequency the spectrum comprises the so-called “lattice modes”, dominated by translational and rotational movements of full molecules around the molecular centre of mass.

The rigid molecule approximation assumes that inter- and intramolecular interactions are completely uncoupled, so that all of the intramolecular force field terms can be ignored when calculating the lattice mode region of the vibrational spectrum. Therefore only the molecular degrees of freedom, six for each molecule, need to be taken into account in the minimisation of the energy and in the calculation of the vibrational eigenfrequencies.

The approach described in chapter 3.2 needs to be modified to include centre of mass translation and rotation, and the mass normalisation in Equation 3.3 has to be replaced by molecular mass and moments of inertia. The procedure is described by Walmsley [112] and Califano [85].

This simplification can dramatically reduce the dimensionality of the dynamical matrix and the number of needed parameters of the force field, and it lowers the computational cost of the calculation; for a crystal structure with  $Z$  molecules in the unit cell, rigid-molecule lattice dynamics leads to  $6Z - 3$  optical vibrational modes at  $\mathbf{k} = 0$  instead of the  $3N - 3$  modes for a system with  $N$  atoms.

Another advantage of this method is that the electronic distribution, a very important parameter in the description of a molecular system, does not change due to the fixed molecular conformation: it is therefore possible to use a multipole expansion of the electronic density with high level *ab initio* methods and use it throughout the calculation.

The main drawback of the rigid molecule approach is due to the imperfect separation between lattice mode and internal modes. This approximation is valid for a number of molecular systems, like benzene [84] (where the lowest internal vibration is separated from the highest lattice mode by almost  $300\text{ cm}^{-1}$ ), but not strictly true even for relatively small systems like paracetamol. This will be discussed in later chapters.

### *The distributed multipole analysis*

The distributed multipole analysis (DMA) is a method to describe the molecular electron density as a distribution of multipole sites, typically coincident with the atomic positions [113]. It was developed to provide an effective way of describing the electrostatic interaction between two or more molecules.

The electronic distributions of isolated molecules are derived from *ab initio* (or DFT) calculations using Gaussian basis sets in the majority of molecular chemistry packages (such as the one we used, GAUSSIAN 03 [114]). The molecular orbitals are expressed as a linear combination of atomic orbitals, and the electronic density can be ultimately written as a product of atomic orbitals.

The product of two Gaussian functions centred at different origins can be expressed as another Gaussian centred on an intermediate point (which is effectively the best choice for a multipole expansion). The molecular orbitals contain an angular part as well, in the form of spherical harmonics: the product of spherical harmonic functions, of angular momentum  $k$  and  $k'$  respectively, can be expressed as a Clebsch-Gordan series of spherical harmonics spanning angular coefficients  $l$  so that

$$|k - k'| \leq l \leq k + k' \quad (3.19)$$

The maximum exponent in the corresponding multipole expansion is limited by the maximum angular factor: for example two spherical  $s$  orbitals ( $k = k' = 0$ ) can be described by a charge, an  $s$  and a  $p$  ( $k = 1$ ) by a charge and a dipole, and so on.

The DMA procedure provides a way to move the origin from this point to the nearest distributed multipole site by recalculating the multipole expansion coefficients. While the new expansion would not be finite (as the original one) it is found to be quickly convergent; the computational time required is negligible compared to the calculation of the electronic density. The DMA expansion is inherently dependent on the molecular geometry, and therefore it is difficult to apply in simulations where the molecular configuration changes, but it is an excellent choice for rigid molecules calculations, providing a significantly better description than methods relying only to partial charges in the description of the electronic density [115].

### 3.5 SOLID STATE ELECTRONIC STRUCTURE CALCULATIONS

The basic requirement to be able to perform a calculation of the electronic structure is the separation of the nuclear coordinates from the electronic coordinates, using the Born-Oppenheimer approximation. Under this assumption, that holds if the sets of solutions of the electronic Schrödinger equation are well separated, the calculation of the electronic wavefunction of a crystal can be viewed as a system of interacting electrons in a field of nuclei. The equilibrium geometry of the crystal is the configuration where there is no net force acting on each nucleus, and can be found by performing a minimisation of the energy of the crystal with respect to its geometrical parameters (unit cell and atomic positions).

The forces necessary to build the dynamical matrix can be obtained by distortion of the nuclear equilibrium structure: moving an atom away from the equilibrium position generates a restoring force, that does not depend only on the equilibrium electronic distribution, but on its variation as well [116], due to the parametric dependence of the electron wavefunction on the nuclear perturbation. The generation of the dynamical matrix therefore requires the calculation of the electronic density (a computationally expensive task) for a high number of nuclear configurations, and this contributes to the overall significant computational cost of such a calculation.

The calculation of the electronic charge distribution is usually performed using density functional theory (DFT). Hohenberg and Kohn [117] proved that only the electron density (rather than the wavefunction) is necessary to describe the ground state of a system. Furthermore, for a system of interacting electrons in an external potential  $V$  (such as the potential generated by the atomic nuclei in the crystal) there exists a universal functional  $F$  of the electron density  $n$ , independent of the external potential, such that the energy  $E$  is defined as

$$E[n] = F[n] + \int V_e[\mathbf{r}]n[\mathbf{r}]d\mathbf{r} \quad (3.20)$$

The problem behind this formulation is that the functional  $F$ , although guaranteed to exist, is not known exactly.

A practical approach to overcome this difficulty was suggested by

Kohn and Sham [118]: we can consider a system of non-interacting electrons with an external potential  $v_{\text{eff}}$  capable of generating the same charge distribution of our starting system. In this way we can express the functional as a sum of physically recognisable contributions: an exact kinetic energy  $T$ , a Coulomb interaction potential between the electrons  $J$ , the external potential  $V_e$  (the nuclear charges), and a quantity called the exchange-correlation potential  $V_{\text{xc}}$ , containing all the unknown terms:

$$E[n] = T[n] + J[n] + V_e[n] + V_{\text{xc}}[n] \quad (3.21)$$

Since we are looking at a minimum for  $E$  we can use a variational approach to obtain a set of equations we can work with.

To do so, we can express the total charge distribution as a sum of the individual non-interacting wavefunctions  $\phi_i$  for each electron, and work out an effective one electron Hamiltonian the orbitals have to obey:

$$\left( -\frac{1}{2}\nabla^2 + v_{\text{eff}} \right) \phi_i = \varepsilon_i \phi_i \quad (3.22a)$$

$$\begin{aligned} v_{\text{eff}}(\mathbf{r}) &= \frac{\delta}{\delta n} V_e[n] + \frac{\delta}{\delta n} J[n] + \frac{\delta}{\delta n} V_{\text{xc}}[n] \\ &= v(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{\text{xc}}(\mathbf{r}) \end{aligned} \quad (3.22b)$$

where the terms in 3.22b come from the variational minimisation procedure [119], so that the effective potential arises as the functional derivative of  $E$  with respect to the electron density. To avoid confusion we will indicate the functional and the potential respectively with upper and lower case.

The form of Equations 3.22 is similar to what one obtains in the case of a Hartree approach (the best product of orbitals) and Hartree-Fock (best antisymmetrised product of orbitals), with the difference that the exchange-correlation term  $v_{\text{xc}}$  includes the correlation between electrons.

The biggest challenge in the research community has been to come up with a good way to approximate and express the exchange-correlation potential. Again, this term can be split into two contributions: the non-classical exchange interaction coming from the Pauli exclusion principle and the correlation between electrons.

The simplest (and historically the first) approach used is to treat the

electron system like a uniform electron gas of fixed density  $n$ . The exchange contribution, derived by Dirac in 1928 [120] is

$$V_x = \sqrt[3]{\frac{3}{\pi}} n^{4/3}(\mathbf{r}). \quad (3.23)$$

There is no known analytic expression of the correlation contribution, except for the limit of very low [121] and very high density [122, 123], but there are several tabulated values for intermediate densities of interest calculated by quantum Monte Carlo simulations [124]. Different interpolation techniques can be used to obtain different functional forms. For example Perdew and Zunger suggested the form (known as PZ81 [125])

$$V_c^{\text{PZ}}(r_s) = \begin{cases} -\frac{0.1423}{1 + 1.9529\sqrt{r_s} + 0.3334r_s} & r_s \geq 1 \\ -0.048 + 0.0311 \log r_s + \\ -0.0116r_s + 0.020r_s \log r_s & r_s < 1 \end{cases} \quad (3.24)$$

having defined the Wigner–Seitz parameter  $r_s$  as the radius of a sphere whose volume is the effective volume of an electron:

$$\frac{4}{3}\pi r_s^3 = n^{-1} \quad (3.25)$$

while Vosko, Wilk and Nusair (vwn) proposed what is generally considered to be the most accurate uniform-gas correlation available [126]

$$V_c^{\text{vwn}} = \frac{A}{2} \left( \log \frac{r_s}{Z(\sqrt{r_s})} + \frac{2b}{Q} \arctan \frac{Q}{2\sqrt{r_s} + b} + \right. \\ \left. - \frac{bx}{Z(x)} \log \frac{r_s + x^2 - 2x\sqrt{r_s}}{Z(\sqrt{r_s})} + \right. \\ \left. + 2 \frac{bx(b + 2x)}{QZ(x)} \arctan \frac{Q}{2\sqrt{r_s} + b} \right) \quad (3.26)$$

where the function  $Z$  and the parameters  $A, b, c, Q, x$  are

$$Z(z) = z^2 + bz + c, \quad A = 0.0621814, \quad x = -0.409286, \\ Q = \sqrt{4c - b^2}, \quad b = 13.072, \quad c = 42.7198$$

The method described above – Local Density Approximation (LDA), as it is known – is based on the assumption that the exchange-correlation

functional depends solely upon the value of the electron density at each point of space: this is exact in the limit of a slowly varying electron density. It provides a good description of several systems despite its simplicity, also because of cancellation of errors: the exchange energy is underestimated by 10 % and correlation energy (much smaller in absolute value) is overestimated by a factor 2 or more [119]. Nonetheless, the LDA approach is unsatisfactory for molecular systems, where the variation in charge is not smooth, particularly where hydrogen bonds are present [127].

More elaborate functional forms have been developed to improve on the LDA approach. One way to do so is to develop a functional that depends not only on the local density  $n$ , but on its variation as well: this is called the generalised gradient approximation (GGA), and it corrects the exchange functional in Equation 3.23 by a function  $F$ :

$$V_x^{\text{GG}} = -\frac{3}{4}\sqrt{\frac{3}{\pi}} \int n^{4/3} F(s) \, \text{d}\mathbf{r}, \quad s = \frac{|\nabla n(\mathbf{r})|}{n^{4/3}} \quad (3.27)$$

The choice of the function  $F$  defines the structure of the functional. One of the first, the 1-parameter Becke potential (B88) [128],

$$F^{\text{B88}}(s) = 1 - \beta \frac{s^2}{1 + 6\beta s \sinh^{-1} s}, \quad (3.28)$$

with  $\beta = 0.0042$ , was fitted to reproduce rare gas data. Becke shows that it provides an improvement of two orders of magnitude when compared with the LDA exchange energies.

Other formulations take into account more parameters or higher density derivatives (meta-GGA approaches [129]). It is worth mentioning that there are approaches, like the PBE potential from Perdew, Burke and Ernzerhof [130],

$$F^{\text{PBE}}(s) = 1 + a - \frac{a}{1 + bs^2} \quad (3.29)$$

where the parameters are only fit to known limiting cases, and not to experimental properties, giving more of an *ab initio* property than other potentials.

There are different types of correction for the correlation potential as well, with the functional by Lee, Yang and Parr (LYP) being a popular choice. An exchange correlation potential is therefore a combination of an exchange and a correlation functional (for example BLYP is a combin-

ation of the B88 exchange and of the LYP correlation).

Another approach is to consider a weighted blend of functionals (not necessarily from DFT context, i.e. taking exchange from the Hartree-Fock formulation) to construct a different one. One of the most used exchange-correlation potentials, B3LYP [131, 132], is an example of such approach:

$$V_{xc}^{\text{B3LYP}} = a_0 (V_x^{\text{HF}} - V_x^{\text{LDA}}) + a_x (V_x^{\text{GGA}} - V_x^{\text{LDA}}) + a_c (V_c^{\text{GGA}} - V_c^{\text{LDA}}) + V_{xc}^{\text{LDA}} \quad (3.30)$$

where the 3 (hence the name) empirical parameters  $a_0$ ,  $a_x$  and  $a_c$  are determined by fitting the predicted values to a set of atomisation energies, ionisation potentials, proton affinities, and total atomic energies.

### 3.5.1 Dispersion interactions using DFT

The main limitation of current functionals is their unsatisfactory description of the dispersion attraction between molecules, which is often the dominant contribution to the stability of organic molecular solids.

Rare gas systems, for example, are experimentally (and computationally, by post Hartree-Fock methods) characterised by slight attraction between the atoms, but DFT functionals predict repulsion [133]. The long-term interaction, characterised by a  $R^{-6}$  long range limit, is not accounted for. For this reason structural optimisation of molecular organic crystals routinely leads to non-physical unit cell dimensions due to the missing attraction forces; while calculated molecular bonds reproduce the experimental values, intermolecular interactions are typically strongly underestimated.

As an example, the unit cell volume of the explosive material RDX, modelled with the PBE functional [134, 135] gives an energy-converged unit cell with a 20 % larger volume than the experimentally determined unit cell. As this structural distortion relates to large changes in intermolecular contact distances, calculated frequencies of vibrational modes in the terahertz region cannot be accurately modelled in such an energy-minimised structure.

A common workaround when applying DFT to organic solids where the dispersion attraction is dominant is to avoid the optimisation of the unit cell, constraining the lattice vectors to their experimentally deter-

ined values and thus fixing the unit cell volume.

A treatment using post Hartree-Fock perturbation techniques like the Møller-Plesset perturbation theory or coupled cluster methods would be unfeasible, due to the high computational cost. A more affordable solution to the dispersion problem in DFT is to supplement the functional by a set of parameterised atom-atom  $R^{-6}$  terms of the same form as those included in force fields, leading to methods known as DFT-D [136]. A functional can be written as

$$E^{\text{DFT-D}} = E^{\text{DFT}} - s_6 \sum_{i \neq j}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{\text{dmp}}(R_{ij}) \quad (3.31)$$

where the summation is over every possible pair of atoms of the  $N_{at}$  present in the system. The scaling constant  $s_6$  depends on the functional used, while the dumping function  $f_{\text{dmp}}$  avoids singularities for small  $R$  and overcorrections near the nuclei:

$$f_{\text{dmp}} = \frac{1}{1 + e^{-(R_{ij}/R_{\text{vdw}}-1)}} \quad (3.32)$$

with the parameters  $R_{\text{vdw}}$  as Van der Waals radii.

It is clear that the parameterisation is not unique: different choices of the correction are possible, by using different parameters  $C$ .

The main advantage of the DFT-D approach is that while it provides a correct dispersion force it does not significantly increase the computational cost over pure DFT [134]. The effectiveness of the method is strictly related to a correct choice of the parameters  $C^{ij}$ . We further discuss this issue in section 4.2.1.

### 3.5.2 Basis set and implementation

Consider the standard general Schrödinger equation of a system of electrons in a periodic field

$$\left( \hat{T} + \hat{V}(\mathbf{r}) \right) \phi(\mathbf{r}) = E\phi(\mathbf{r}) \quad (3.33)$$

with the total kinetic energy operator  $\hat{T}$  and a potential  $\hat{V}$  with lattice periodicity.

The general properties of a system subject to a periodic potential is

simplified by the Bloch theorem [91]: the electronic wavefunction can be written as a product of a plane waves of wave vector  $\mathbf{k}$  within the first Brillouin zone, multiplied by an appropriate periodic function  $u$ :

$$\phi(\mathbf{k}, \mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{k}, \mathbf{r}) \quad (3.34)$$

The initial problem is now simplified to a calculation of the function  $u$  just over the Brillouin zone.

In practice, the electronic density in a DFT calculation is expanded as a linear combination of basis functions, with the expansion coefficients to be determined as a part of the calculation. While an infinite number of functions might be necessary for mathematical completeness, the number of basis functions that can be practically included is obviously finite. In order to control the resulting unavoidable errors due to the incompleteness of the basis set, it is therefore crucial to choose the right set and the right number of functions.

Plane waves are the obvious choice for a basis set of a periodic system and they are used by a variety of computational packages. They have a simple mathematical form and they do not need complicated expressions to ensure orthonormality of the basis set. Furthermore, the relation between  $\mathbf{k}$  and the kinetic energy provides a straightforward approach to systematically improving the basis set completeness, by adding higher kinetic energy plane waves until the calculated properties of interest converge to the required level.

From the computational point of view, the integrals containing the products between  $e^{i\mathbf{k}\cdot\mathbf{r}}$  terms and function  $V$  are just the Fourier transform of  $V$ , for which efficient subroutines like Fast Fourier Transform (FFT) have been developed.

Among the disadvantages of plane waves are their poor description of localised states (it is necessary to introduce pseudopotentials for the description of tightly bound core electrons – see next section) and the dependence of the basis functions on the dimensions of the unit cell. The latter means that large changes in the unit cell volume on energy minimisation will change the kinetic energy cutoff and therefore the consistency of the calculation.

The calculation of lattice properties needs the calculation of integrals over all the points of the reciprocal space of the crystal. This is usually performed numerically by choosing a number of sample  $k$ -points in a

regular grid, called a Monkhorst-Pack grid. The amount of  $k$ -points is a trade off between speed and accuracy: higher number of points means a better sampling, at a price of a slower calculation.

### *Pseudopotentials*

The use of plane waves, although providing many advantages, is flawed by the so called “variational collapse” problem [137]: the description of strongly localised atomic states, of radius  $r$ , requires wave vectors in reciprocal space of the order of  $2\pi/r$ . For a system like crystal silicon the number of required basis set functions to achieve an accurate description of the system would be in the order of  $10^6$  to describe its  $1s$  orbital [91], involving dealing with matrix of order  $10^6 \times 10^6$ , an impossibly high value for computational purposes.

From the chemical point of view we are usually interested in the valence electron density, responsible for the chemical properties of bonding. Inner core electrons, tightly bound to the nuclei, are not strongly affected by the differences in the chemical environment. For this reason, these electron are not usually considered explicitly in the calculation, but they are considered as “part” of the nucleus core.

A practical approach is to define an effective potential  $V_{\text{pseudo}}$  to describe the valence electrons in the field of nuclei and inner core electrons, without caring to describe the true electron density in the atom core region.

As a first step, each of the atom types is treated individually to find a numerical solution of the all-electron atomic problem with potential  $V$ . As a second step we need to look for the functional form of the pseudopotential, fitting the parameters  $\lambda_i$  (as few as possible) that reproduce the energy levels and wavefunction behaviour out to a cutoff radius  $r_c$ :

$$V_{\text{pseudo}}(r) = \begin{cases} V_1(r, \lambda_i) & r \leq r_c \\ V(r) & r > r_c \end{cases} \quad (3.35)$$

Following on from the first very simple forms (constant for  $r < r_c$ ) as suggested by Ashcroft [138] for the sodium atom, there has been a lot of research and improvements. One great advance in the quality of simulations has been the use of norm-conserving pseudopotentials [139, 140]: these enforce the condition that the normalised wavefunc-

tion and pseudowavefunction have to be the same for  $r > r_c$ , meaning the electron charge inside the core radius have to be the same. This approach involves dealing with wavefunctions rather than potentials, with the effect that they are usually dependent on the angular momentum  $l$  as well.

Using different pseudopotentials means also having to use a different number of plane waves to achieve a correct representation of a system; soft and Vanderbilt's ultrasoft pseudopotentials release the norm-conserving criteria in order to have a smoother function [141]. Although computationally slightly more expensive because of the inclusion of other terms [142], its use is justified by the need of a smaller wavevector cutoff for big systems, which overcomes its complexity. However, there is usually a trade-off between needing less wavefunctions and being able to use them in different chemical environment (reduced "transferability" of the pseudopotential) [143].

## ANALYSIS METHODS

---

FOR THE SIMULATION of the absorption spectra we need a practical implementation of the ideas we discussed in the previous chapter. We describe in more detail the programs we choose for these calculation: a package that uses the empirical forcefield ideas (DMACRYS), and one based on the plane waves, periodic DFT approach (CASTEP). We report the information needed for the simulations, and the tools we used and develop to analyse the vibrational eigenvectors and to compare the results with experimental data.

### 4.1 THE ATOM-ATOM CALCULATIONS

The programs we used for our lattice dynamics calculations are DMAREL and its evolution DMACRYS<sup>1</sup>. These programs [144, 145] model crystals of rigid molecules, allowing lattice energy minimisation and the calculation of vibrational properties. They are designed to use anisotropic atom-atom model intermolecular potentials and the DMA electrostatic model.

The information on the unit cell and crystallographic atomic coordinates is obtained from the CSD, and it is stored in the form of a SHELX file. This contains information on the unit cell, the symmetries and the inequivalent atoms [146]. There is no information on the bonds between the atoms, which is directly calculated by NEIGHCRY (a program that helps the generation of the input file for DMACRYS) according to a custom cutoff distance between atoms.

The rigid molecule approximation is effectively taken into account by the use of local coordinates: each of the atomic coordinates and atomic multipoles of a molecule is written according to a set of local axis system,

---

<sup>1</sup> The main difference between the two programs lies in the ability to consider crystals with bigger unit cells.

and does not vary throughout the minimisation: the only change is in the position and the orientation of the local axis system with respect to the crystallographic axes (or global axis system).

The multipole expansion is obtained from the molecular charge density matrix calculated with the quantum chemical program GAUSSIAN 03, for which we used DFT calculation with the B3LYP functional and 6-31G\*\* basis set<sup>2</sup> [150–153]. The orbitals are stored (alongside with other information) in an ascii .chk checkpoint file, and converted to multipoles using the program GDMA2 [154].

The interatomic potential used is an empirical Buckingham atom-atom repulsion-dispersion potential. The older calculations were performed with the FIT potential, an evolution of the Williams parameterisation already mentioned at page 41, with the parameters for polar hydrogen parameterised by Price and co-workers [155, 156]. In our most recent calculations we used the newer w99 potential [157–159].

Among the advantages of the w99 potential is a wider range of atom types, effectively describing the chemical difference between differently bonded species (i.e., different parameters for carbon atoms with different coordination). The centres of interaction for the hydrogen atoms are shifted 0.1 Å from the atom along the bond to better model the effective position of the electron density; this is automatically taken into account by NEIGHCRY in the setting up of the calculation. The parameters are reported in Table 4.1.

The standard minimisation process simultaneously relaxes both the unit cell parameters and the molecular positions and orientations, looking for a minimum of crystal enthalpy (constant pressure). The symmetry of the unit cell is taken into account to reduce the dimensionality of the minimisation problem: constraints are eventually applied to preserve the space group of the initial structure.

The electrostatic contributions are summed up to infinity using the Ewald summation method [160] for charge-charge, charge-dipole and dipole-dipole interaction, while for the higher order terms a direct space cutoff of 15 Å is used. The same cutoff is applied to the non-electrostatic terms as well.

---

<sup>2</sup> The 6-31G\*\* is limited up to Krypton in GAUSSIAN: for molecules containing Iodine we used the CEP-121 basis set [147–149]

Potential type	Atom pair	$A^u$ kJ mol <sup>-1</sup>	$B^u$ Å <sup>-1</sup>	$C^u$ kJ Å <sup>6</sup> /mol	Description of the atom type
W99	C <sub>2</sub> ··· C <sub>2</sub>	103 235	3.60	1435.09	Carbon atoms bonded to two atoms
	C <sub>3</sub> ··· C <sub>3</sub>	270 363	3.60	1701.73	Carbon atoms bonded to three atoms
	C <sub>3</sub> ··· C <sub>3</sub>	131 571	3.60	978.36	Carbon atoms bonded to four atoms
	H <sub>1</sub> ··· H <sub>1</sub>	12 680	3.56	278.37	Hydrogen atoms bonded to a carbon atom
	H <sub>2</sub> ··· H <sub>2</sub>	361.30	3.56	0	Hydrogen atoms in an alcoholic group
	H <sub>3</sub> ··· H <sub>3</sub>	115.70	3.56	0	Hydrogen atoms in a carboxylic group
	H <sub>4</sub> ··· H <sub>4</sub>	764.90	3.56	0	Hydrogen atoms in a N–H group
	N <sub>1</sub> ··· N <sub>1</sub>	96 349	3.48	1407.57	Nitrogen atoms involved in a triple bond
	N <sub>2</sub> ··· N <sub>2</sub>	102 369	3.48	1398.15	As N <sub>1</sub> , but with no other nitrogen bonded
	N <sub>3</sub> ··· N <sub>3</sub>	191 935	3.48	2376.55	Nitrogen atoms bonded with one hydrogen only
	N <sub>4</sub> ··· N <sub>4</sub>	405 341	3.48	5629.82	Nitrogen atoms bonded with two hydrogen atoms
	O <sub>1</sub> ··· O <sub>1</sub>	241 042	3.96	1260.73	Oxygen atoms bonded with one atom only
	O <sub>2</sub> ··· O <sub>2</sub>	84 623	3.96	1285.87	Oxygen atoms bonded with two atoms
	FIT	C ··· C	369 743	3.60	2439.8
N ··· N		254 529	3.78	1378.4	All nitrogen atoms
O ··· O		230 064.1	3.96	1123.59	All oxygen atoms
H <sub>1</sub> ··· H <sub>1</sub>		11 971	3.74	136.4	Hydrogen atoms bonded to a carbon atom
H <sub>2</sub> ··· H <sub>2</sub>		5029.68	4.66	21.50	Hydrogen atoms bonded to a nitrogen atom
H <sub>3</sub> ··· H <sub>3</sub>		2263.3	4.66	21.50	Hydrogen atoms bonded to a oxygen atom

Table 4.1. Parameters  $A^u$ ,  $B^u$ ,  $C^u$  of the exp-6 w99 (top) and FIT (bottom) potentials. Only the parameters for homonuclear pairs are reported: the heteronuclear parameters are obtained by the combination rules  $A^{ik} = (A^i A^{kk})^{1/2}$ ,  $B^{ik} = 1/2(B^i + B^{kk})$  and  $C^{ik} = (C^i C^{kk})^{1/2}$

#### 4.1.1 Implementation details

Usually the determination of hydrogen atom positions via X-ray diffraction is not reliable, because of their low scattering power and because the corresponding electron density is smeared out and asymmetrical [161]. Furthermore, hydrogen atoms frequently have larger librational amplitudes than other atoms.

The experimental hydrogen atom positions were therefore recalculated with a DFT calculation on the isolated molecule, keeping all the heavier atom positions fixed, and the resulting coordinates put back in the SHELX file and used for the DMACRYS calculation.

The lattice energy minimisation process uses the Newton-Raphson method [143] in order to obtain a relaxed structure. It is crucial that the process of minimisation does not drastically change the unit cell, as large differences would indicate that the method was unsuitable (i.e. underestimation of specific interactions due to bad parameters in the potential, or inaccurate electrostatics). A useful parameter to assess the similarity between the initial and final crystal structure is the Gavezzotti's discrepancy factor  $F$  [162]:

$$F = \left(\frac{1}{2}\Delta\theta\right)^2 + (10\Delta x)^2 + \left(100\frac{\Delta a}{a}\right)^2 + \left(100\frac{\Delta b}{b}\right)^2 + \left(100\frac{\Delta c}{c}\right)^2 + (\Delta\alpha)^2 + (\Delta\beta)^2 + (\Delta\gamma)^2, \quad (4.1)$$

where  $\Delta\theta$  is the total rigid body rotational displacement of the molecules (in degrees),  $\Delta x$  is their total rigid body translational displacement (in angstroms), and the remaining terms correspond to the variation of the unit cell length  $a$ ,  $b$ ,  $c$  (in angstroms) and angles  $\alpha$ ,  $\beta$ ,  $\gamma$  (in degrees). The weighting factors are necessary to keep the relative contributions on the same scale.

A value  $F = 0$  means that the initial and final structures are identical; if  $F < 50$  we considered the minimisation to be well behaved. As an example a value  $F = 50$ , split to each term, averages to  $\approx 6$  per degree of freedom: this corresponds to  $2.2^\circ$  for a unit cell angle, 2.2% for a unit cell axis, 0.22 Å for  $\Delta x$  or  $4.4^\circ$  for  $\Delta\theta$ . Larger values were obtained in the minimisation of hydrates, due to bigger reorientation factors of the small water molecules.

The standard output of phonons eigenvectors from DMACRYS is rather unfriendly: a 3-column list of numbers of all the  $x$ ,  $y$ ,  $z$  molecular translations and  $R_x$ ,  $R_y$ ,  $R_z$  rotations, the latter in the reference set of the principle axes of inertia, themselves expressed in local coordinates. Furthermore, while the vibrational frequencies and eigenvectors are determined by DMACRYS, the calculation does not provide an estimate of the infrared intensities.

For analysis purposes we used and expanded a pre-existing utility for the analysis of DMAREL output, RUDOLPH (Rigid Units Display Of Lattice Phonons [163]), capable of extracting the relevant information from the DMACRYS output, expressing the overall data in global Cartesian coordinates and writing it into a `.xyz` format, readable by the program JMOL [164], a powerful Java based visualiser. Furthermore, RUDOLPH uses the fixed electrostatic distribution and the phonon eigenvectors to calculate the unit cell electric dipole derivative (proportional to the infrared absorption) with respect to each eigenvector.

## 4.2 THE SOLID STATE DFT CALCULATIONS

The package we use for our solid state quantum mechanical calculations is CASTEP [165]. It uses plane waves as a basis set for valence electron densities, and a wide number of possible pseudopotentials to describe core electrons.

The relevant processing parameters for a CASTEP job is contained in two files, `.cell` and `.param`, separating the information about the structure from the directives to the program.

The `.cell` file contains all the information about the atomic positions and the unit cell, symmetry (which the program will try to guess if coordinates are given with sufficient accuracy), the location of the pseudopotentials to be used and the number of  $k$ -points to use for the integration. The `.param` file has the information on the task to be performed (geometry optimisation or phonon calculation), the functional to be used, the energy cutoff for the number of wave planes, and the tolerances used for convergence of the lattice energies, atomic forces and atomic positions.

It is important to check for the convergence of lattice energy with re-

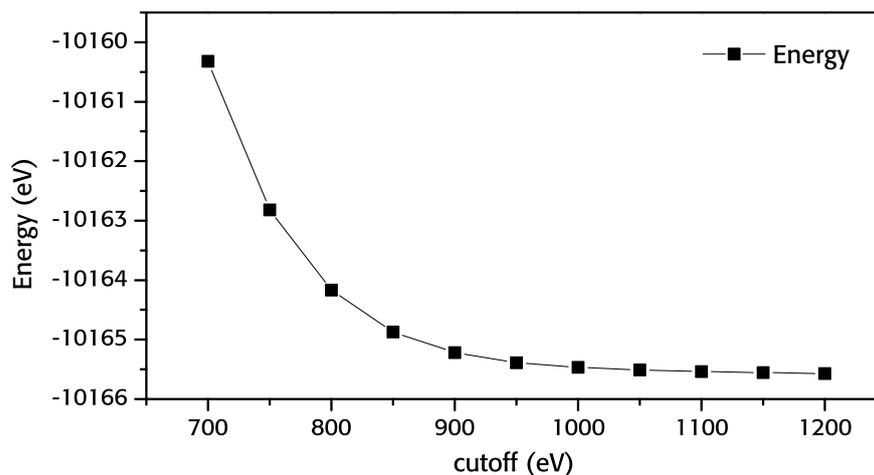


Figure 4.1. Energy of paracetamol form I at different basis set cutoff energies, PBE xc functional

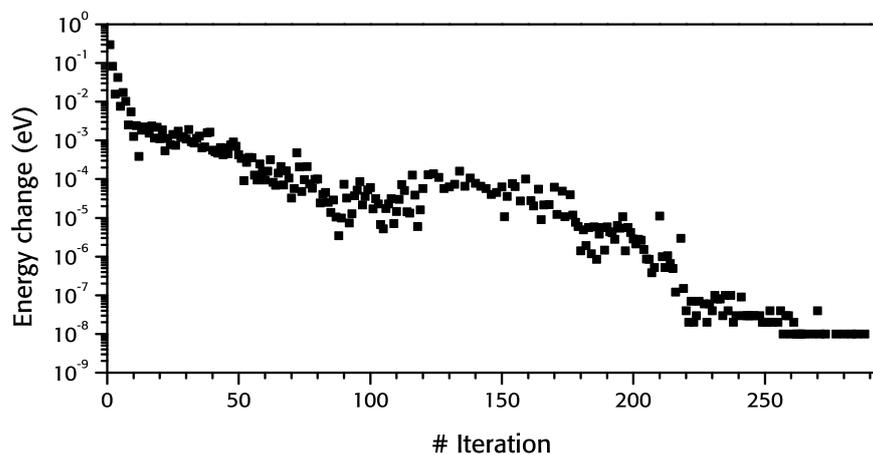


Figure 4.2. Variation of the lattice energy after every minimisation step for paracetamol form I, CASTEP 4.1, PBE functional, 1050 eV basis set cutoff

spect to several parameters, running a series of fast jobs to determine the best choice: as we mentioned before, a small number of  $k$ -points (or a low plane wave energy cutoff) makes the calculation faster, but can lower the precision of the calculation, and vice versa.

For example, figure 4.1 shows the lattice energy of paracetamol form I as a function of the cutoff energy. It can be noted that a basis set cutoff energy of 1000 eV provides a reasonably converged lattice energy, which is considerably higher than the 300 eV that it is needed for simpler systems, like  $\text{SiO}_2$ . As a comparison, the total energy variation during minimisation for paracetamol form I (see figure 4.2) is about 0.5 eV.

For consistency and ease of comparison between our data and other published papers we have used a cutoff energy of 1200 eV, unless stated otherwise; this cut-off energy resulted in well converged lattice energies in all our systems. The Monkhorst-pack grid of  $k$ -points was fixed to be at least  $0.05 \text{ \AA}^{-1}$ .

As for DMACRYS, it is necessary to run a minimisation job before the calculation of the phonon structure: in this case, since there are no rigid molecule constraints, the number of modes will be higher, and the calculation much longer.

The output file for a phonon calculation is a `.phonon` file containing geometry information, the phonon eigenvector and eigenvalues, and the corresponding infrared absorption.

The program is complemented with a pack of PERL utilities to convert the output files in other formats. In particular we used the CASTEP2SHELX to convert the CASTEP geometry to a crystallographic output, and the PHONON2XYZ to have a file useful for visualisation.

#### 4.2.1 Dispersion corrections

In some of our earliest calculations, using the available versions of CASTEP (4.1–5.0), there was no possibility to introduce a dispersion correction to the DFT calculation: all the spectra were obtained by constraining the unit cell dimensions.

Starting from version 5.5 (released in 2010) it is possible to apply a dispersion correction in the calculation, in four different flavours for the choice of the  $c_6$  coefficients [166–169]: we opted for the predefined option developed by Grimme [167].

The  $c_6$  values in this approach are derived from the London formula for dispersion [170], starting and improving a previous parameterisation by Wu and Yang [171]. By using calculations of atomic ionisation potentials  $I_p$  and static dipole polarisabilities  $\alpha$ , the  $c_6$  coefficient for an atom  $a$  is given by

$$c_6^a = 0.05NI_p^a\alpha^a \quad (4.2)$$

where  $N$  has values 2, 10, 18, 36, and 54 for atoms from rows 1–5 of the periodic table. The constant in the equation was chosen to fit binding values of rare gas dimers and a training set of molecules. Values for pair

of elements  $a$  and  $b$  are obtained by use of the geometric mean

$$c_6^{ab} = \sqrt{c_6^a c_6^b} \quad (4.3)$$

### 4.3 COMPARISON WITH THE EXPERIMENTAL DATA

The eigenvalues of our phonon calculation, together with the corresponding infrared absorption, can be plotted as a Lorentzian function (as shown in Equation 3.8, page 32) for an easier comparison with the experimental data. The main choice we have to make is to decide the width parameter  $\gamma$ , as a different choice has a great impact on the resulting simulated spectra.

We report in figure 4.3 the DMACRYS calculated spectrum of theophylline form I (CSD reference code BAPLOT01). Some of the spectral features are already too weak or too close to each other to be separated from the signal at  $\gamma = 1 \text{ cm}^{-1}$ ; at  $\gamma = 4 \text{ cm}^{-1}$  some of the features become shoulders, while at  $\gamma = 16 \text{ cm}^{-1}$  we can count only three peaks with broad shoulders.

Since we are not interested in how the area under the curve changes with  $\gamma$ , we have normalised all calculated spectra so that the most prominent absorption has 1 as fixed height.

In the experimental peaks at low temperature  $\gamma$  are usually in the range of  $2 \text{ cm}^{-1}$  to  $3 \text{ cm}^{-1}$ , while for room temperature comparisons we have used  $6 \text{ cm}^{-1}$  as a maximum value.

### 4.4 RIGID-MOLECULE ANALYSIS

The main challenge we have to face while comparing the output of a phonon calculation from simulations using two different methods is the comparison of the data: first of all, lattice energy minimisation with DMACRYS and CASTEP results in different unit cell dimensions and atom positions from each simulation, since we are using two very different methods; furthermore, we have two different sets of phonon eigenvectors (atomic vs molecular) to compare with each other.

We are interested in answering a set of questions:

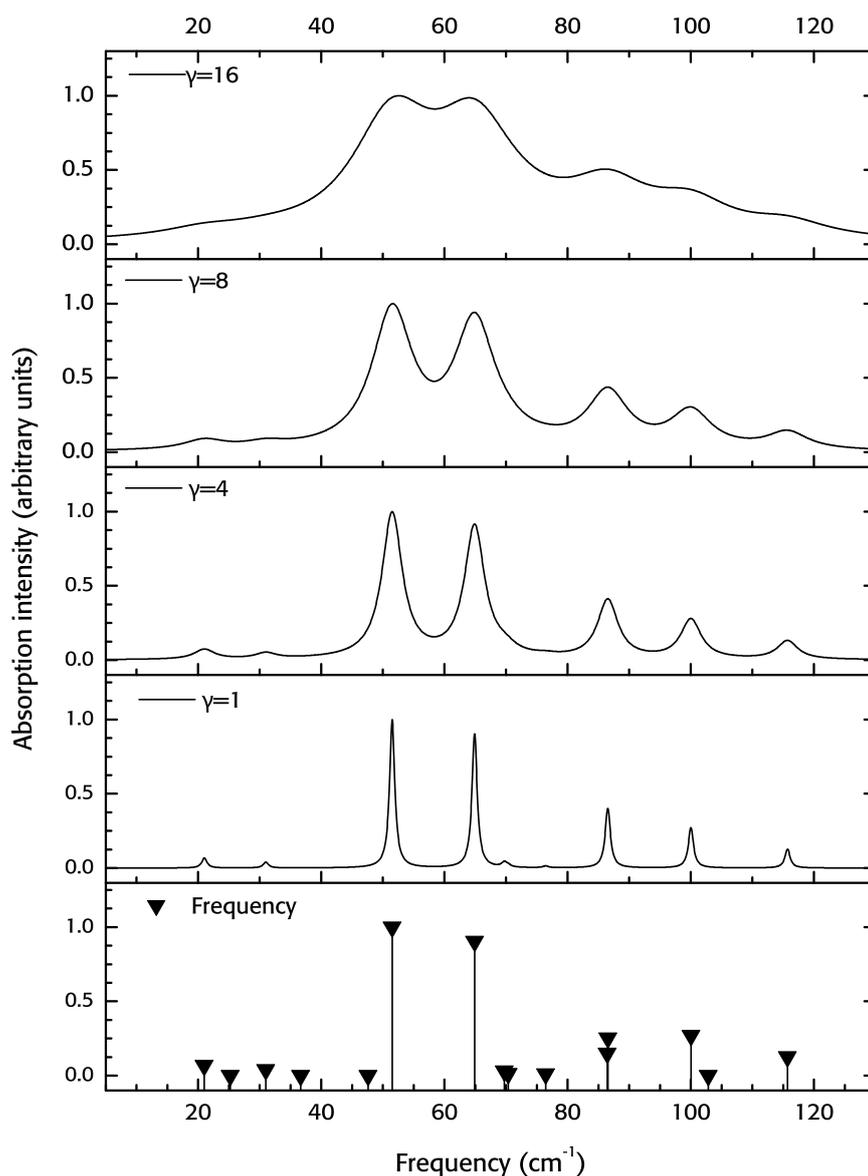


Figure 4.3. Calculated non-zero absorption frequencies of theophylline form I with DMACRYS, w99 forcefield. From the bottom: the position of the absorption features and Lorentzian envelopes with  $\gamma = 1, 4, 8, 16 \text{ cm}^{-1}$

1. How similar are the two calculated sets of eigenvectors?
2. Is the rigid molecule approximation valid in the terahertz range for our systems? If not, when does it fail?
3. What is the range for the validity of the rigid molecule approximation?

To answer the first question we developed a method to quantify the similarity between the eigenvectors, by converting the DMACRYS rigid-

molecule eigenvectors to atomic displacement eigenvectors. We use the fact that minimisation with both methods results only on slight variation of atomic and molecular positions within the unit cell.

The two programs use different conventions for the expression of the unit cell in the orthogonal coordinate system; while `DMACRYS` puts the  $c \parallel z$ , the  $x$  axis parallel to the reciprocal  $a$  axis and the  $y$  axis forming a right-handed set with  $x$  and  $z$ , `CASTEP` has  $a \parallel x$ ,  $b$  in the  $xy$  plane and  $z$  forming a right-handed axis system; furthermore, upon minimisation the unit cell is not constrained to these initial orientations.

For this reason we performed a preliminary rotation to express all calculated eigenvectors in the `CASTEP` notation, so that the orthogonal coordinates and displacements of the atoms are comparable.

In `DMACRYS` a phonon eigenvector for a crystal with  $Z$  molecules within each unit cell is composed of a set of  $6Z$  vectors, each with  $6Z$  components (3 translations and 3 rotations for each molecule). This set of vectors is by construction a basis for rigid displacements of molecules in the crystal for  $\mathbf{k} = 0$ , orthonormal for the properties of the eigenvalue system in Equation 3.5.

The actual translations and rotation are derived by scaling the eigenvectors respectively by molecular mass and moment of inertia, therefore losing their orthonormality. Atomic displacements can be obtained straightforwardly from the molecular centre of mass translation (the same for every atom in a molecule) and the rotation (in the reference frame of the axes of inertia, from the centre of mass of each molecule). Finally, each atomic displacement has to be weighted by atomic mass and the whole vector renormalised.

The resulting mass-weighted vectors are still a complete set over the possible rigid movements of the molecules in the unit cell. By projecting `CASTEP` eigenvectors over the `DMACRYS` space we can quantify each vector contribution, and the overall percentage of each `CASTEP` eigenvector that can be described by rigid-molecule motions.

By the process exposed before we have no guarantee that the resulting mass-weighted `DMACRYS` vector expressed as atomic displacements are still orthogonal to each other; therefore the formulas for expansion will be slightly more cumbersome. By defining the `DMACRYS` vectors  $|d_i\rangle$  and `CASTEP` vectors  $|c_i\rangle$ , we can write the expansion as

$$|c_i\rangle = \sum_j a_j |d_j\rangle + b |\text{non-rigid}\rangle \quad (4.4)$$

where the non-rigid contribution comprises of the sum of all the internal vibrations over all the molecules.

By standard projection on each DMACRYS vector  $\langle d_k|$  we obtain the set of equations

$$\begin{aligned} \langle d_k|c_i\rangle &= \sum_j a_j \langle d_k|d_j\rangle + b \langle d_k|\text{non-rigid}\rangle \\ &= \sum_j a_j Q_{kj} + 0 \end{aligned} \quad (4.5)$$

where we have used the fact that the DMACRYS rigid-molecule vectors must be orthogonal to any non-rigid component. The set of coefficients  $a$  can be obtained by resolving the set of Equations 4.5, inverting the matrix  $Q$ , to obtain:

$$a_k = \sum_j Q_{kj}^{-1} \langle d_j|c_i\rangle \quad (4.6)$$

where  $a_k$  is a direct measure of the similarity between vector  $c_i$  and  $d_k$ .

The actual amount of rigid body contribution is better expressed by the the *square* of the norm of the resulting vector:

$$\begin{aligned} \langle c_i|c_i\rangle &= \langle \text{rigid}|\text{rigid}\rangle + \langle \text{non-rigid}|\text{non-rigid}\rangle \\ 1 &= \sum_{jk} a_j a_k \langle d_j|d_k\rangle + b^2 \langle \text{non rigid}|\text{non rigid}\rangle \\ 1 &= \sum_{jk} a_j a_k Q_{jk} + b^2 \langle \text{non rigid}|\text{non rigid}\rangle \end{aligned} \quad (4.7)$$

In all of our calculations, the transformation of the vectors left the orthonormality untouched. While we have not proven it<sup>3</sup>, this allows us to use the simplified relations arising from  $Q_{ij} = \delta_{ij}$ .

The whole process of conversion of the coordinates is again performed using RUDOLPH (see above), expanded with a subroutine section for this purpose. The actual comparison is performed with the aid of a spreadsheet, constructing a table with the values  $a$  for all the vectors as the

<sup>3</sup> The demonstration is straightforward for the translation part, but dependence of the rotational part on the principle axes of inertia makes it complicated to handle

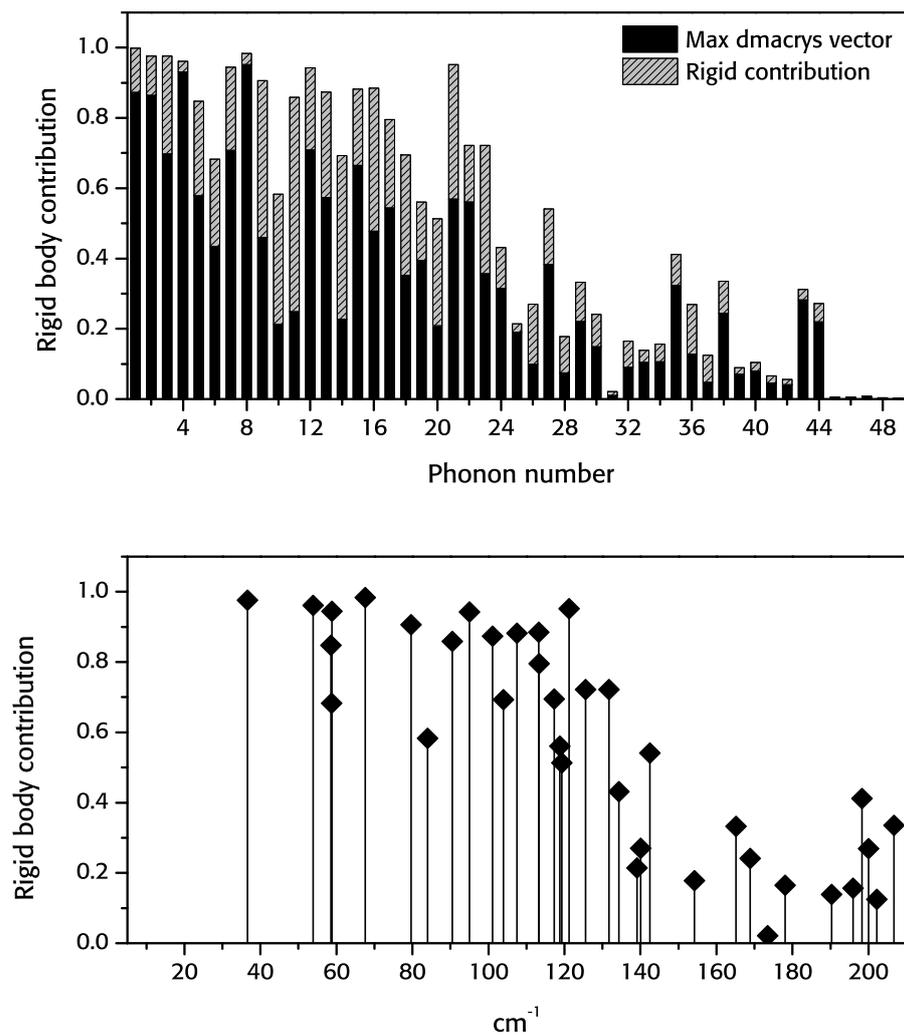


Figure 4.4. Graphic representation of the amount of rigid body contribution in the phonon normal modes calculated using CASTEP 5.5, dispersion corrected PBE, for paracetamol form I. Top, stack graph with the most representative DMACRYS phonon. Bottom, plot against the frequency of the phonon (non IR active vibrations as well)

equation above shows. The resulting matrix has a block structure, since there is no possible mixing of phonons with different symmetries.

For visual inspection of the amount of the rigid body contribution in the CASTEP eigenvectors we used two types of graph: one is a scatter plot of the amount of rigid body contribution against the phonon frequency, which is useful to check the rigidity of the eigenvector motions across the spectrum; the other is a bar chart, highlighting similarity between CASTEP and DMACRYS eigenvectors. Figure 4.4 shows examples of these two types of plot, which will be used in later chapters.

## MODELLING POLARISATION BY THE CRYSTAL ENVIRONMENT

---

**A**T PRESENT IN DMACRYS, our rigid molecule force field based package, the electron density distribution of each molecule in the unit cell is considered to be the same as an isolated molecule; this is an approximation due to the difficulty of calculating the induction effect due to the crystal environment, as mentioned in chapter 3.4.

In this work we are interested in calculating the effect of the change in molecular electron distribution due to induction, the resulting influence on the intermolecular forces in the crystal, and ultimately the phonon eigenvectors and eigenvalues. The polarisation effect is known to be non-negligible for the electronic distribution of organic molecules (see [172]), but we expect it to be particularly important for hydrate crystal structures, because water has a significantly enhanced electric dipole in the condensed state [173, 174].

### 5.1 METHODS

Of course, to fully consider the effective electronic distribution and the distributed multipoles from an *ab initio* point of view we would have to use a periodic electronic structure calculation, which rather defeats the purpose of using an atom-atom method. The computational cost of a calculation considering the induction effect on a “probe” molecule with a surrounding shell of atoms around it quickly gets too high, due to the more-than-linear dependence of computational cost with the number  $n$  of electrons of DFT and even higher cost for other correlated electronic structure methods.

The problem is circumventable by treating the molecules around the probe as a classical distribution of charges, rather than treating the en-

tire system (central molecule and surroundings) quantum mechanically. The `CHARGE` option in `GAUSSIAN 03` [175, 176] allows the calculation of the electronic properties of a molecule in a field of electric charges, without dramatically increasing the computational cost of the calculation.

Using this feature it is possible to use the atomic charges resulting from an isolated molecule calculation, calculate the position of a cloud of atoms surrounding this molecule in the crystal, and put the calculated charges in the corresponding atomic positions; a new calculation would then provide a new set of charges, and this is repeated until we obtain a set of self consistent charges, as shown in diagram of Figure 5.1. Finally, a distributed multipole analysis is performed on the final charge density matrix, and the resulting atomic multipoles are used in the `DMACRYS` calculation<sup>1</sup>.

It is only necessary to iterate the charges on each molecule in the asymmetric unit cell, because the symmetry in the crystal guarantees that the charge distribution on all other molecules is identical to one of the molecules in the asymmetric unit. We have to be careful in achieving self consistency in case there is more than one molecule in the asymmetric unit cell; in each iterative cycle we use the charges of the molecules from the previous one, and update all of them at the end of it, so that all the molecules are calculated consistently with each other.

Initially we set up a `FORTRAN 90` program to calculate the `GAUSSIAN` input for the cloud of atomic charges (which we will refer simply as the cloud) from the `DMACRYS` output and from the knowledge of the atomic charges; since all the remaining operations (extract the charges to be put back, checking the convergence for each molecule) were performed manually, the whole process was quite cumbersome. Recently we have expanded the `PERL` script `SANTHA` – see next section – to automatically execute all the steps.

The type of charge to use should describe as closely as possible the field produced by the electronic distribution of the molecule; for this reason `DMA` charges are not ideal, because they are accurate only if we include higher terms of the expansion, which are not implementable in the `GAUSSIAN` cloud, while Mulliken charges [177] are not meant to describe the electrostatic potential outside the molecule [178].

<sup>1</sup> The molecular orientation is kept fixed throughout the calculation using the same local axis set used in `NEIGHCRY`, since the multipoles are orientation-dependent

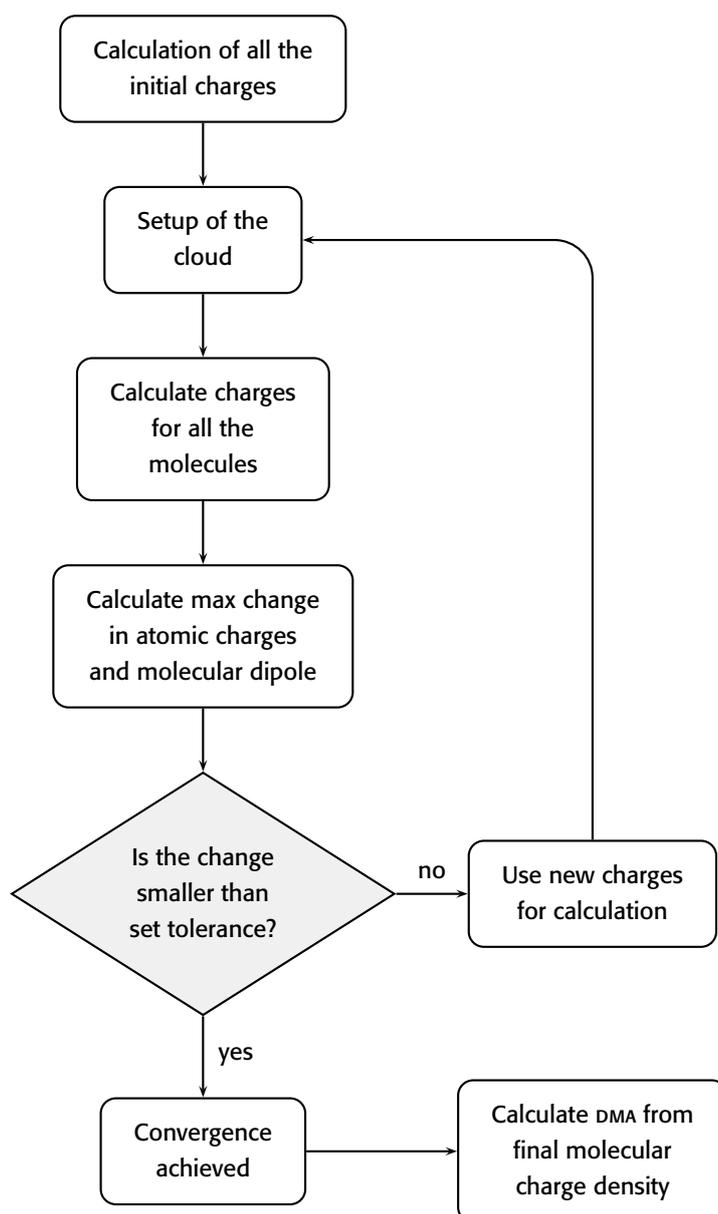


Figure 5.1. Iterative calculation of the electron distribution for each of the molecules in the system

We decided to use the method due to Breneman, also called CHELPG (CHarges from Electrostatic Potentials using a Grid based method) [179], where atomic charges are fitted to reproduce the molecular electrostatic potential at a number of points on a grid around the molecule. Its main setback is in the treatment of big systems because the evaluation of the potential is not affected significantly by the innermost atoms, whose fitting is therefore not meaningful: however this is not relevant for our molecules, relatively small and without buried atoms.

It can be noticed that the converged charges have a sharp variation

only up to 10–15 Å, finally converging below 2 % at 25 Å. The molecular dipole, instead, is subject to larger changes; for this reason, we have decided to choose 30 Å as the cloud radius for all of our calculations.

The tolerance parameter depends on what degree of approximation we are interested in during the calculation: the maximum possible variation of the electronic charges detected (or at least available to the user) in GAUSSIAN is  $1 \times 10^{-6} e$ , but our calculations do not show a detectable variation in predicted IR intensity for a tightness smaller than  $1 \times 10^{-4} e$ , (see Table 5.2) which is the one we selected.

There are some factors to keep into account in this calculation: since we want to keep the charge neutrality we have to include full molecules<sup>2</sup>; furthermore, we should carefully check that the atomic charges do not depend too much on the shape of the cloud, as might be the case with highly polar systems.

The parameters that need to be set for these calculations are:

- radius of the cloud;
- tolerance for the achievement of convergence;
- type of atomic charges to use

To decide the radius of the cloud we performed a series of test calculations to check at which point the molecular dipoles and charges are unaffected by further increasing the radius. For all of our systems we have seen the behaviour visible in Table 5.1 for paracetamol form I, with one molecule per asymmetric unit. The calculated atomic charges shown are converged within  $5.0 \times 10^{-5} e$ , which required 5 iterations for all charge cloud radii tested.

A molecule is considered part of the surrounding cloud if the relative distance between their centre of mass is smaller than the radius  $r_c$  of the cloud. We did not include the case  $r_c = 5 \text{ \AA}$  in the table, as only one molecule would have been part of the cloud.

The number of surrounding molecules to be considered increases dramatically with the radius of the cloud (more than 10 000 atoms for  $r_c = 30 \text{ \AA}$ ), and this allows the shape to be a better approximation of a surrounding sphere.

---

<sup>2</sup> We did not handle ionic systems, for which it might be needed to use full unit cells

		Cutoff radius for the cloud / Å				
		10	15	20	25	30
		Number of surrounding molecules				
		19	72	173	334	581
Initial charges <i>e</i>		Charge variation with the increased radius percent variation				
C	0.305 02	1.57	0.50	0.08	0.20	-0.11
C	-0.195 28	19.00	-3.25	0.72	-0.71	0.40
C	-0.205 80	11.12	-1.95	1.65	-0.22	0.04
C	0.357 00	12.21	-1.33	0.27	-0.27	0.11
C	-0.287 18	-6.70	3.12	-0.33	0.55	-0.25
C	-0.206 26	7.81	2.34	-1.44	0.27	-0.03
C	0.721 15	7.36	-0.41	0.26	0.04	-0.03
C	-0.435 67	10.51	1.64	-0.15	0.34	-0.10
N	-0.630 35	2.15	0.36	-0.09	0.06	-0.05
O	-0.585 14	21.50	-2.30	1.03	-0.32	0.07
O	-0.562 94	25.60	-2.88	0.85	-0.41	0.22
H	0.169 48	28.19	-1.26	-0.03	-0.31	0.20
H	0.150 80	-0.24	2.99	-1.18	0.41	-0.13
H	0.151 20	-1.51	-1.90	1.60	-0.58	0.31
H	0.140 82	13.17	-3.01	2.08	-0.29	0.02
H	0.418 74	25.86	-2.77	0.70	-0.51	0.21
H	0.331 84	17.38	-1.48	0.60	-0.36	0.10
H	0.130 32	-5.35	4.23	-1.98	1.49	-0.60
H	0.107 02	13.47	-0.54	2.78	0.90	-0.47
H	0.125 24	38.01	4.59	-1.11	-0.30	0.45
Molecular dipole Debye		Dipole variation with the increased radius % variation				
4.8314		48.3	28.4	-3.5	1.9	-0.6

Table 5.1. Test calculation of the Breneman charges and molecular dipole, corrected for induction induced by charge cloud of different radius. Each percentile variation is referred to the previous radius of the cloud. The system is paracetamol form I

## 5.2 INDUCTION EFFECT ON PHONON ABSORPTION

### INTENSITIES: THE SANTHA SCRIPT

The calculation of the intensities in a phonon calculation, implemented in RUDOLPH and used in previous studies [60, 163, 180], is very basic: the program uses the knowledge of the phonon eigenvectors to rotate the molecules in the unit cell and calculates the variation of the unit cell

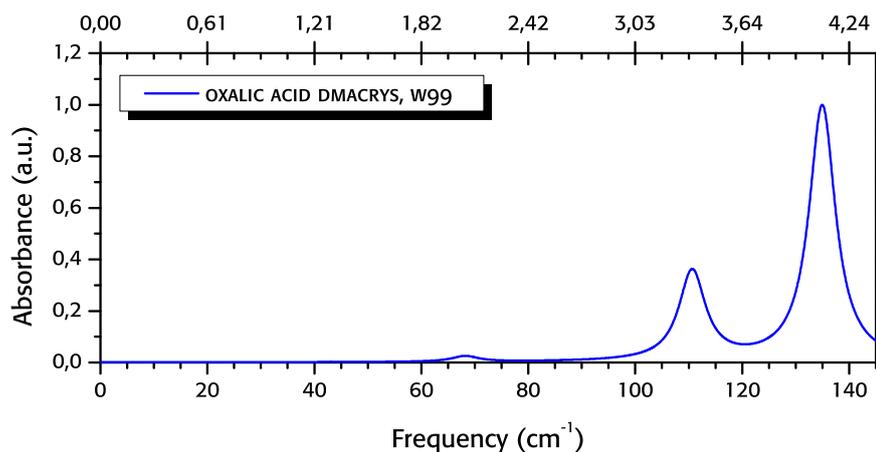
		Convergence parameter / $e$			
		$10^{-1}$	$10^{-2}$	$10^{-3}$	$10^{-4}$
		Number of GAUSSIAN calculations			
		74	146	151	243
Frequency $\text{cm}^{-1}$	Initial absorption	Calculated absorption a.u.			
54.46	0.028	0.103	0.105	0.105	0.105
73.27	0.088	0.219	0.220	0.220	0.220
83.92	1.000	0.976	0.983	0.981	0.979
91.11	0.854	1.000	1.000	1.000	1.000
95.47	0.778	0.917	0.912	0.912	0.912
99.16	0.353	0.337	0.341	0.341	0.340
103.93	0.471	0.557	0.560	0.560	0.560
130.58	0.065	0.285	0.286	0.283	0.284
144.22	0.285	0.075	0.077	0.077	0.076

Table 5.2. Effect of the variation of the convergence parameter of the charges in the cloud on the calculated absorption intensities of the terahertz active phonons. The intensities are normalised to the largest intensity in each calculation. The system is paracetamol form I

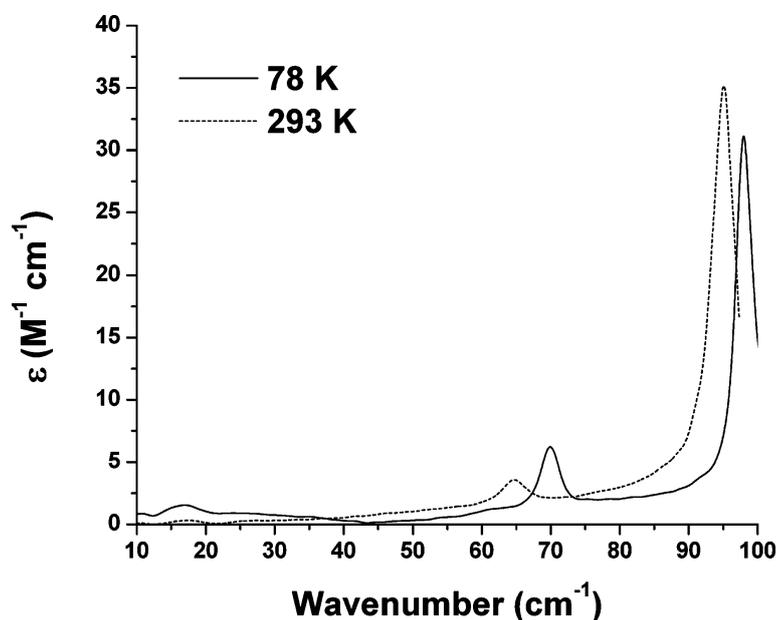
dipole as a sum of all the fixed molecular dipoles, obtained from the DMACRYS output. There is certainly a contribution to the molecular dipoles due to the variation of the electronic distribution along the eigenvector of the vibration, but this is inaccessible without a recalculation of the electronic charges.

The script SANTHA (Scripting Accurate Numbers in TeraHertz Absorption) was written to simulate the effect of the molecular motion along the phonon eigenvectors on the electronic clouds. The principle is similar to what was exposed in the previous section, but on a larger scale: the molecules are displaced along the direction of the phonon eigenvectors; for each molecule, a cloud of point charges is constructed, so that the effects of induction are considered. The script constructs a GAUSSIAN input file according to the information gathered and standard parameters, launches the program and retrieves the relevant information from the output files. The converged dipoles are used to calculate the total unit cell dipole, and the total variation versus the displacement is used to evaluate the absorption intensity.

One of the advantages is the possibility to calculate the absorption intensity for molecules without a permanent electric dipole; with the static dipole method implemented in RUDOLPH the calculated intensity of all



(a) Calculated spectrum using DMACRYS, w99 forcefield and the SANTHA script



(b) Experimental terahertz spectra. From [181]

Figure 5.3. Calculated and measured (78 K and 293 K) terahertz spectra of oxalic acid

phonon modes for such molecules would be zero, while SANTHA has access to the induced dipoles. We have verified this by calculating the phonon spectrum of oxalic acid (CSD code OXALAC04, see Figure 5.2), a centrosymmetric molecule. The results for the calculation are reported in Figure 5.3a, while the experimental spectrum is in Figure 5.3b.

To use the script we have to consider a number of parameters:

- size of the displacement along the eigenvector;
- number of points for the evaluation of the dipole derivative;

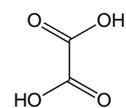


Figure 5.2.

- tolerance for the achievement of convergence;

The size of the displacement is a trade-off between two factors: we need to use molecular displacement small enough so that the numerical derivative is accurate, but at the same time the displacement should be large enough for the change in dipole to be accurate, and we need a few points (3 at least, or 5, with positive and negative displacement) to have a good estimate.

We need to be careful with the choice of the tolerances for the point charge cloud, which must be smaller than the typical changes in atomic charge with phonon displacement if we want to have meaningful results.

The effect of the polarisation model introduced in this chapter will be discussed further in the results section.

### 5.2.1 Eigenvector symmetry analysis

The calculation of absorption intensities with this method is very computationally demanding compared to a normal RUDOLPH run, since hundreds (or even thousands) of GAUSSIAN calculations are needed to obtain this data. We therefore considered a number of symmetry features to reduce the number of operations. As we mentioned before, for the calculation of the cloud charges we only need to calculate the electronic distribution of the symmetry independent molecules. While this is correct for the molecules sitting in the unperturbed equilibrium positions, the total symmetry is broken by the phonon displacements.

Phonon eigenvectors themselves transform as irreducible representations under the symmetry group of the crystal, and this information can be used to reduce the number of calculations.

From a group analysis we could check whether or not a phonon mode belongs to the same irreducible representation of the electric dipole (linear functions). However, the current implementation does not utilise a group analysis of the crystal: rather, we used more intuitive arguments that are more easily implemented in the code. For example, we retrieved all the matrix transformations necessary to transform equivalent molecules into each other; if the translation and rotation components of the eigenvectors<sup>3</sup> of two molecules are equivalent under this transform-

<sup>3</sup> Keeping into account that the rotation is a *pseudovector*, and it transforms in a different way than a vector under inversion

ation, it means that the environment around the molecules is forced to remain the same during a phonon mode displacement, and the electrostatic distribution in these molecules will not vary, halving the number of GAUSSIAN calculations. With the same idea we check whether the variation of the unit cell dipole is forced to be zero by symmetry, or it is an even function with respect to the displacement, guaranteeing that its derivative at equilibrium is zero and avoiding any calculation at all. This is a practical implementation of symmetry selection rules for infrared absorption.

Furthermore, it is usually the case that the dipole variation is an odd function of the molecular displacements, again halving the number of calculations needed.

These three symmetry implementations easily save at least one order of magnitude in the number of GAUSSIAN calculations, and consequently shorten the analysis time. The symmetry subroutines, consisting mainly of small matrix transformations to be performed only once, have a negligible computational cost.

The current implementation is unable to deal with degenerate irreducible representations, excluding some of the higher symmetry space groups. We plan to eventually extend the program to be able to deal with these situations; one simple way would be to directly read the symmetry information that is produced by NEIGHCRYST.

### 5.2.2 Subroutine implementation

The PERL script relies on the output from RUDOLPH. The information needed consists of the molecular orthogonal coordinates, the lattice vectors (for the generation of the point charge cloud), and the phonon eigenvectors and eigenvalues, already expressed in the global orthogonal reference system.

Furthermore, the subroutine asks for several parameters: the cloud size, the molecular displacement size and number, number of equivalent molecules in the system and the level of theory for the GAUSSIAN molecular electron density calculation.

According to the numbering in the diagram of Figure 5.4 we can distinguish the following phases of the calculation:

1. In the symmetry setup section, the transformation matrices are

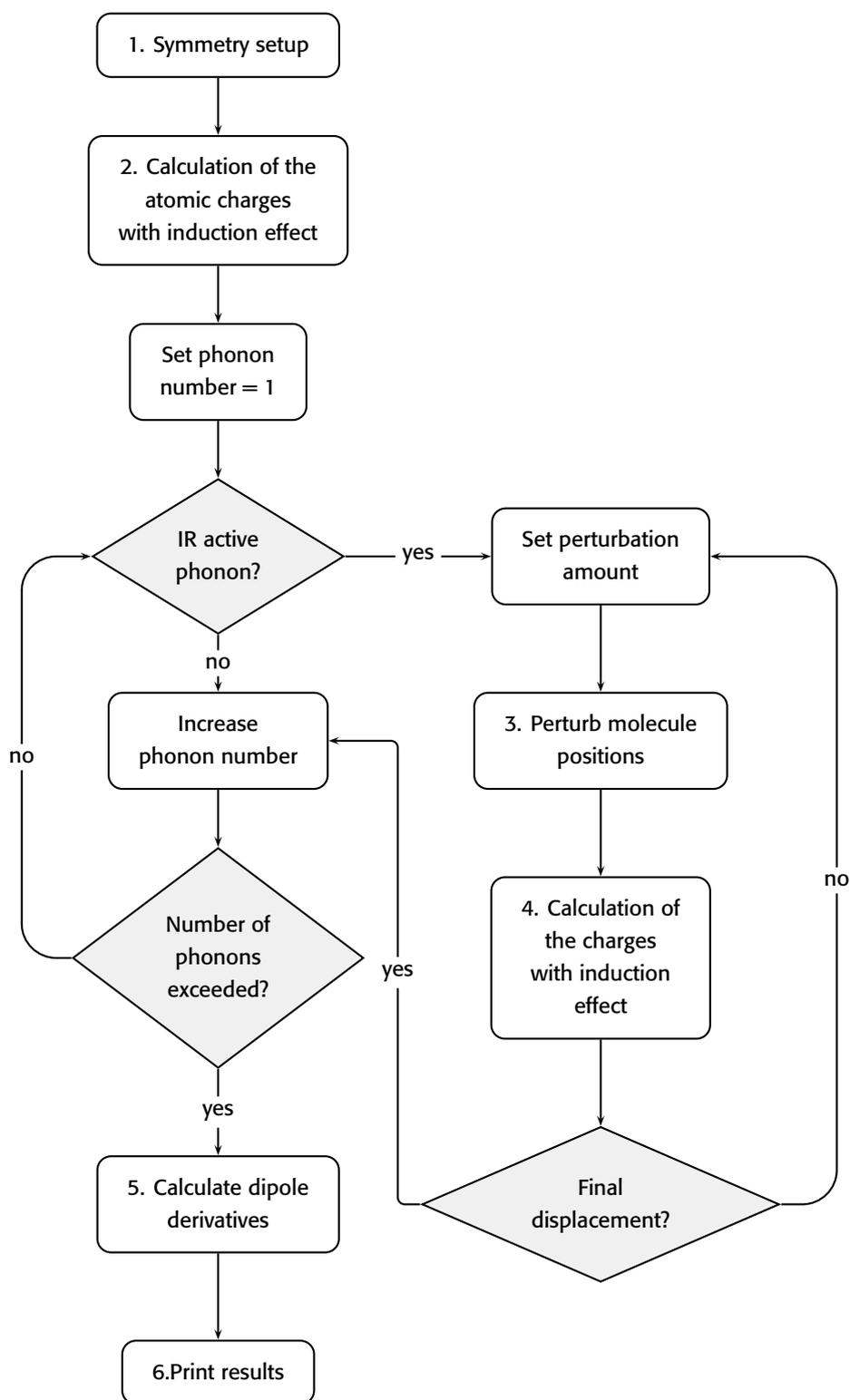


Figure 5.4. Iterative calculation of the terahertz absorption intensities using the SANTHA script

constructed by comparing the molecular coordinates; an array of equivalent molecules under each eigenvector perturbation is created;

2. The induction-affected charges are calculated according to the subroutine in the diagram of Figure 5.1;
3. The position of the atoms in the unit cell and in the surrounding cloud are perturbed according to the eigenvectors; for consistency, the molecules that are included in the cloud are kept the same during all the perturbations, regardless of the fact that new molecules might be shifted in or out of the cutoff radius;
4. The induction-affected charges are again calculated according to the subroutine in the diagram of Figure 5.1, but taking into account the reduced symmetry;
5. Absorption intensities are calculated as a numerical derivative of the unit cell dipole, and normalised to set the largest intensity to unity;
6. Output files are produced, containing frequency, intensity and preferential direction for the dipole change.

### 5.2.3 Raman intensity calculations

The *SANTHA* subroutine (although not tested specifically for this) is also capable of calculating approximate Raman intensities, often a complementary tool to understand the vibrational properties of a material.

The Raman effect occurs when light interacts with the electron cloud and the bonds of a molecule in a crystal. A photon excites the molecule from the ground state to a virtual energy state, and when it relaxes it emits a photon, returning to a different vibrational state. The difference in energy between the original state and this excited state leads to a shift in the emitted photon's frequency.

Since this is a different phenomenon than infrared absorption, there are also different selection rules for a phonon mode to be active; the discriminating parameter is the total polarisability of the unit cell, which has to vary for a phonon mode to be active. The polarisability transforms as irreducible crystal group representation spanned by the quad-

ratio function (not linear as the dipole moments, see previous section). Consequently, for centrosymmetric crystal there is an exclusion rule: no phonon mode can be both IR and Raman active. We expanded our symmetry script to check for Raman active modes as well, taking into account the full polarisability tensor for all the symmetry transformation.

The calculation of molecular polarisabilities is also a standard feature in GAUSSIAN calculations, although considerably slower than a normal energy calculation. Its variation, necessary for the Raman intensity calculation, is calculated using the same subroutine we used for the unit cell dipole derivative.

## SYSTEMS STUDIED

---

**F**OR OUR STUDY we selected a few molecules whose crystal structures are available in the CSD. We selected our systems according to the requirements imposed by our programs of choice and from the objectives of our study. In particular we were looking for molecules of pharmaceutical interest, with the additional requirements:

**SMALL MOLECULES**, up to 25-30 atoms, and a total number of atoms in the unit cell smaller than 200. They should contain only hydrogen, carbon, nitrogen, oxygen or halogen atoms, because of the limitation on the parameterised potential we use in DMACRYS; the limit on the number of atoms is an important parameter in CASTEP, to keep the computation time reasonable;

**RIGID MOLECULES**, for a correct behaviour of DMACRYS. We tried to select systems with increasing degree of intramolecular flexibility to test the limits of our rigid-molecule lattice dynamics in DMACRYS. This is not an issue for the CASTEP description.

One of the main interests in the research has been how polymorphism affect the terahertz spectra, and how we can appreciate the differences from the experimental and computational point of view. We therefore looked for molecules that crystallise in more than one form.

From the point of view of our research, we selected molecular systems according to the following goals, expanded in the relative results chapters:

**CHAPTER 8** deals with the effect on the spectra due to the flexibility of the molecules in the crystal, and tries to test how much we can stretch the DMACRYS program to deal correctly with it. For this study we needed a set of molecules with different degrees of rigidity;

CHAPTER 9 deals with the effect water molecules have on the measured and calculated spectra, looking for significant features that can help to identify hydrate crystal structures. For this study we needed molecular systems that crystallise with and without water within the unit cell, and where the crystal structures of both the neat and hydrate crystals are known;

CHAPTER 11 deals with the effect that small changes in the molecular structure have on the measured and calculated spectrum. For this study we needed a series of similar molecules, which are able to crystallise in similar unit cells.

There are of course several possible choices: a 2006 study of the CSD [182] found about 2000 polymorphic compounds, while a different study by the same author [183] found more than 8000 molecules where both an unsolvated and a solvate crystal structure are known, with half of the solvates being hydrates. The molecular systems we have chosen are described in more detail in the following sections.

## 6.1 PARACETAMOL

Paracetamol – IUPAC *N*-(4-hydroxyphenyl) acetamide, see Figure 6.1 – is a very common analgesic and antipyretic drug. There are six different crystal forms listed in the CSD, with different stabilities.

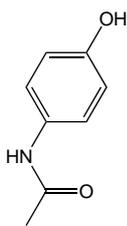


Figure 6.1.

There are many (32!) different determinations of the anhydrous forms, unified under the code HXACAN, under different pressures and temperature conditions. There are three known polymorphs: the stable monoclinic form I (HXACAN06) and the orthorhombic form II (HXACAN08) have been known for a long time, while the crystal structure of the metastable orthorhombic form III (HXACAN29) has only recently been determined [184].

There are three known hydrate crystal structures: a monohydrate form (HUMJEE), a high pressure dihydrate form (WAFNAT) and a trihydrate form (XOMWOL). None of these hydrates is stable at standard conditions: the monohydrate is reported to lose water after 5 minutes exposure to air [185], while the trihydrate is even more elusive [186].

We did not calculate the spectrum for the dihydrate, a high pressure only form, and we will not refer to it in the following chapters.

### 6.1.1 Anhydrous forms

The polymorphism of paracetamol is interesting from the pharmaceutical point of view because form II has better tableting properties compared to form I [187], without affecting the solubility rate [17]. The different tableting properties are due to the different type of packing, with form I lacking the flat, hydrogen bonded sheets that characterise form II (see Figure 6.2).

From the topological point of view the hydrogen bonding networks in the two forms of paracetamol are very similar, with the formation of interconnected 4-molecule rings (Figure 6.3) across the molecular sheets, involving all the oxygen and the nitrogen atoms in the molecule. A common way of characterising the ring is to consider the behaviour of neighbouring molecules in the ring, whether it is an hydrogen bond donor (*d*) or acceptor (*a*). The ring can be classified *ddaa* for form I, *dddd* for form II.

Form III has striking similarity with form II, with the same layered structure and almost superimposable unit cell parameters. The main difference is in the arrangement of the molecular sheets: in form II two subsequent layers are superimposable by symmetry [184], while this is not true for form III, in which  $Z' = 2$  (Figure 6.4).

### 6.1.2 Hydrate forms

The two hydrate forms show very different structures, each characterised by its own hydrogen bond network between water and paracetamol molecules.

Each water molecule in the monohydrate acts as a bridge connecting

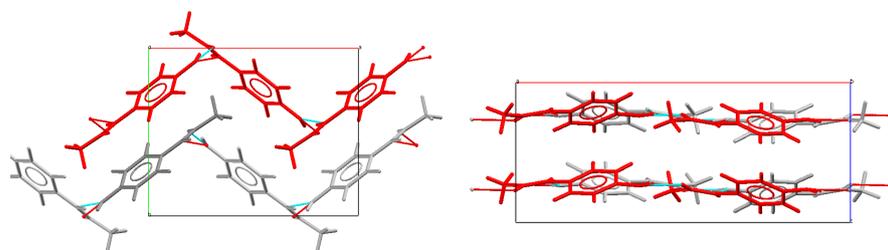


Figure 6.2. Comparison of the crystal packing of paracetamol form I (left) and form II (right) as viewed down the *b* axis. Molecules closer to the observer are coloured in red

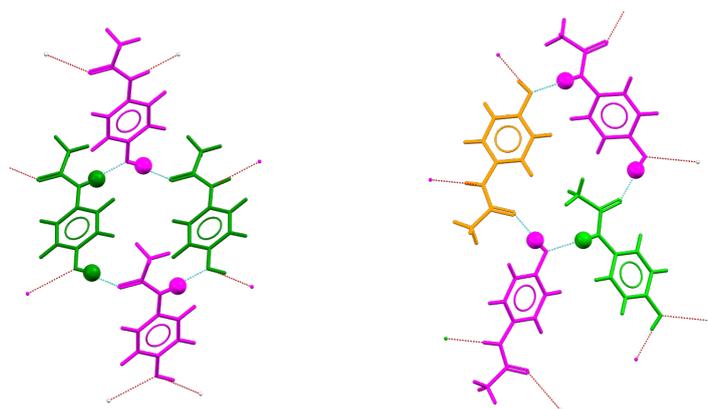


Figure 6.3. Hydrogen bond network in paracetamol form I (left) and form II (right). Molecules with the same symmetry type are in the same colour. Hydrogen bonds between molecules are dotted lines, coloured in azure if both molecules are shown. The donated hydrogen atoms are highlighted as spheres

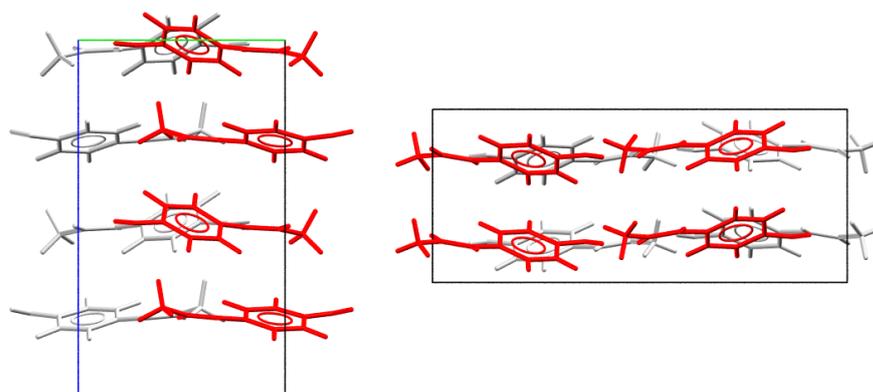


Figure 6.4. Comparison of the crystal packing of paracetamol form II (left) and form III (right). The molecules closer to the observer are coloured in red

three paracetamol molecules, and it is embedded as in a pocket. In the trihydrate, instead, each water molecule interacts with at least another water molecule, forming infinite interconnected channels: in this case paracetamol could be described as residing in a pocket created by the water molecules. This connectivity is depicted in Figure 6.5, showing in different colours symmetry inequivalent water molecules: the hydrogen bonding network between water molecules is dashed in red.

Structurally, the paracetamol molecules in the trihydrate are organised in parallel sheets, while they keep a zigzag arrangement in the monohydrate (similar to form I). More on the structures of paracetamol hydrates is discussed in the results section, chapter 9.2.2 and 9.2.3.

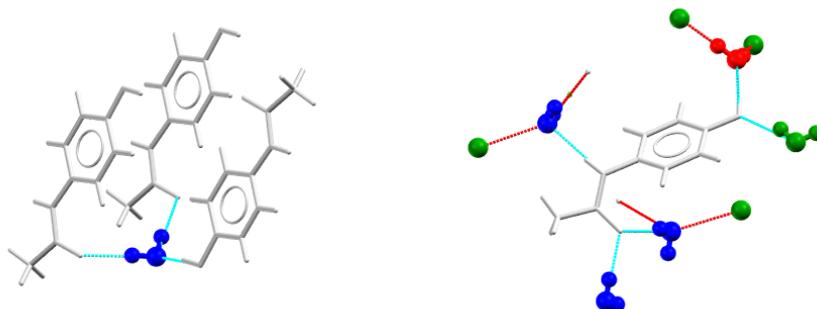


Figure 6.5. First neighbours of a water molecule paracetamol monohydrate monohydrate (HUMJEE, left) and first neighbours of a paracetamol molecule in the trihydrate (XOMWOL, right)

## 6.2 THEOPHYLLINE

Theophylline – IUPAC 1,3-dimethyl-7H-purine-2,6-dione, see Figure 6.6 – is a drug of the xanthine family, structurally similar to the caffeine molecule; it is highly effective as a bronchodilator, and it is used in the treatment of acute respiratory diseases such as asthma; its use is not widespread [188] because it has a narrow therapeutic range ( $10 \text{ mg l}^{-1}$  to  $20 \text{ mg l}^{-1}$ ) and side effects may appear with a blood concentration higher than  $15 \text{ mg l}^{-1}$ . The polymorphism of theophylline, which affects its bioavailability [189], hinders its use as an oral medication [190].

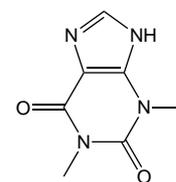


Figure 6.6.

Five neat forms of theophylline have been reported in the literature<sup>1</sup> and characterised in the literature [192, 193], but the crystal structures have been determined only for one of these (form II); a stable monohydrate form, is known as well, which interconverts easily with form II if heated over 340 K [192].

There are two entries in the CSD: an anhydrous form (BAPLOT01) and a monohydrate form (THEOPH01).

The anhydrous form is characterised by a chain of hydrogen bonds between the molecules [194], while the hydrate structure shows dimer formation between theophylline molecules, together with an infinite chain of water molecules in the *a*-axis direction [195].

On the left, Figure 6.7 shows the linear chain in the anhydrous form, together with the unit cell; on the right, the different hydrogen bonded dimers of theophylline, coloured the same. The water chains run per-

<sup>1</sup> A sixth form has been recently discovered as well [191]

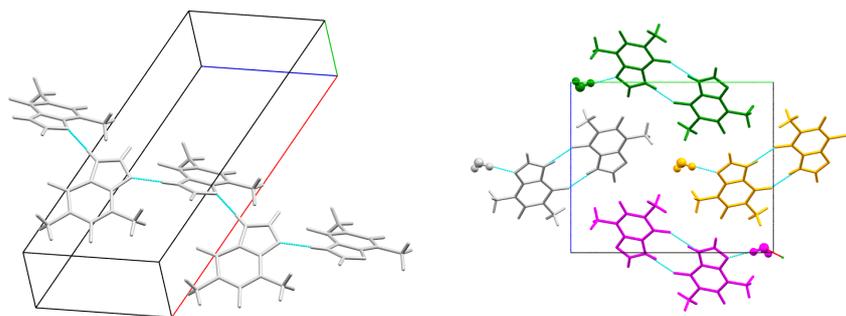


Figure 6.7. Crystal packing of anhydrous theophylline (BAPLOT01, left) and monohydrate (THEOPH01, right). In the monohydrate hydrogen bonded pairs are drawn with the same colour

pendicularly to the plane of the figure, each water molecule connected via hydrogen bond to each other.

In the theophylline hydrate structure reported in the CSD there is disorder for both a water hydrogen (50% occupancy, hydrogen bonding to different theophylline molecules in the unit cell) and the methyl group (rotation of  $60^\circ$  with 64/36% occupancy). Indeed, the disorder of water is fundamental to correctly interpret the hydrogen bond and interaction between water molecules. This will be discussed in more detail in chapter 7.3.

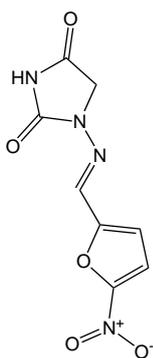


Figure 6.8.

### 6.3 NITROFURANTOIN

Nitrofurantoin – IUPAC 1-[(5-nitro-2-furyl) methylideneamino] imidazolidine-2,4-dione, see Figure 6.8 – is an antibiotic, sometimes used to cure infection in the urinary tract; bioavailability problems related to the existence of polymorphism were reported [196]: tablets of the stable form  $\beta$  can partially convert (in humid atmosphere, over a period of 4-8 weeks) to either the forms  $\alpha$  and hydrate I, both of which are characterised by a lower solubility rate.

The CSD lists 4 forms: the stable anhydrous LABJON (form  $\beta$ ) and LABJON01 (form  $\alpha$ ), and the hydrate forms HAXBUD (form I) and HAXBUD01 (form II).

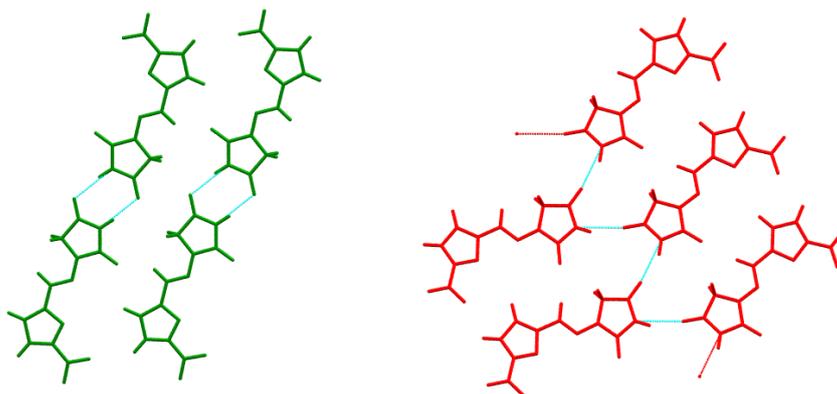


Figure 6.9. Connectivity of nitrofurantoin molecules in forms  $\alpha$  (right) and  $\beta$  (left)

### 6.3.1 Anhydrous forms

Both anhydrous forms are characterised by molecular stacking layers, which in form  $\beta$  are parallel to the  $20\bar{4}$  plane, while in form  $\alpha$  they are orthogonal to the  $b$  axis. There is a similar separation between the stacking in both structures.

The arrangement of molecules within a sheet is different: in form  $\alpha$  LABJON the hydrogen bonds form an 8-membered ring, a dimer structure with a double strong  $\text{NH} \cdots \text{O}$  hydrogen bond. In form  $\beta$  the same  $\text{NH} \cdots \text{O}$  bond is present, but with an infinite chain of molecules within a molecular plane (Figure 6.9).

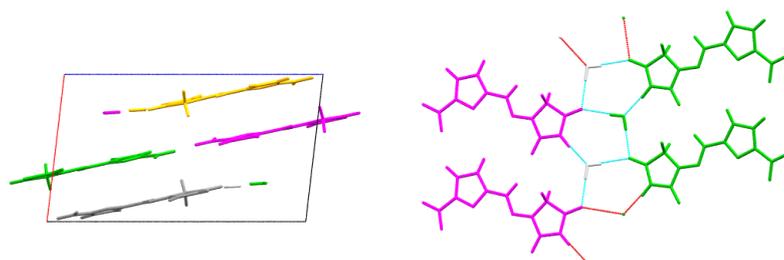


Figure 6.10. Views of nitrofurantoin hydrate form I (HAXBUD01) to show the planar arrangement along the  $b$  axis (left) and the molecular connectivity in each sheet (right). Different colours are used for symmetry inequivalent molecules in each sheet

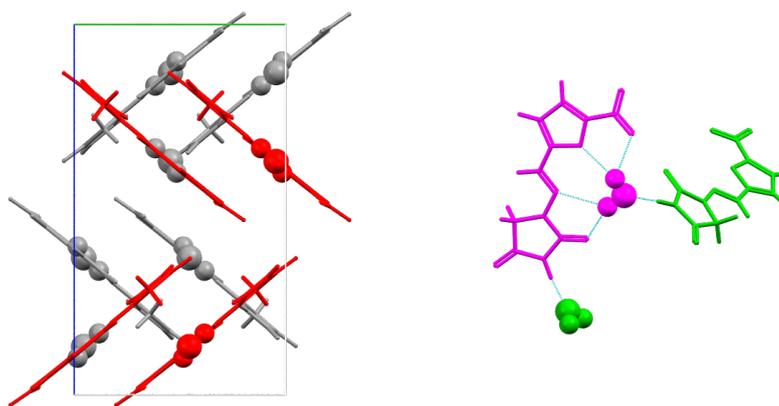


Figure 6.11. Views of nitrofurantoin hydrate form II (HAXBUD) along the  $b$  axis (left, red molecules closer to the viewer) and molecular diagram to show the connectivity (right, different colours are used for symmetry inequivalent molecules). Water atoms are drawn as spheres

### 6.3.2 Hydrate forms

The two hydrate forms are characterised by a completely different structure: in polymorph I there is still organisation in parallel sheets, with an internal network of hydrogen bonds between nitrofurantoin and water molecules (see Figure 6.10). In polymorph II the contact between nitrofurantoin molecules is mediated by a molecule of water, but this time the effect of the interaction is that each of them is tilted by an angle near to  $90^\circ$  with respect to one another.

The structure of the hydrate and its implications for the phonon eigenvectors is discussed more in depth for the two forms in chapter 9.

## 6.4 CARBAMAZEPINE

Carbamazepine – IUPAC 5H-dibenzo[b,f]azepine-5-carboxamide, see Figure 6.12 – is an anticonvulsant and mood-stabilizing drug, and it is one of the most used drugs for the treatment of epilepsy and bipolar disorder [197–199]. It was discovered in 1953 and it has been marketed and used in several countries since the early seventies for its anti-epileptic properties.



Figure 6.12.

From the structural point of view, carbamazepine exhibits wide polymorphism, with as many as five different anhydrous forms and a dihydrate form; for this reason it is commonly used as a prototypical sys-

tem for polymorphism studies.

Anhydrous carbamazepine is identified in the database with the code CBMZPN, with 11 different entries. Each form is identified with a Roman numeral (despite different and sometimes contradictory nomenclature sometimes used in the literature, see for example [200]).

Forms I–IV exhibit the same type of dimer structure, with a strong N–H ··· O double bridge of hydrogen bonds between molecules: the difference is thus mainly due to the different arrangement of this dimer in the structure, together with the length  $d$  of the bridge and the angle  $\alpha$  (see Figure 6.13) which are not found to vary too much between different forms (see Table 6.1). Form V [201] was cleverly synthesised in 2010 by templating the growth of carbamazepine from the vapour phase onto the surface of a crystal of a similar molecule (dihydrocarbamazepine) in order to obtain a predicted, but not previously observed, hydrogen bond chains (Figure 6.14)

The P-monoclinic form III (CBMZPN02) is the only suitable form for drug formulation, because it is thermodynamically stable at ambient temperature and has the highest bioavailability [197]; form I (CBMZPN11) is the stable form at high temperature, form II (CBMZPN03) is trigonal and characterised by partial solvent inclusion [202], and the unstable form IV (CBMZPN12) is C-monoclinic.

The dihydrate has 4 different crystal structure determinations, under the FEFNOT code name. Two different crystal polymorphs are reported, allegedly depending on the starting anhydrous form used in the preparation: however a more accurate determination of the crystal structures [200] demonstrated the two structures to be equivalent. The structure of the hydrate and its implications for the phonon eigenvectors is discussed more in depth in section 9.2.6.

## 6.5 BENZOIC ACID

Benzoic acid – see Figure 6.15 – is the simplest aromatic carboxylic acid, used in medicine for its antifungal properties. It is also used as a food preservative, with the code number E210, and as a salt with further denominations E211 – E214.

The CSD code for benzoic acid is BENZAC, with 8 different determina-

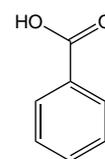


Figure 6.15.

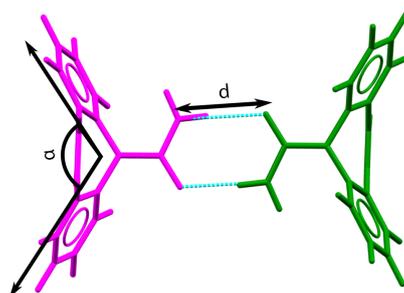


Figure 6.13. Distance  $d$  between nitrogen and oxygen in the hydrogen bond and angle  $\alpha$  between the phenyl rings

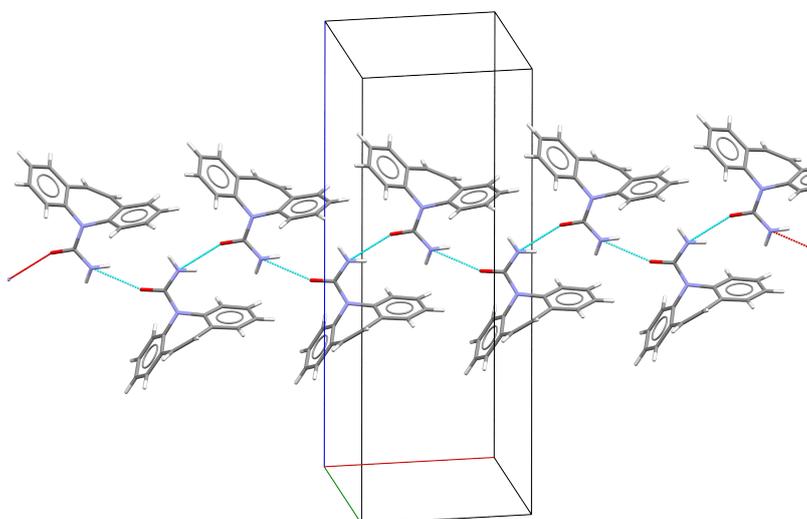


Figure 6.14. Hydrogen bond chain in carbamazepine form V, hydrogen bonds are dotted lines. Unit cell axes are shown

Form	CSD refcode	$Z'$	$d$ (Å)	$\alpha$ (°)
Form I	CBZPN11	4	2.831	116.3
			2.884	116.3
			2.894	116.3
			2.892	116.4
Form II	CBZPNO3	1	2.890	116.1
Form III	CBZPNO2	1	2.928	117.2
Form IV	CBZPN12	1	2.874	117.5
Dihydrate	FEFNTO2	1	2.867	116.7

Table 6.1. Variability between the different forms of carbamazepine for the hydrogen bond distance  $d$  and the angle  $\alpha$  between phenyl rings, as defined in Figure 6.13

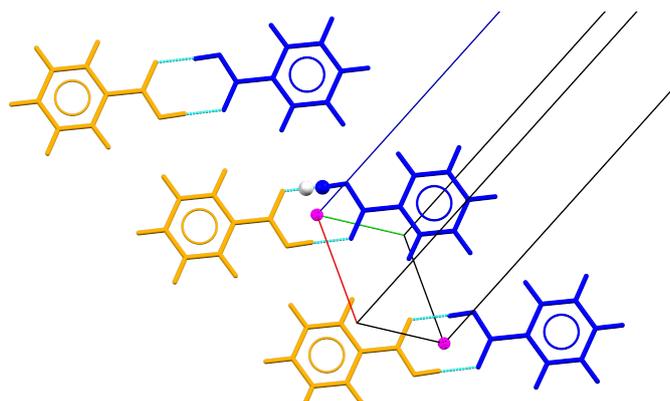


Figure 6.16. Crystal structure of benzoic acid. Molecules of the same symmetry equivalence are shown with the same colours. The alternate position for the acid hydrogen is drawn as a white sphere; the symmetry inversion elements for two of the dimers (on the corners of the unit cell) are coloured in magenta

tions, and no reported polymorphism or hydrate structures. The carboxylic group mediates a strong hydrogen bond interaction  $\text{OH} \cdots \text{O}$ , so that the structure comprises of dimers of benzoic acid molecules.

The position of the acidic hydrogen on an isolated molecule does not affect its energy, due to its symmetry; in the condensed state, however, each dimer sits on one of the crystal centres of inversion (Figure 6.16) or there are two possible configurations, one slightly energetically favoured over the other. As a consequence the crystal structure is disordered.

A 1996 article [203] analysed the structural changes with the temperature varying from 20 K to 175 K, and the resulting temperature dependent crystal structures are recorded in the database with numeric codes from BENZAC07 to BENZAC11. The ratio between the hydrogen positions for the two possible sites changes from 0.87:0.14 (20 K) to 0.62:0.38 (175 K).

Such a subtle difference in the structure as the slight change in the position of a hydrogen was found to have a big effect in the measurement and modelling of the terahertz spectrum [180].

## 6.6 SUBSTITUTED ACETANILIDES

We selected a family of molecules where only one functional group is changed: the IUPAC name of this compounds would generally be N-(4-X-phenyl) acetamide, – see Figure 6.17 – where X is the substituted

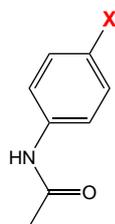


Figure 6.17.

group.

These molecules, as well as exhibiting small molecular changes, crystallise with a very similar packing. Therefore they represent a very good system to investigate the effect of small molecular changes on their terahertz spectra.

The molecules reported here were chosen from a larger set of substituted acetamide derivatives, and the experimental samples were provided by Suzanne Huth and Micheal Hursthouse (University of Southampton).

It can be noted from Table 6.3 that we can cluster three of the structures together for similarity, CLACTN and BRAC TN (almost superimposable) and ACTOLD02.

As substitution *X* group we selected halogen atoms, the hydroxyl group (OH) and methyl group (CH<sub>3</sub>).

We already extensively described the hydroxyl-derivative, paracetamol (section 6.1, Figure 6.1). The other systems are obviously going to have different reference codes in the database, as they are different molecules. The nomenclature is reported in Table 6.2. Note that there is no crystal for the iodo-derivative in the CSD, as the crystal structure determination we used comes from a private communication: we will call it IIACTN for brevity throughout the work.

### 6.6.1 Isostructural crystals

The crystal packing of these molecules is very similar – with the notable exception of paracetamol – and it represents a very good model system to analyse the effect of slight changes of structure on the terahertz spectrum. The crystal packings are equivalent to the one reported in Figure

Functional group	CSD refcode	Polymorphic?
Hydroxy-	HXACAN	✓
Methyl-	ACTOLD	✓
Fluoro-	SAZLEL	✗
Chloro-	CLACTN	✗
Bromo-	BRAC TN	✗
Iodo-	(IIACTN)	✗

Table 6.2. Crystal structure code of the acetanilides derivatives used in our work. The iodo-derivative is not present in the database

-X	CSD refcode	Space group	Unit cell parameters				$\delta$ (°)
			a (Å)	b (Å)	c (Å)	$\beta$ (°)	
CH <sub>3</sub>	ACTOLD01	P2 <sub>1</sub> /c	11.67	9.58	7.55	106.1	—
	ACTOLD02	Pna2 <sub>1</sub>	9.89	12.95	6.54	90	116.51
F	SAZLEL	Cc	4.73	17.06	9.63	92.87	99.22
Cl	CLACTN02	Pna2 <sub>1</sub>	9.71	12.75	6.53	90	115.19
Br	BRACTN	Pna2 <sub>1</sub>	9.70	13.02	6.67	90	114.34
I	(IIACTN)	P2 <sub>1</sub> /c	9.58	10.30	9.43	109.28	98.29

Table 6.3. Structural information on the acetanilide derivatives, substituted group X indicated. The unit cell angles  $\alpha = \gamma = 90^\circ$  for all the structures. ACTOLD01 is not isostructural with the other forms

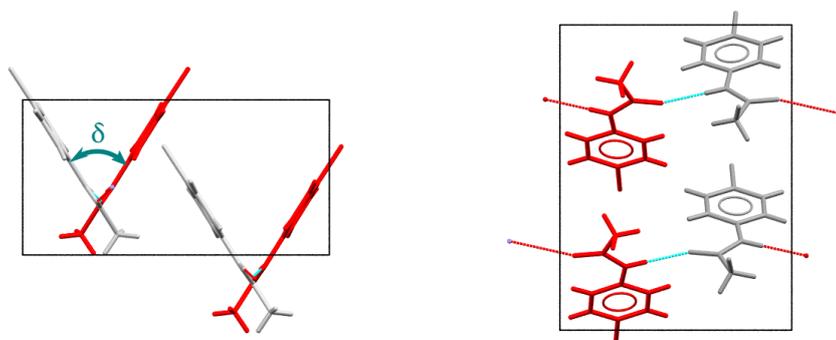


Figure 6.18. Molecular packing of the form CLACTN02 through the *b* (left, with red molecules closer to the observer) and *a* (right) axes. On the right, the hydrogen bonds are shown

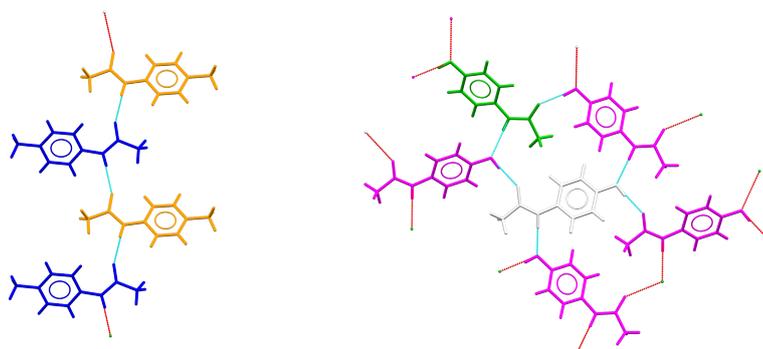


Figure 6.19. Comparison of the in-plane connectivity of N-(4-methyl-phenyl)acetamide (left) and paracetamol form II (right)

6.18 for CLACTNO2 as an example. The main variability is represented by the unit cell parameters and the apparent “angle”  $\delta$  between two molecules (visible on the left); more precisely – since the molecules are shifted with respect to one another – it is the tilting angle between the plane containing the phenyl rings of two neighbouring molecules.

The hydrogen bond network is different from the one in paracetamol, since one of the groups involved (OH) is missing: the molecules are linked by an infinite  $\text{C}=\text{O} \cdots \text{H}-\text{N}-\text{C}$  chain.

The structures SAZLEL and IIACTN have a similar angle  $\delta$ , but the differences in the unit cell parameters and space group make their superimposition difficult, except for the similar molecular stacking along the  $c$  axis.

The polymorphic form of the methyl derivative ACTOLDO1 has instead a planar structure more similar to that of paracetamol form II, but with a different hydrogen bond network, due to the absence of the -OH group: the molecules are therefore arranged in hydrogen bonded chains instead of an interconnected plane.

## EXPERIMENTAL AND CALCULATED TERAHERTZ SPECTRA

---

**I**N THIS CHAPTER we present the comparison between the experimental spectra to the calculated DMACRYS and CASTEP spectra. We discuss the influence of the forcefield used in DMACRYS and the effect of the dispersion correction in CASTEP whenever it is possible to have a comparison (paracetamol form I and nitrofurantoin  $\beta$ ).

The effect and the importance of considering the disorder in the crystal structure is analysed in the case of benzoic acid and theophylline hydrate.

We report the experimental uncertainties for all the measurements we performed; for clarity we only indicate them once. Some of the reported spectra (theophylline monohydrate and carbamazepine form I and III) are older determinations acquired by Axel Zeitler.

The spectra are typically plotted in the range that is experimentally accessible (generally  $4\text{ cm}^{-1}$  to  $115\text{ cm}^{-1}$ ), unless we want to highlight relevant calculated peaks at higher frequency.

For all the experimental spectra we selected the clearest determination, which is the lowest temperature one except for the helium-cooled systems. The temperature dependent spectra are reported in appendix A.

### 7.1 BENZOIC ACID

Part of the calculations<sup>1</sup> and all of the experimental measurements of benzoic acid in this section were performed by Ruoy Li and they are published [180].

The CSD experimental structure we used as a starting point for our cal-

---

1 Non-dispersion corrected CASTEP, PBE functional

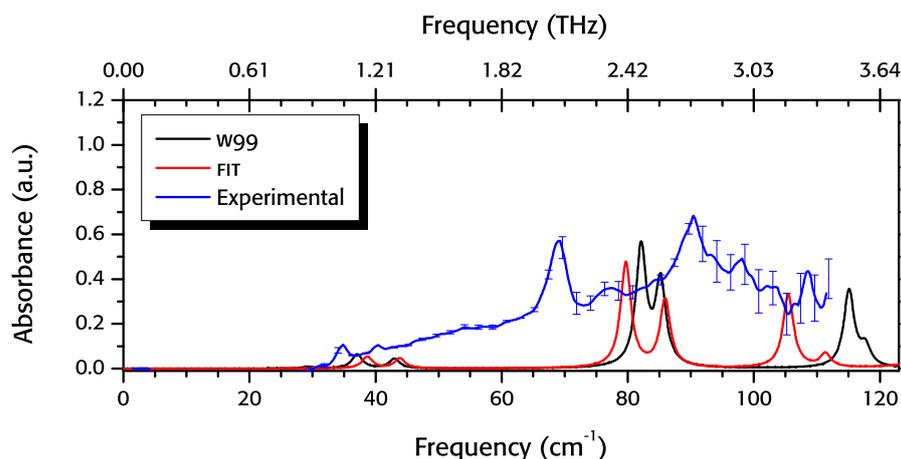


Figure 7.1. Experimental spectrum of benzoic acid collected at 110 K, compared with DMACRYS calculations with the w99 and FIT forcefields

calculations is BENZAC09, determined by neutron diffraction at 100 K. The structure is disordered over one hydrogen atom, with a 0.67:0.33 ratio; we retained only the lower energy dimer configuration in the calculations reported in sections 7.1.1 and 7.1.2 and we refer to it as configuration A (visually it is the hydrogen atom coloured in blue in Figure 6.16).

The low temperature spectrum (Figure 7.1) shows two weak absorption peaks at  $35\text{ cm}^{-1}$  and  $40\text{ cm}^{-1}$ , and the two main signals at  $69\text{ cm}^{-1}$  and  $91\text{ cm}^{-1}$ . The error bars at higher frequency make it impossible to unambiguously identify any of the other absorptions. A measurement by Takahashi at 4 K [204] shows that there is a peak at  $78\text{ cm}^{-1}$  and four absorptions in the range  $95\text{ cm}^{-1}$  to  $130\text{ cm}^{-1}$ .

### 7.1.1 DMACRYS calculations

The geometry optimisation did not have a significant effect on the unit cell parameters (see Table 7.1): the  $a$  vector shrank slightly while the other two dimensions expanded to some extent.

The two calculations (Figure 7.1) show seven active phonon absorptions, with the lower two matching very well with the experimental values. The higher calculated absorption frequencies at  $80\text{ cm}^{-1}$  to  $85\text{ cm}^{-1}$  can be related to the experimental  $70\text{ cm}^{-1}$  to  $78\text{ cm}^{-1}$  doublet; the higher FIT absorption peaks have a similar shape with respect to the  $90\text{ cm}^{-1}$  to  $110\text{ cm}^{-1}$  experimental structure compared to the w99 features, but

	Crystal parameters variation				
	Exp. (110 K)	w99	Change/%	FIT	Change/%
$a / \text{\AA}$	5.43	5.33	-1.86	5.37	-1.13
$b / \text{\AA}$	5.04	5.13	1.74	5.19	2.90
$c / \text{\AA}$	21.75	21.86	0.50	22.09	1.55
$\beta / ^\circ$	98.04	100.56	2.57	100.37	2.37
Vol. / $\text{\AA}^3$	589.4	587.3	-0.4	604.9	2.6
$F$ value		26.85		28.17	

Table 7.1. Variation of the lattice parameters for the DMACRYS geometry optimisation of benzoic acid. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$

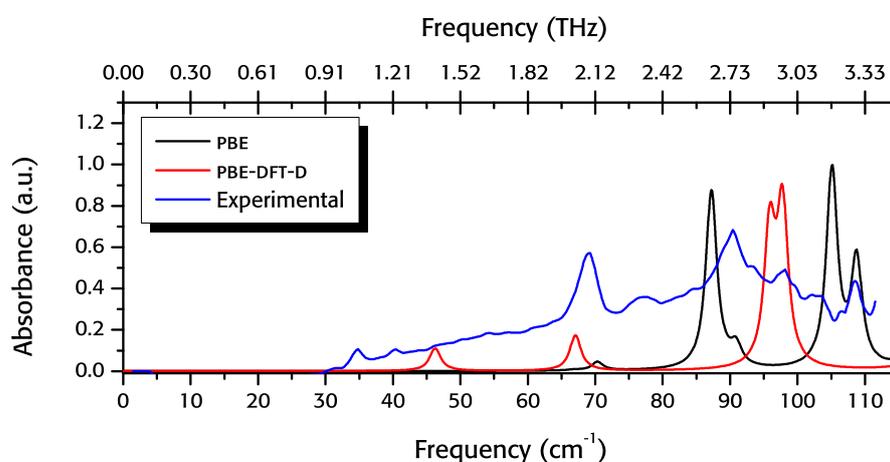


Figure 7.2. Experimental spectrum of benzoic acid collected at 110 K, compared with CASTEP calculations with PBE and dispersion corrected PBE functional

there is not a good agreement in frequency.

All of the calculated peaks are found at higher frequencies than the assigned experimental determinations.

### 7.1.2 CASTEP calculations

Relaxation of the unit cell was allowed only in the dispersion-corrected PBE calculation, with a decrease of 6.98 % in volume from the experimental value; there is a contraction of all the unit cell parameters (Table 7.2), while for the DMACRYS calculations such contraction was only observed for the  $a$  vector.

The calculated absorption frequencies are in good agreement with the

	Experimental (110 K)	Crystal parameters variation	
		PBE-D	Change /%
$a / \text{\AA}$	5.43	5.24	-3.50
$b / \text{\AA}$	5.04	4.94	-1.98
$c / \text{\AA}$	21.75	21.25	-2.29
$\beta / ^\circ$	98.04	99.37	1.35
Volume / $\text{\AA}^3$	589.4	542.19	-6.98

Table 7.2. Change in the lattice parameters for the CASTEP geometry optimisation of benzoic acid, dispersion corrected PBE. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$

experimental values, especially at the higher frequencies, but the peak at  $78 \text{ cm}^{-1}$  is missing from both simulations.

The two calculations do not agree well with the low frequency experimental doublet: this feature is present in the PBE calculation at  $50.7 \text{ cm}^{-1}$  and  $66.9 \text{ cm}^{-1}$ , but the intensity is too low to be seen in Figure 7.2. In the PBE-D calculation the doublet is compressed to two almost superimposable absorptions ( $42.2 \text{ cm}^{-1}$  and  $42.6 \text{ cm}^{-1}$ ).

### 7.1.3 Effect of the disorder

As we mentioned, about 30 % of the acid hydrogen atom positions are disordered at 100 K. We performed DMACRYS and CASTEP calculations with all the dimers swapped to the higher energy conformation (configuration B), and we have seen significant changes in the final spectrum.

The effect of the disorder is reported by Li *et al.* [180] by considering a series of supercells containing 12 dimers<sup>2</sup>. For every supercell, four randomly chosen dimers were chosen to be in the B configuration, thus keeping the ratio fixed at 2 : 1 for configuration A : B, and we analysed the resulting spectra. The results show an improved agreement with the experimental data, and demonstrate that even such a subtle variation within the unit cell has a measurable effect (Figure 7.4).

The DMACRYS spectra resulting from these calculations are reported in Figure 7.3; we can see how every disordered spectrum is different from each other, and the detailed structure of the peaks (for instance, the separation of the two peaks at 2.5 THz) is different from the weighted aver-

<sup>2</sup> In this way “real” disorder is not considered: there is a bigger unit cell with a reduced symmetry, but it is still a periodic calculation

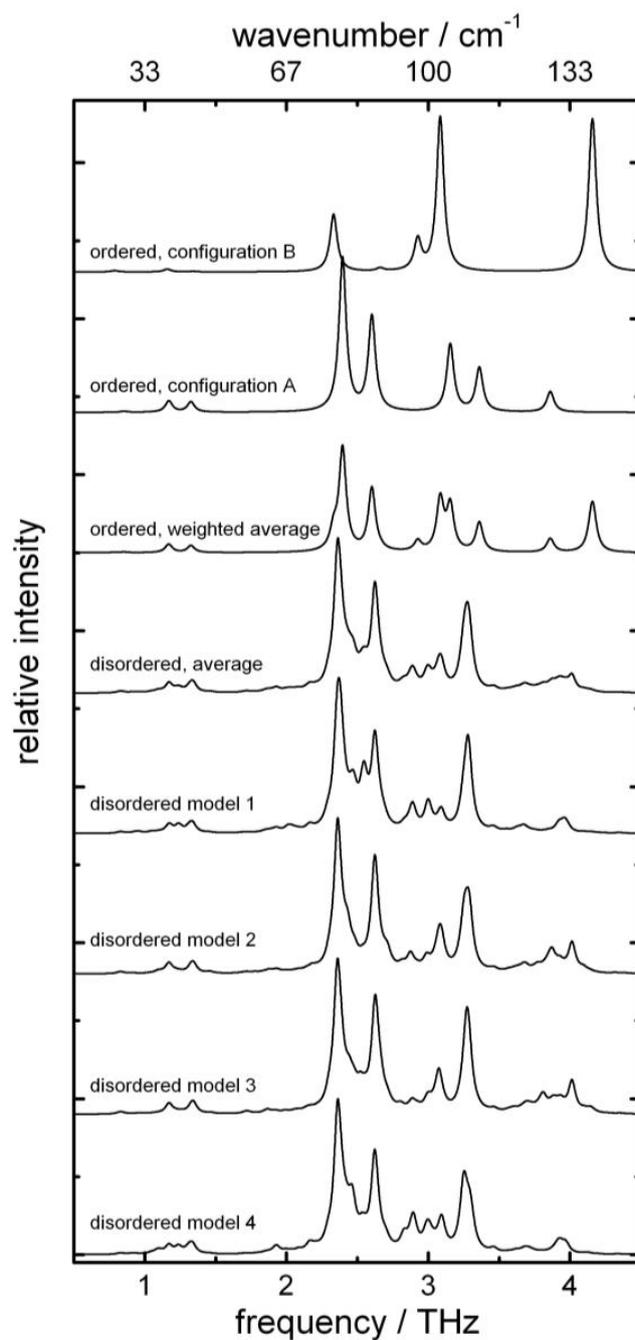


Figure 7.3. Simulated terahertz absorption spectra from the ordered models of the benzoic acid crystal structure and from the four models with a 2 : 1 ratio of dimers in configurations *A* and *B*. The weighted average spectrum (third from the top) is a simple 2 : 1 average of the spectra calculated from the two ordered crystal structure models. The average disordered spectrum (fourth from the top) is an average over the spectra calculated from the four disordered models. The peak shape for each calculated vibrational mode is assumed to be Lorentzian, with a full width at half maximum of 0.1 THz. From [180]

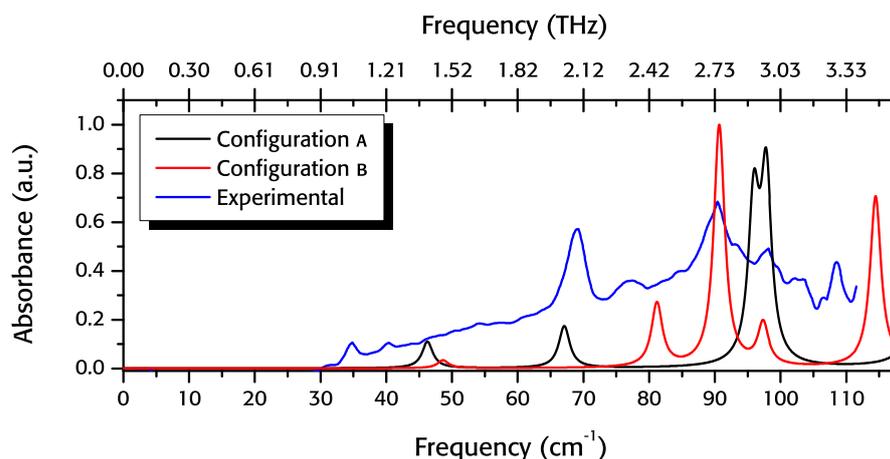


Figure 7.4. Experimental spectrum of benzoic acid collected at 50 K, compared with CASTEP calculations with dispersion corrected PBE functional, configuration A and B

age of the spectra of the two ordered configurations.

A calculation with a  $2a \times 3b \times c$  supercell (720 atoms) such as the one reported above would be infeasible with CASTEP because of the sheer size of the system. We complement here analysis with the dispersion corrected calculation for both of the unit cell configurations A and B. The calculations confirm that the energy of configuration A is slightly smaller than configuration B: the energy difference between the two configurations is  $1.39 \text{ kJ mol}^{-1}$ , slightly smaller than the value for the DMACRYS calculations,  $1.7 \text{ kJ mol}^{-1}$ .

The spectra from these CASTEP calculations is reported in Figure 7.4. Although it would be incorrect to consider the total spectrum of the disordered system as a weighted sum of the two ordered structures (since the phonons depend on the local and global environment of the dimers, and the DMACRYS calculations seem to suggest that the disorder is not organised in macroregions) we cannot fail to notice that the two spectra seem to be complementary to each other, with all of the experimental peaks having a calculated counterpart at the same frequency (except for the two low frequency features at  $35 \text{ cm}^{-1}$  and  $40 \text{ cm}^{-1}$ ).

## 7.2 ANHYDROUS THEOPHYLLINE

The experimental spectrum of theophylline was recorded with a range of temperatures from 80 K to 240 K. The experimental sample was obtained by Sigma-Aldrich (code T1633-50G, purity > 99 %).

The experimental measurement, together with the DMACRYS simulated spectra, is reported in Figure 7.5. The experimental peaks are reliable only up to about  $110\text{ cm}^{-1}$  due to experimental uncertainties. We can clearly identify three regions, each with several peaks: a low frequency feature centred around  $35\text{ cm}^{-1}$ , a broad peak at  $61\text{ cm}^{-1}$  with three minor shoulders, and the higher frequency region from  $90\text{ cm}^{-1}$  to  $120\text{ cm}^{-1}$ , comprising of 3-4 peaks.

Another low temperature spectrum is reported in the literature [57], which suggest that the shoulder at  $51\text{ cm}^{-1}$  in our measured spectrum is a peak.

### 7.2.1 DMACRYS calculations

There is very little variation between the experimental crystal structure and the one calculated in DMACRYS (optimised from BAPLOT01, room temperature determination), in particular using the FIT forcefield.

The absorption frequencies calculated with FIT and w99 are very close to each other up to  $90\text{ cm}^{-1}$ , and close in intensity (with the exception of the peak at  $62\text{ cm}^{-1}$ , which is very weak with the FIT forcefield, but quite strong in the w99 calculation).

The lowest experimental absorption feature can be assigned to the

	Exp. (295 K)	Crystal parameters variation			
		w99	Change/%	FIT	Change/%
$a / \text{\AA}$	24.61	24.49	-0.47	24.67	0.23
$b / \text{\AA}$	3.83	3.92	2.27	3.92	2.35
$c / \text{\AA}$	8.50	8.35	-1.79	8.51	0.17
Vol. / $\text{\AA}^3$	801.4	801.08	-0.04	823.5	2.76
<i>F</i> value		14.09		8.45	

Table 7.3. Variation of the lattice parameters for the DMACRYS geometry optimisation of anhydrous theophylline, with FIT and w99 forcefield. Symmetry constraints: unit cell angles  $\alpha = \beta = \gamma = 90^\circ$ . The initial structure parameters are taken from BAPLOT01

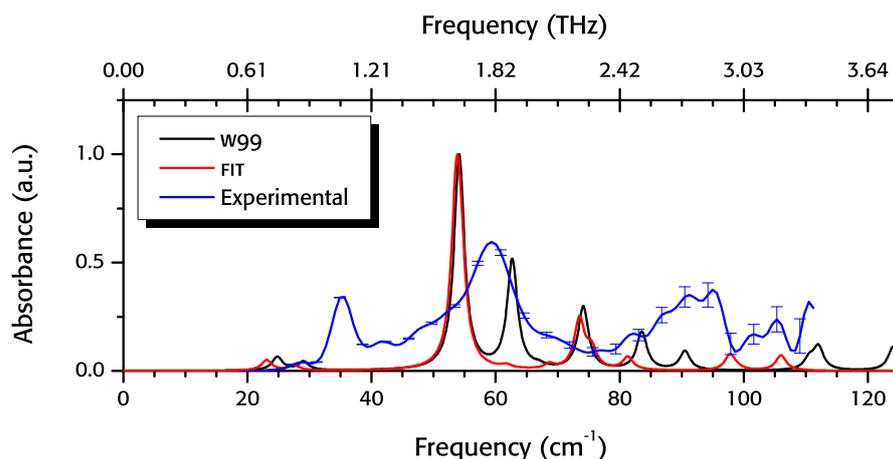


Figure 7.5. Experimental spectrum of anhydrous theophylline collected at 100 K, compared with DMACRYS calculations, w99 and FIT forcefield

	Experimental (295 K)	Crystal parameters variation	
		PBE-D	Change /%
$a / \text{\AA}$	24.61	24.142	-1.90
$b / \text{\AA}$	3.83	3.62	-5.48
$c / \text{\AA}$	8.50	8.47	-0.35
Volume / $\text{\AA}^3$	801.4	741.06	-7.52

Table 7.4. Change in the lattice parameters for the CASTEP geometry optimisation of anhydrous theophylline, dispersion corrected PBE. Symmetry constraints: unit cell angles  $\alpha = \beta = \gamma = 90^\circ$ . The initial structure parameters are taken from BAPLOT01

lowest two calculated frequencies, that are located  $\approx 10 \text{ cm}^{-1}$  lower in frequency – leaving one of the experimental features unassigned. There is a good agreement between the experimental and calculated frequencies in the central group of features, but their relative absorption intensities are poorly reproduced.

In both of the calculated spectra there is one peak just below  $50 \text{ cm}^{-1}$  with intensity much higher than the others, while in the experimental spectrum the three absorption groups display similar magnitude for the absorbance.

### 7.2.2 CASTEP calculation

The minimisation with CASTEP shows a 7.5 % contraction in the unit cell volume, mainly due to the decrease of the  $b$  axis (Table 7.5).

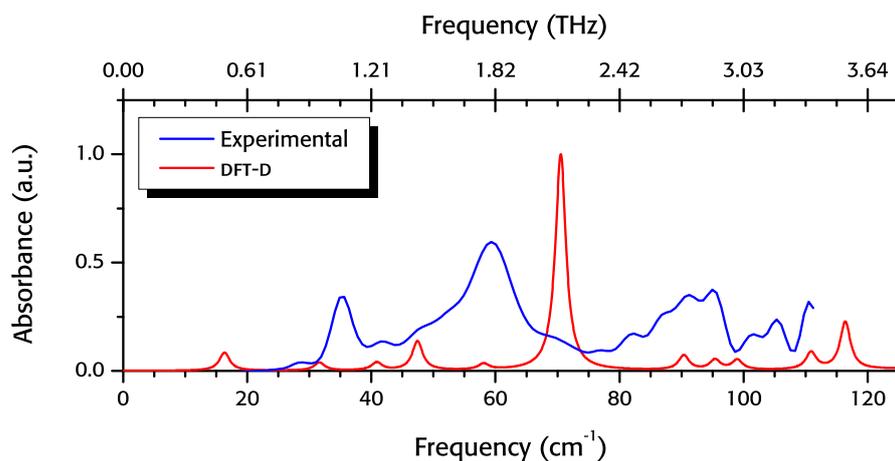


Figure 7.6. Experimental spectrum of anhydrous theophylline collected at 100 K, compared with a CASTEP calculation, dispersion corrected PBE functional

As found in the DMACRYS calculations, there is one clearly predominant absorption phonon in the experimental window. However, this mode is at much higher frequency ( $71\text{ cm}^{-1}$ ) compared to the two DMACRYS calculations.

The lowest absorption frequency ( $16\text{ cm}^{-1}$ ) is about  $12\text{ cm}^{-1}$  lower than the lowest experimental value; this peak, along with the two at  $32\text{ cm}^{-1}$  and  $41\text{ cm}^{-1}$ , can be assigned to the lowest frequency experimental group.

The following three calculated peaks have a frequency similar to that of the central experimental group of features, although the assignment cannot be made, because of the poor agreement between calculated and measured intensities.

The high frequency region shows a decent match both with position and relative intensity, with the highest frequency peak (at  $117\text{ cm}^{-1}$ ) displaying double the intensity of the previous four.

### 7.3 THEOPHYLLINE MONOHYDRATE

The experimental spectra of theophylline monohydrate was recorded at low temperature by Axel Zeitler, and is reported in literature [57].

The spectrum (Figure 7.7) comprises of a weak low frequency absorption frequency at  $30\text{ cm}^{-1}$ , a doublet feature centred around  $60\text{ cm}^{-1}$ , a

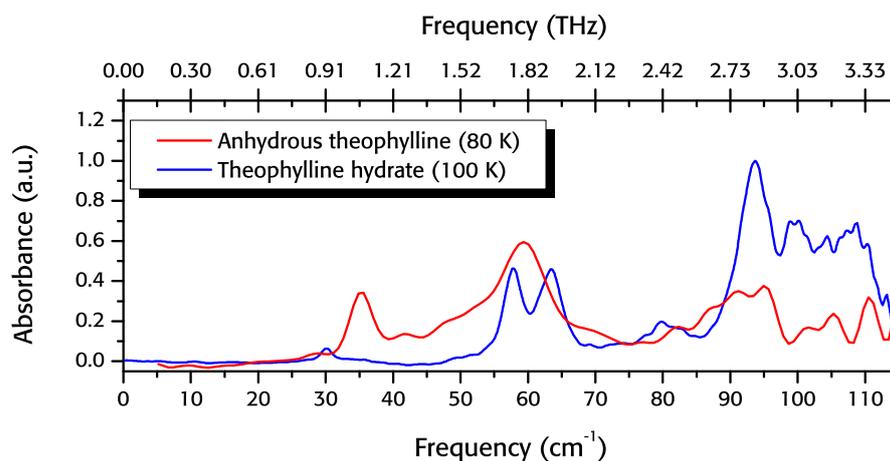


Figure 7.7. Comparison of the experimental spectra of anhydrous and theophylline monohydrate, respectively at 100 K and 80 K

broad absorption frequency at  $80\text{ cm}^{-1}$  and a series of absorption features from  $90\text{ cm}^{-1}$  to  $120\text{ cm}^{-1}$ . The main differences with the spectrum of the neat theophylline crystal are in the region below  $50\text{ cm}^{-1}$ .

### 7.3.1 DMACRYS calculations – symmetry constrained

The structure of theophylline hydrate recorded in the database, THEOPH01, reports disorder of the hydrogen atoms in water and in one of the methyl groups, with a 50 % occupancy for two sites, as reported in chapter 6.2.

Water molecules in the structure are symmetry related by a centre of inversion. With this constraint, the hypothetical positions of the hydrogen atoms in water are not optimal: the water-water close contacts are either  $\text{H} \cdots \text{H}$  or  $\text{O} \cdots \text{O}$ , instead of forming the favoured hydrogen bond  $\text{H}-\text{O} \cdots \text{H}$ . The geometry optimisation with DMACRYS cannot reach this molecular configuration due to the symmetry constraints imposed by the unit cell crystal group: the rotation (see the arrows in Figure 7.8) can only change the close contact from  $\text{H} \cdots \text{H}$  to  $\text{O} \cdots \text{O}$ .

The geometry optimisation with DMACRYS, therefore is constrained by symmetry in a configuration that cannot achieve optimal hydrogen bonds: the structure is trapped in a higher energy state. The results of the optimisation is summarised in Table 7.5. The  $F$  values are very high, mainly due to the complete repositioning of the water molecules and the structural changes necessary to relocate them within the structure, as

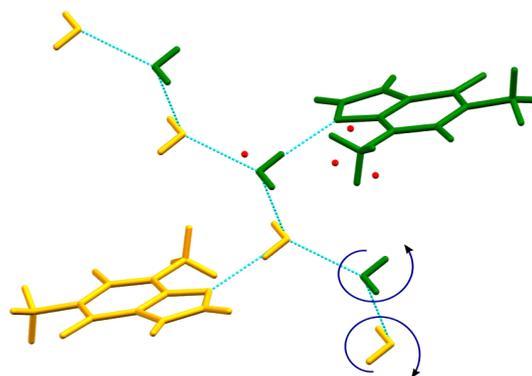


Figure 7.8. Hydrogen bond network of the water molecules within the THEOPH01 crystal structure. Molecules with the same symmetry equivalence are in the same colours, the hydrogen bonds are coloured in blue. Alternative positions in two of the molecules for the disordered hydrogen are coloured in red

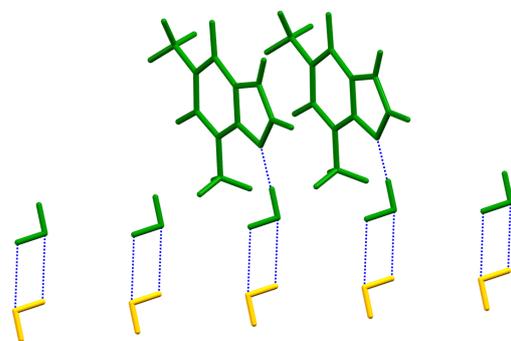


Figure 7.9. Molecular packing of theophylline monohydrate after the minimisation with DMACRYS, w99 forcefield. Molecules with the same symmetry equivalence are in the same colours, the hydrogen bonds are coloured in blue

shown in Figure 7.9. The water molecules reorganise from one column to two parallel columns. The required change in the shape of the unit cell of the crystal results in a compression along the  $a$  axis and on expansion along the  $b$  axis, with a slight increase in volume.

The minimisation converged for both the forcefields, and the minimised structure has an energy decrease upon lattice minimisation of almost  $60 \text{ kJ mol}^{-1}$ ; the calculated spectrum is reported in Figure 7.10. The calculated FIT structure, reached a stationary point, but not a real minimum: as a result the phonon spectrum had imaginary frequencies. Oddly (since the w99 optimised structure is very different from the experimental determination), the calculated spectrum reproduces the experimental spectrum remarkably well, apart from the low energy

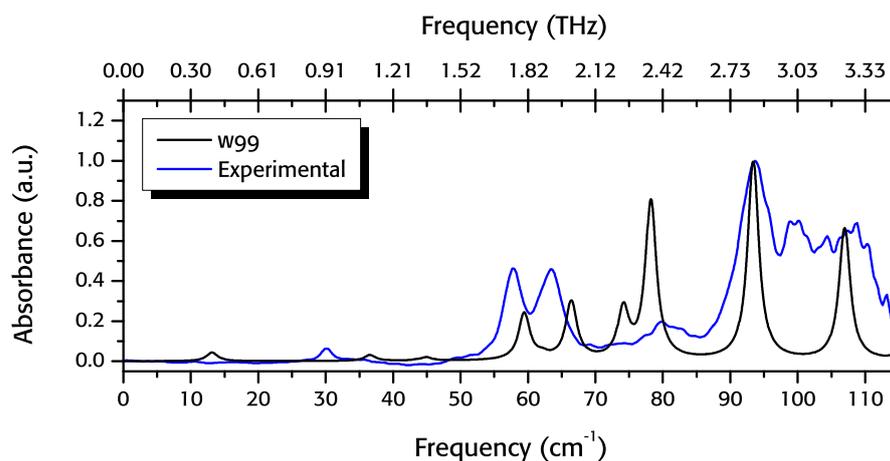


Figure 7.10. Experimental spectrum of theophylline monohydrate collected at 80 K, compared with DMACRYS calculation with full symmetry imposed on the structure, using the w99 forcefield

	Exp. 295 K	Crystal parameters variation			
		w99	Change/%	FIT	Change/%
$a / \text{\AA}$	4.47	4.01	-10.25	4.11	-7.92
$b / \text{\AA}$	15.35	16.09	4.76	15.24	-0.74
$c / \text{\AA}$	13.12	14.26	8.70	14.70	12.04
$\beta / ^\circ$	97.79	96.92	-0.88	96.87	-0.93
Vol. / $\text{\AA}^3$	891.86	913.32	2.40	915.11	2.60
$F$ value		571.58		607.71	

Table 7.5. Variation of the lattice parameters for the DMACRYS geometry optimisation of theophylline monohydrate with full symmetry imposed on the structure, using FIT and w99 forcefield. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$ . The initial structure parameters are taken from THEOPHO1. Note that the FIT final structure does not represent a real minimum for the energy

calculated absorption at  $16 \text{ cm}^{-1}$  and the intensity of the  $80 \text{ cm}^{-1}$  feature.

### 7.3.2 DMACRYS calculations – symmetry relaxed structure

If we look back at the structure of the monohydrate in Figure 7.8, we see that the water molecules can form a reasonable hydrogen bond network if the hydrogen positions of each of the “green” water molecules occupy the alternative disordered site. The disorder in the crystal structure, therefore, is not random: each water is equally likely to be in either of the two orientations, but all the molecules within a channel will be

# Structure	Initial energy / kJ mol <sup>-1</sup>	Final energy / kJ mol <sup>-1</sup>
0 – w99	-103.09	-147.04
0 – FIT	-100.37	-157.40
1 – w99	-149.38	-160.70
2 – w99	-146.95	-159.37
3 – w99	-147.18	-158.81
4 – w99	-149.64	-160.12
1 – FIT	-157.54	-166.04
2 – FIT	-154.73	-164.62
3 – FIT	-157.56	-165.71
4 – FIT	-157.54	-166.04

Table 7.6. Initial and final lattice energy for the optimisation of the crystal structure of theophylline hydrate in the symmetry constrained (0) and in the four possibilities for the methyl disorder (1–4), calculated with DMACRYS, w99 and FIT forcefield

able to form the energetically favoured hydrogen bond.

The arrangement of molecules described above breaks the centre of inversion symmetry; therefore we had to construct a unit cell with reduced symmetry for the minimisation with DMACRYS.

Since each of the molecules within the unit cell has one disordered group, we have 16 possible combinations for the choice of unit cell. However, the choices for the water are not independent, since we want to force the formation of a hydrogen bond chain, and half of the remaining eight choices are equivalent under inversion symmetry. We therefore performed four calculations (one for each possible configuration of the methyl groups within the reduced symmetry unit cell). The disorder sites for the hydrogen atoms in the structure are shown in Figure 7.11, with their van der Waals radii.

The final energies of the four structures, reported in Table 7.6, are within a 2 kJ mol<sup>-1</sup> window: the disorder of the methyl group, therefore, does not have such a big effect on the calculations, but the elimination of the symmetry inversion from the unit cell lowered the energy by more than 10 kJ mol<sup>-1</sup> in comparison with the symmetry constrained calculation.

The summary of the structural changes for all the possible disordered structures is reported in Table 7.7. In general, the Gavezzotti factor  $F$  is quite high, but much lower in comparison with the symmetry constrained calculations. There is a big variation along the  $c$  axis, which in

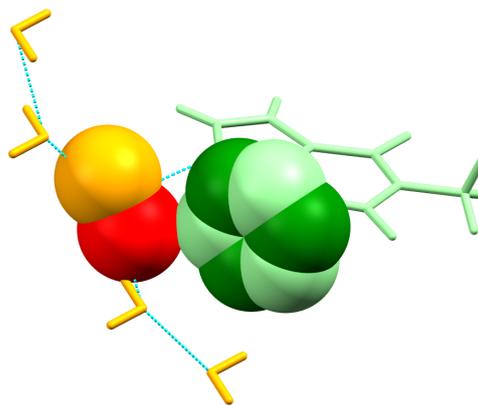


Figure 7.11. Disordered sites for the hydrogen in water (shades of red) and in the theophylline molecule (shades of green)

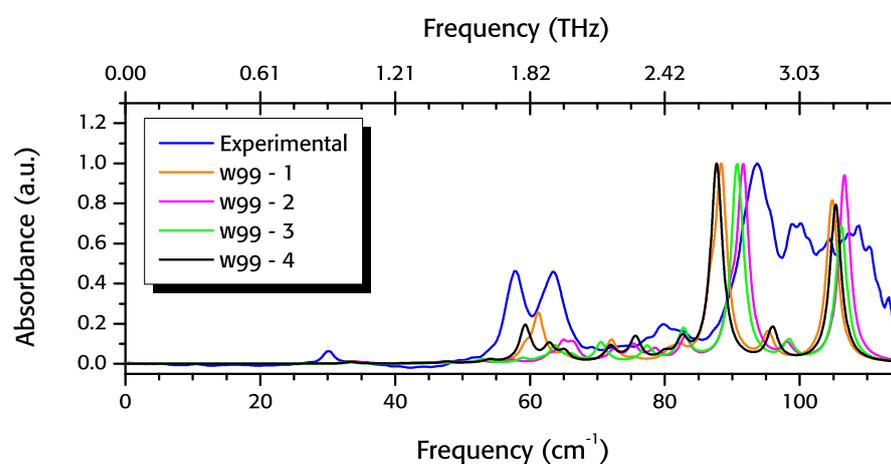


Figure 7.12. Experimental spectrum of theophylline monohydrate collected at 80 K, compared with the four disorder methyl configurations, with DMACRYS and the w99 forcefield

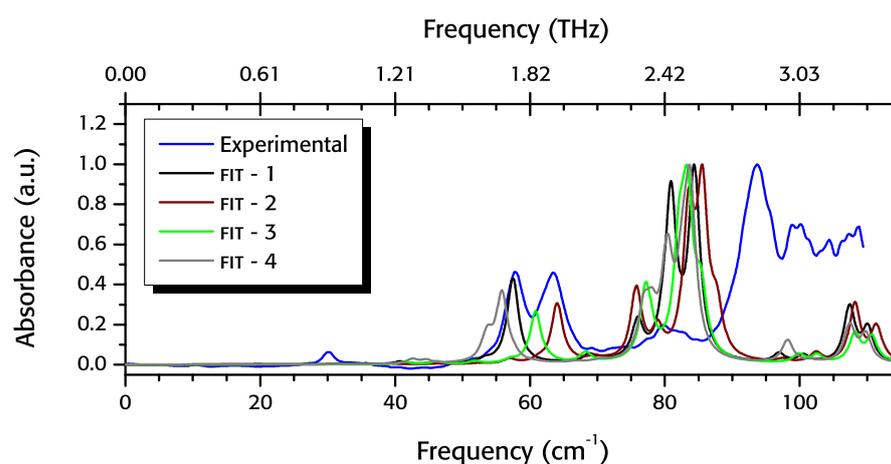


Figure 7.13. Experimental spectrum of theophylline monohydrate collected at 80 K, compared with the four disorder methyl configurations, with DMACRYS and the FIT forcefield

	Crystal parameters variation				
	Exp.	w99	Change/%	FIT	Change/%
Structure 1					
$a / \text{\AA}$	4.47	4.25	-4.86	4.31	-3.51
$b / \text{\AA}$	15.35	14.97	-2.50	15.21	-0.89
$c / \text{\AA}$	13.12	14.39	9.69	14.20	8.24
$\beta / ^\circ$	97.79	97.70	-0.08	97.11	-0.69
Volume / $\text{\AA}^3$	891.86	907.60	1.76	924.63	3.67
$F$ value		218.51		151.66	
Structure 2					
$a / \text{\AA}$	4.47	4.30	-3.57	4.36	-2.41
$b / \text{\AA}$	15.35	14.98	-2.42	15.22	-0.85
$c / \text{\AA}$	13.12	14.19	8.15	14.05	7.10
$\beta / ^\circ$	97.79	97.43	-0.36	97.01	-0.80
Volume / $\text{\AA}^3$	891.86	908.35	1.84	925.91	3.81
$F$ value		161.02		115.36	
Structure 3					
$a / \text{\AA}$	4.47	4.27	-4.28	4.34	-2.69
$b / \text{\AA}$	15.35	14.97	-2.47	15.24	-0.75
$c / \text{\AA}$	13.12	14.25	8.64	14.04	6.97
$\beta / ^\circ$	97.79	97.47	-0.32	96.98	-0.83
Volume / $\text{\AA}^3$	891.86	905.17	1.49	923.06	3.50
$F$ value		176.64		110.61	
Structure 4					
$a / \text{\AA}$	4.47	4.21	-5.77	4.29	-3.78
$b / \text{\AA}$	15.35	14.95	-2.62	15.21	-0.89
$c / \text{\AA}$	13.12	14.49	10.39	14.19	8.15
$\beta / ^\circ$	97.79	97.76	-0.03	97.09	-0.72
Volume / $\text{\AA}^3$	891.86	903.45	2.40	921.31	3.30
$F$ value		248.93		149.11	

Table 7.7. Variation of the lattice parameters for the DMACRYS geometry optimisation of theophylline hydrate, with FIT and w99 forcefield. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$ . The initial structure parameters are taken from THEOPH01, and each of the structure has one of the possible methyl configurations

the w99 calculations is balanced with a variation along the other two axes to leave the volume almost unchanged.

The variation in the structure during the energy minimisation is much less pronounced in the case of the FIT forcefield calculations; anyway, the

agreement of the terahertz absorption is better with the w99 forcefield.

In Figure 7.12 we can notice that the higher frequency peaks agree very well with the experimental features, and it is very similar in shape among the four calculations; the greatest variability is in the description of the experimental doublet at  $60\text{ cm}^{-1}$ , with one of the structures (number 4, the second lowest energy crystal structure) resembling the experimental shape much better.

In contrast, the FIT calculated spectra do not display a very good agreement: there is only one major calculated peak around  $60\text{ cm}^{-1}$  instead of the two measured ones, while the experimental peak at  $90\text{ cm}^{-1}$  is calculated shifted to lower frequencies by a value between  $10\text{ cm}^{-1}$  and  $15\text{ cm}^{-1}$ .

## 7.4 PARACETAMOL FORM I

The experimental spectra of paracetamol form I was measured for a range of temperatures from 10 K to 200 K in a helium cooled chamber, with the determination at 29 K being the clearest.

All the measurements are characterised by a very strong background. The strongest absorption peaks (Figure 7.14) are at  $51\text{ cm}^{-1}$ ,  $62\text{ cm}^{-1}$ ,  $75\text{ cm}^{-1}$ ,  $93\text{ cm}^{-1}$  and  $107\text{ cm}^{-1}$ , increasing monotonically in absorbance. There are at least two other strong absorptions at  $124\text{ cm}^{-1}$  and  $136\text{ cm}^{-1}$ , while other peaks are unclear due to experimental uncertainties.

### 7.4.1 DMACRYS calculations

We used the CSD structure HXACAN06, determined by neutron scattering at 100 K as a starting point for our geometry minimisations. This crystal structure determination shows a big librational motion of the methyl group and suggests the possibility of disorder within the structure [205].

The  $F$  value for the geometry optimisations of the crystal structure is very low, with the FIT minimised structure exhibiting a larger deviation of the axis lengths from the experimental values than the w99 calculated values (although the final volumes are almost the same).

The calculated spectra for these calculations are reported in Figure 7.14. We can see that there is a good agreement among the calculated

peaks and intensities, except for the two highest features in the experimental range.

There is partial agreement between the experimental data and the DMACRYS calculations. In the region from  $40\text{ cm}^{-1}$  to  $80\text{ cm}^{-1}$  one of the absorptions is missing; if we assume that the  $94\text{ cm}^{-1}$  experimental peak corresponds to the  $86\text{ cm}^{-1}$  calculated peak, the whole upper frequency absorption region is shifted to higher frequency by  $12\text{ cm}^{-1}$ .

#### 7.4.2 CASTEP calculations

Relaxation of the unit cell was allowed only in the dispersion-corrected CASTEP calculation, with a volume decrease of 8.47 %. The biggest contraction is along the *b* direction, although all of the optimised axes are decreased in value.

The dispersion corrected calculation is the one showing worst agreement with the experimental data. There are only two, very weak absorptions at  $58\text{ cm}^{-1}$  and  $80\text{ cm}^{-1}$  instead of the three experimental values. The higher group of vibrational modes in the range  $105\text{ cm}^{-1}$  to  $130\text{ cm}^{-1}$  is found at a frequency  $20\text{ cm}^{-1}$  higher than the group of absorption seen in the experimental spectrum.

We reported two non-dispersion corrected calculations in Figure 7.15, with different energy cutoff for the basis set (1050 eV vs 1200 eV). We can see how the intensities significantly differ between the two spectra, with the 1200 eV showing a better agreement with the experimental data.

## 7.5 N-PHENYL-ACETAMIDE DERIVATIVES

The spectra of the N-phenyl-acetamide derivatives are reported in chapter 11.

## 7.6 NITROFURANTOIN FORM $\beta$

The sample of the stable  $\beta$  form of nitrofurantoin was obtained from Sigma-Aldrich (code N7878-25G, purity > 99 %) and used without fur-

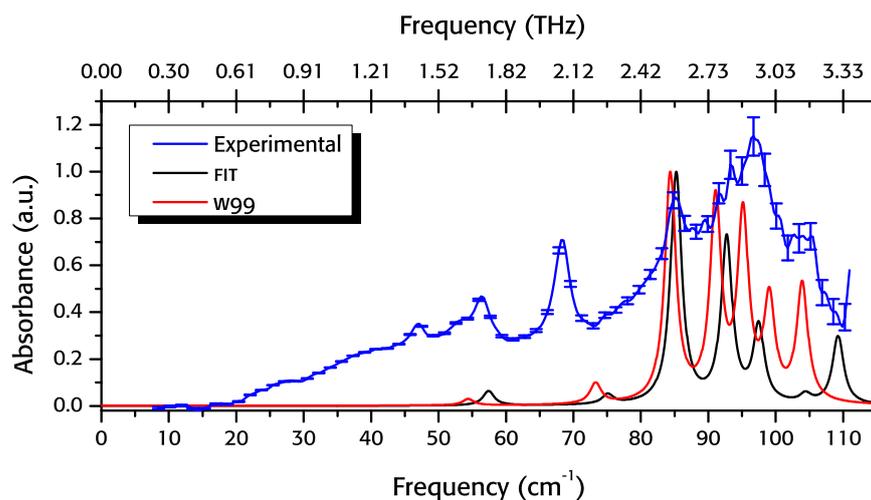


Figure 7.14. Experimental spectrum of paracetamol form I (29 K), compared with DMACRYS calculations with the w99 and FIT forcefields

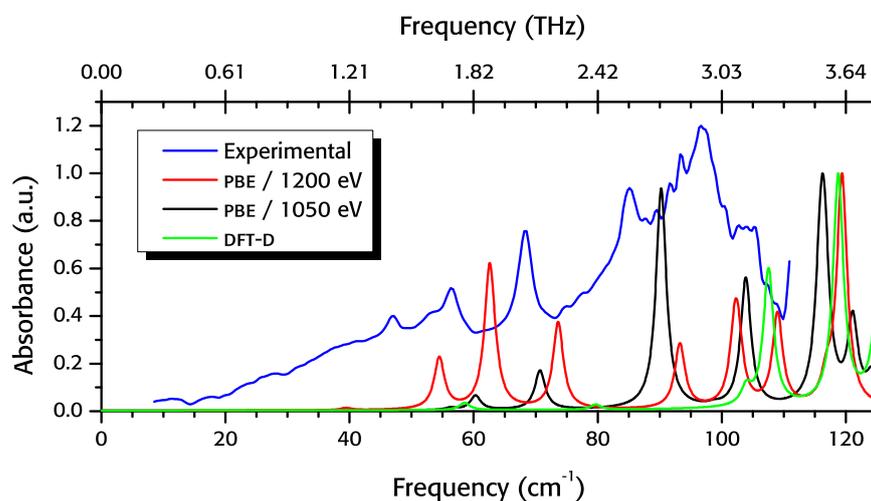


Figure 7.15. Experimental spectrum of paracetamol form I collected at 29 K, compared with PBE and dispersion corrected PBE calculations

	Crystal parameters variation				
	Exp. (100 K)	w99	Change/%	FIT	Change/%
$a / \text{\AA}$	12.76	12.71	-0.36	12.57	-1.49
$b / \text{\AA}$	9.25	9.38	1.43	9.36	1.15
$c / \text{\AA}$	7.08	7.20	1.65	7.28	2.77
$\beta / ^\circ$	115.52	115.00	-0.45	114.57	-0.81
Vol. / $\text{\AA}^3$	754.51	778.51	3.18	778.66	3.20
$F$ value		10.97		19.24	

Table 7.8. Variation of the lattice parameters for the DMACRYS geometry optimisation of paracetamol form I, with FIT and w99 forcefield. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$

	Experimental (100 K)	Crystal parameters variation	
		PBE-D	Change /%
$a / \text{\AA}$	12.76	12.48	-2.19
$b / \text{\AA}$	9.25	8.74	-5.51
$c / \text{\AA}$	7.09	6.95	-1.97
$\beta / ^\circ$	115.52	114.48	-0.90
Volume / $\text{\AA}^3$	754.51	690.53	-8.47

Table 7.9. Change in the lattice parameters for the CASTEP geometry optimisation of paracetamol form I, dispersion corrected PBE. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$ . The initial structure parameters are taken from HXACAN06

ther treatment.

The measurements were performed with helium cooled equipment, and we recorded a series of spectra in the temperature range from 10 K to 75 K. Once again the clearest measurement was not the one at lowest temperature, but the measurement at 29 K, reported in Figure 7.16.

The measurement displays very low experimental uncertainties up to high frequencies. The most prominent experimental features are at  $68 \text{ cm}^{-1}$ ,  $92 \text{ cm}^{-1}$ ,  $104 \text{ cm}^{-1}$  and  $108 \text{ cm}^{-1}$ ; there are also other weaker absorptions and shoulders.

### 7.6.1 The DMACRYS calculations

The minimisation of the structure with DMACRYS led to an expansion of the experimental unit cell by a similar amount as compared with the w99 and the FIT forcefield: the sum of the terms due to the unit cell variation is by far the most important contribution to the  $F$  factor, with 31.08 and 21.11 respectively. The two final structures are almost superimposable, with an equal length of the hydrogen bond (see Figure 6.9) increased from the experimental value of  $1.916 \text{ \AA}$  to  $1.938 \text{ \AA}$ .

The calculated terahertz spectrum with the two forcefields are almost identical to each other, as well (Figure 7.17): there are two major peaks agreeing with the experimental spectrum; the calculated peaks at  $37 \text{ cm}^{-1}$  do not have an experimental counterpart, and the calculations do not find any relevant absorption with frequency higher than  $100 \text{ cm}^{-1}$ . We can assume that these higher frequency observed features correspond to intramolecular absorptions, which are not accounted for

	Exp. (293 K)	Crystal parameters variation			
		w99	Change/%	FIT	Change/%
$a / \text{\AA}$	7.84	8.17	4.19	8.11	3.43
$b / \text{\AA}$	6.46	6.56	1.61	6.57	1.70
$c / \text{\AA}$	18.92	18.29	-3.29	18.44	-2.53
$\beta / ^\circ$	93.18	92.03	-1.23	91.75	-1.53
Vol. / $\text{\AA}^3$	957.66	981.41	2.48	982.92	2.63
$F$ value		47.07		33.65	

Table 7.10. Variation of the lattice parameters for the DMACRYS geometry optimisation of nitrofurantoin form  $\beta$ , with FIT and w99 forcefield. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$ . The initial structure parameters are taken from LABJON

in DMACRYS calculations.

### 7.6.2 The CASTEP calculations

We performed two CASTEP calculations, with and without dispersion correction. The crystal structure of the non-dispersion corrected calculation is almost superimposable with the experimental one.

The relaxation of the unit cell was performed only with the PBE/DFT-D scheme, and the results are reported in Table 7.11. The variation of the unit cell parameters is smaller than the corresponding dispersion corrected calculation for the other systems mentioned in the previous sections (variable from 6 % to 10 %), but still larger than the variation with DMACRYS.

The positions of the calculated peaks in the terahertz spectrum (Figure

	Experimental (293 K)	Crystal parameters variation	
		PBE-D	Change /%
$a / \text{\AA}$	7.84	7.59	-3.19
$b / \text{\AA}$	6.46	6.47	0.15
$c / \text{\AA}$	18.92	18.64	-1.47
$\beta / ^\circ$	93.18	90.82	-2.53
Volume / $\text{\AA}^3$	957.66	915.16	-4.43

Table 7.11. Change in the lattice parameters for the CASTEP geometry optimisation of nitrofurantoin  $\beta$ , dispersion corrected PBE. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$ . The initial structure parameters are taken from LABJON

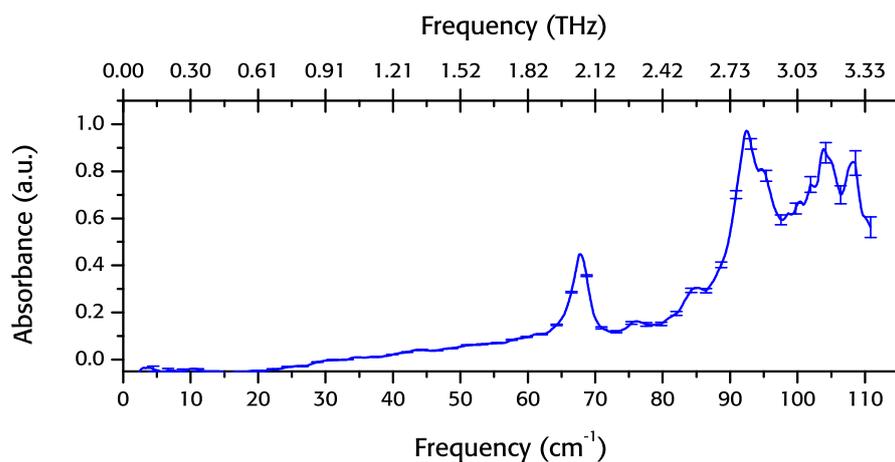


Figure 7.16. Spectrum of nitrofurantoin  $\beta$  collected at 29 K

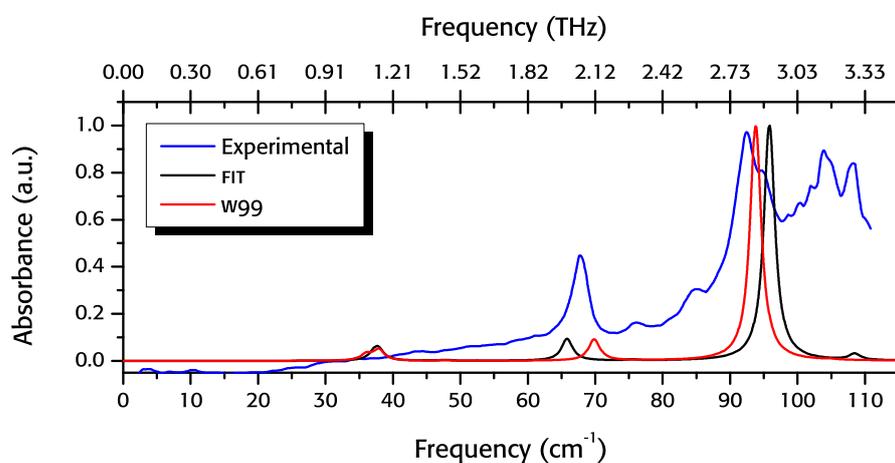


Figure 7.17. Comparison of the experimental spectrum of nitrofurantoin  $\beta$  with the DMACRYS calculated spectra

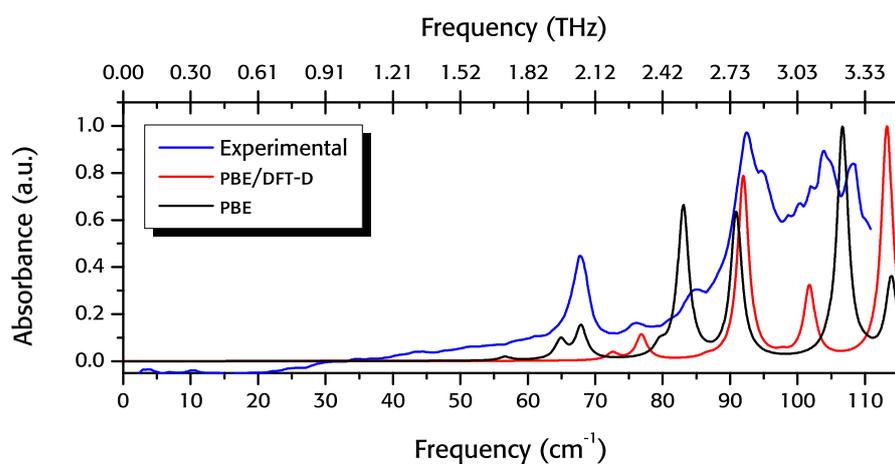


Figure 7.18. Comparison of the experimental spectrum of nitrofurantoin  $\beta$  with the CASTEP calculated spectra

7.18) for the PBE calculation are very close to experimental features; the intensity is not always matching, such as at  $83\text{ cm}^{-1}$  (an experimental shoulder).

The PBE/DFT-D are very similar in shape with the PBE, but shifted at higher frequency by  $8\text{ cm}^{-1}$  to  $10\text{ cm}^{-1}$ . The calculation seems to underestimate the frequencies in the high end of the terahertz spectrum, while underestimating frequencies of the lower energy modes. Furthermore, the comparison of the eigenvectors (Figure 8.14f, page 157) would make us discard the idea of a pure energy shift.

## 7.7 NITROFURANTOIN HYDRATE, FORM II

The orthorhombic polymorph (form II) is the stable nitrofurantoin hydrate form at room temperature, and it was obtained by Shyam Karki by following the experimental procedure outlined by Pienaar *et al.* [206]. The spectra were collected over a range of temperatures from 80 K to 295 K. The clearest determination is the lowest temperature one, in Figure 7.16.

The phonon calculation with CASTEP, PBE/DFT-D, (the biggest we attempted, with 16 molecules, 208 atoms and an initial unit cell volume of  $2166\text{ \AA}^3$ ) found seven imaginary frequencies; the spectrum is therefore not reported in this section. In the time available we were not able to continue the calculations to try to find a true energy minimum.

### 7.7.1 The DMACRYS calculations

The geometry optimisation with the w99 forcefield left the unit cell parameters nearly unchanged (see Table 7.12), and a unit cell volume change of only 2 %.

By far, the biggest contribution to the  $F$  value (104.6) is the rotation of the 8 molecules of water by  $29^\circ$  each. This rotation of the water molecules leads to changes in the hydrogen bond distances relative to the experimentally determined structure: this might be a signal of some inadequacy of the electrostatic part of the intermolecular potential, or in the evaluation of the strength of the four hydrogen bonds (there is a complex interaction of water with the surrounding molecules, see Fig-

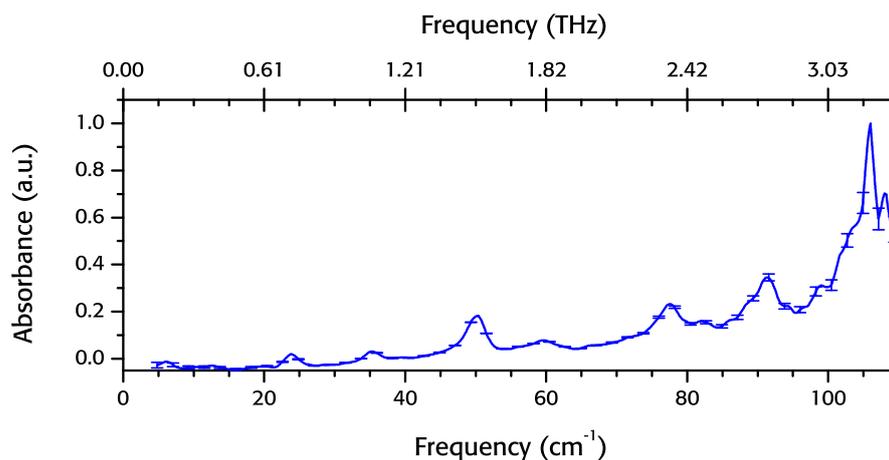


Figure 7.19. Experimental spectrum of nitrofurantoin hydrate form II at 80 K

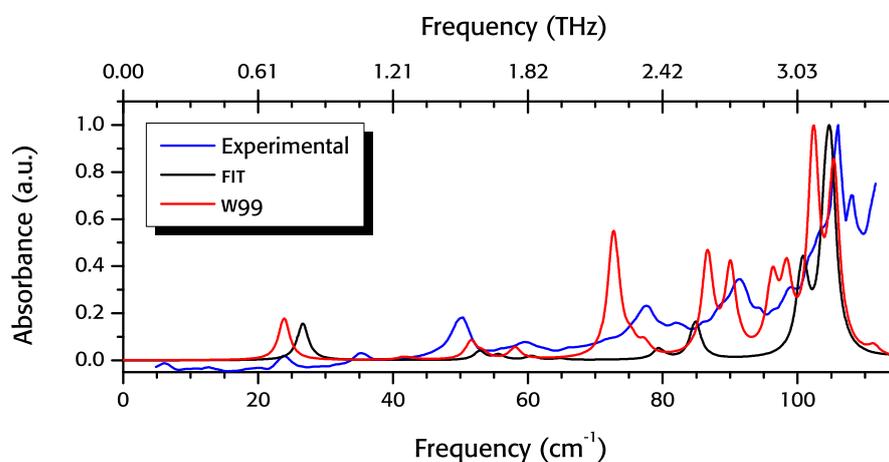


Figure 7.20. Comparison of the spectra of nitrofurantoin hydrate form II generated with DMACRYS, w99 and FIT forcefield with the experimental spectrum. The initial structure is taken from HAXBUD

	Crystal parameters variation				
	Exp. (293 K)	w99	Change/%	FIT	Change/%
$a / \text{\AA}$	12.64	12.74	0.82	12.70	0.46
$b / \text{\AA}$	9.85	9.81	-0.39	9.34	-5.28
$c / \text{\AA}$	17.38	17.68	1.72	18.40	5.88
Vol. / $\text{\AA}^3$	2166.13	2212.88	2.15	2182.43	0.75
$F$ value		118.24		286.24	

Table 7.12. Variation of the lattice parameters and volume for the DMACRYS geometry optimisation of nitrofurantoin hydrate form II, with FIT and w99 forcefield. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$ . The initial structure parameters are taken from the room temperature determination HAXBUD01

ure 6.11, page 84); it might also be due to a poor determination of the hydrogen positions in the experimental structure – since hydrogen atoms can be difficult to locate within the structure.

With the FIT calculation there is a similar rotation of the water molecules (by almost  $38^\circ$ ), associated with a big reorganisation of the axes relative magnitude: although the volume is almost unchanged, the variation of the  $b$  and  $c$  axes are on the order of 5 %.

Despite the reorientation of water in the calculated structure, the agreement of the calculated phonon spectrum with the experimental determinations is very good: all of the peaks in the region from  $20\text{ cm}^{-1}$  to  $110\text{ cm}^{-1}$  have an excellent agreement in position and intensity.

## 7.8 CARBAMAZEPINE FORM I

Carbamazepine form I is not the most stable form at room temperature, but it can be easily obtained by heating form III to 443 K or higher [207]. The measurement of the spectrum was performed at a cryogenic temperature of 7 K by Axel Zeitler [60]. There are several, very clear absorption peaks (Figure 7.21, top): the strongest are at  $33\text{ cm}^{-1}$ ,  $56\text{ cm}^{-1}$ ,  $74\text{ cm}^{-1}$ ,  $86\text{ cm}^{-1}$  and  $100\text{ cm}^{-1}$ .

Since this is the system with the biggest unit cell among the one we studied (Table 7.13), with four molecules in the asymmetric unit cell, we did not attempt a calculation of the phonon spectrum with CASTEP. In the next section we report the analysis with DMACRYS.

### 7.8.1 The DMACRYS calculations

The structural changes upon lattice energy minimisation with DMACRYS were minimal, with the  $F$  value resulting from the FIT and w99 force-field being smaller than 5. The number of absorption frequencies is related to the number of molecules in the unit cell, and for carbamazepine form I this is higher than for other systems: there are 21 IR-active modes, and we can count no less than 16 absorption frequencies with a non-negligible absorption intensity in the spectra (Figure 7.21). The frequencies are not strongly dependent to the forcefield employed, with the FIT values shifted to higher values only very slightly.

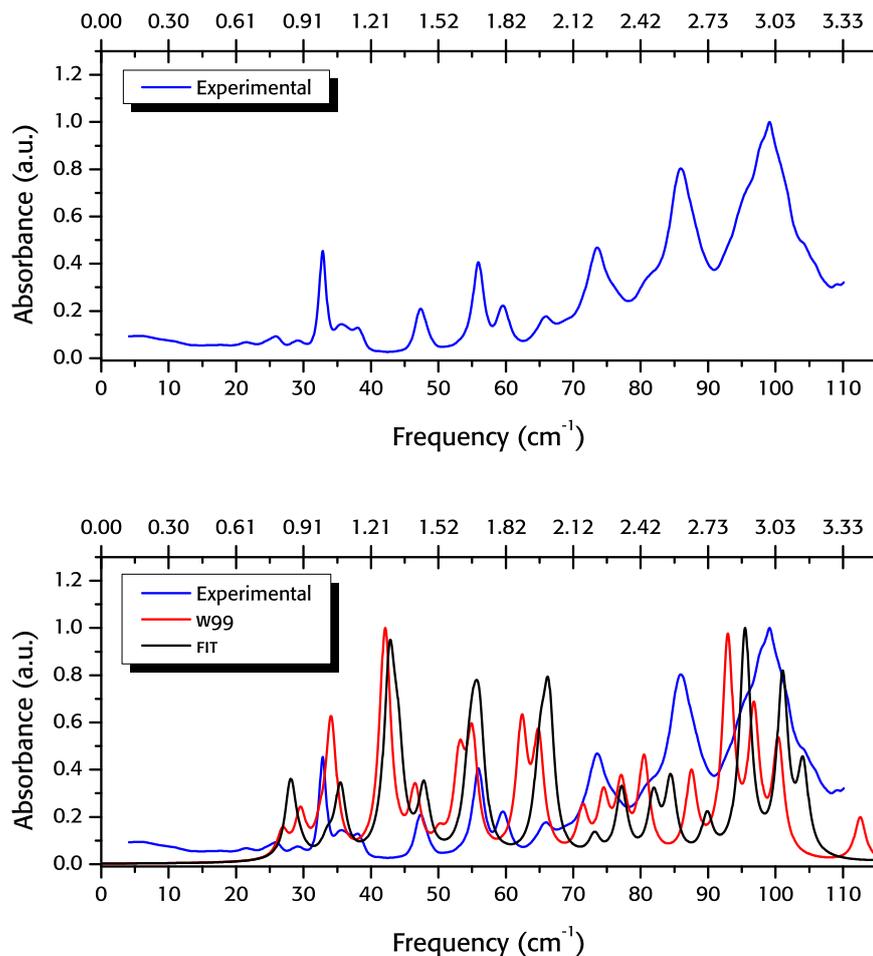


Figure 7.21. Experimental spectrum of carbamazepine form I at 7 K (top) and comparison with DMACRYS, w99 and FIT forcefield (bottom). The initial structure is taken from CBMZPN11

	Crystal parameters variation				
	Exp. (158 K)	w99	Change/%	FIT	Change/%
$a / \text{\AA}$	5.17	5.22	1.04	5.21	0.84
$b / \text{\AA}$	20.57	20.74	0.82	20.67	0.49
$c / \text{\AA}$	22.24	22.29	0.23	22.41	0.74
$\alpha / ^\circ$	84.12	84.43	0.36	84.55	0.51
$\beta / ^\circ$	88.00	87.84	-0.18	87.57	-0.49
$\gamma / ^\circ$	85.19	84.91	-0.32	85.08	-0.12
Volume / $\text{\AA}^3$	2344.82	2394.50	2.11	2394.88	2.13
$F$ value		4.57		4.33	

Table 7.13. Variation of the lattice parameters and volume for the DMACRYS geometry optimisation of carbamazepine form I, with FIT and w99 forcefield. The initial structure parameters are taken from CBMZPN11

	Exp. (158 K)	Crystal parameters variation			
		w99	Change/%	FIT	Change/%
$a / \text{\AA}$	7.53	7.62	1.188	7.64	1.38
$b / \text{\AA}$	11.15	11.09	-0.53	11.05	-0.89
$c / \text{\AA}$	13.91	13.84	-0.51	13.73	-1.26
$\beta / ^\circ$	92.86	92.85	-0.06	93.30	0.47
Vol. / $\text{\AA}^3$	1168.30	1169.81	0.12	1158	-0.38
$F$ value		3.60		9.18	

Table 7.14. Variation of the lattice parameters and volume for the DMACRYS geometry optimisation of carbamazepine form III, with FIT and w99 forcefield. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$ . The initial structure parameters are taken from CBMZPN10

The agreement of the calculated absorption intensities with the experimental values in the range  $20 \text{ cm}^{-1}$  to  $40 \text{ cm}^{-1}$  is better for the w99 forcefield; for both FIT and w99 the strongest calculated absorption is at about  $42 \text{ cm}^{-1}$ , where there is no strong experimental absorption. Other assignments are very difficult, because of the high density of modes and the possible uncertainties in frequency.

## 7.9 CARBAMAZEPINE FORM III

Carbamazepine form III is the stable polymorph and it is used in the drug formulations in tablets [208]. The spectrum measurement was performed by Axel Zeitler at 7 K [60]. The spectrum has different absorption features than form I, allowing for an easy recognition of the crystal form.

There are several characterising absorption peaks (Figure 7.22): the lowest frequency, weak signal at  $32 \text{ cm}^{-1}$  and stronger features at  $45 \text{ cm}^{-1}$ ,  $68 \text{ cm}^{-1}$ ,  $76 \text{ cm}^{-1}$  and  $86 \text{ cm}^{-1}$ . There is a group of peaks around  $100 \text{ cm}^{-1}$ , while the peak at  $115 \text{ cm}^{-1}$  would be in the region of high experimental uncertainties.

### 7.9.1 DMACRYS calculations

The energy minimisation with DMACRYS, as in the case of form I, does not affect very much the unit cell parameters and volume (Table 7.14):

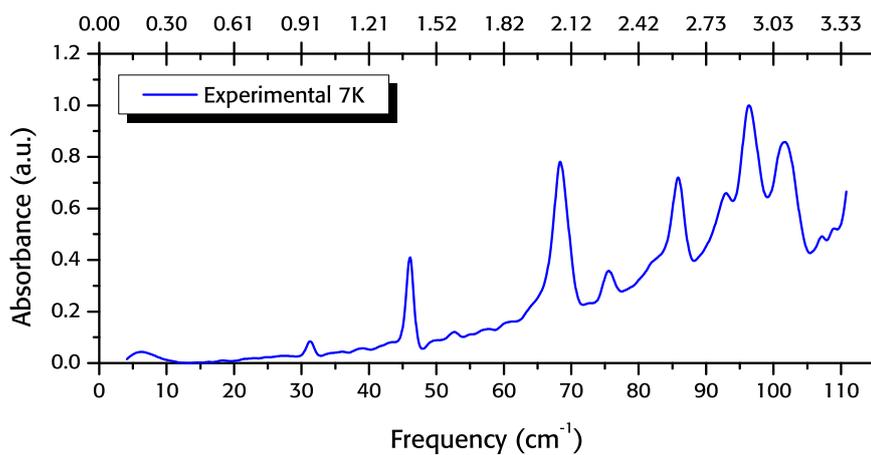


Figure 7.22. Experimental spectrum of carbamazepine form III at 7 K

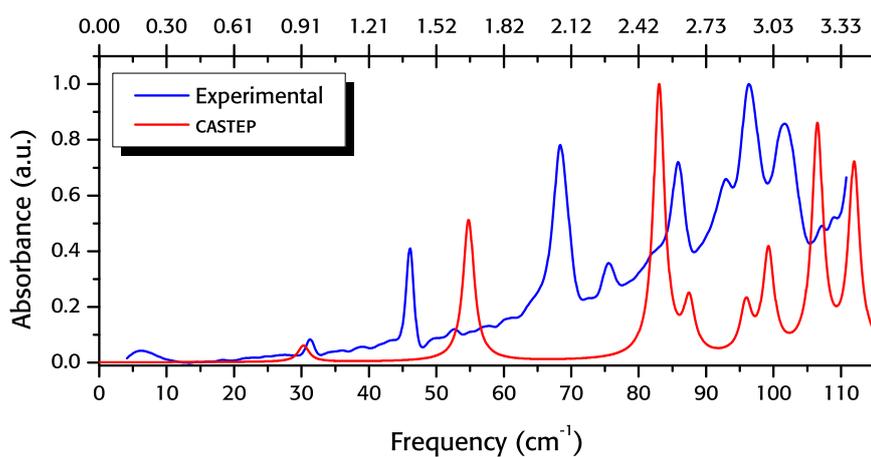
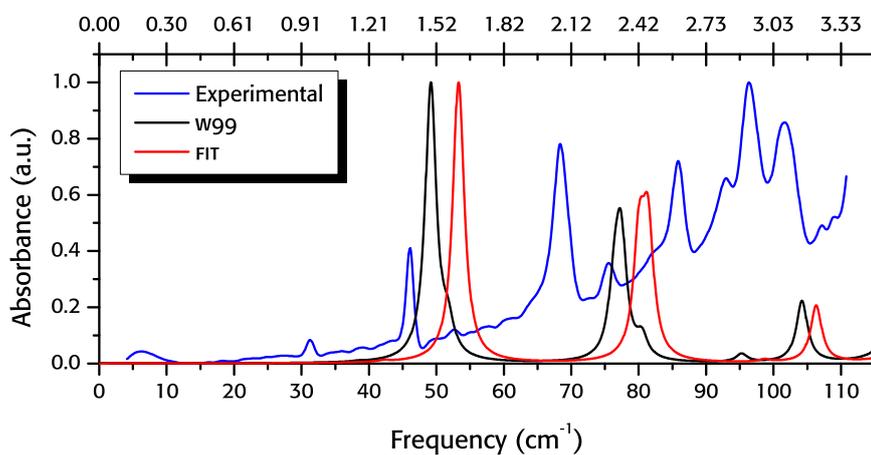


Figure 7.23. Comparison of the experimental spectrum of carbamazepine form III with DMACRYS, w99 and FIT forcefield (top) and CASTEP, PBE/DFT-D (bottom). Initial structure is from CBMZPN02

with both FIT and w99 the  $F$  value is low, less than 15. The FIT calculation has a smaller volume change, but a bigger variation for the individual lattice dimensions.

The absorption at  $31\text{ cm}^{-1}$  has a calculated counterpart, almost invisible (intensity 1/1000 of the strongest calculated) at approximately  $38\text{ cm}^{-1}$  with the two forcefields.

The strongest absorption peaks in the spectra for both the calculations can be assigned to the experimental feature at  $45\text{ cm}^{-1}$ , which is not the strongest experimentally. The lowest experimental absorption is found in the calculations but it is too weak to be seen in Figure 7.23.

The w99 doublet around  $80\text{ cm}^{-1}$  could be assigned to the experimental features at  $68\text{ cm}^{-1}$  and  $76\text{ cm}^{-1}$ , while other assignments would be impossible to judge correctly, due to the scarcity of calculated mode compared to the experimental spectrum.

We can conclude that there is no particular agreement between experiment and calculations for form III for frequency higher than  $60\text{ cm}^{-1}$ .

### 7.9.2 CASTEP calculation

We performed a dispersion corrected CASTEP calculation of the phonon spectrum. The structure undergoes a contraction during the minimisation, with the  $b$  and  $c$  axis shrinking by about 4.5 %, and the total volume contraction of 10.55 % is comparable to what we see in other systems with PBE/DFT-D. In Figure 7.24, as seen along the  $a$  axis direction, the different size of the two structures is shown. The hydrogen bond in the carbamazepine dimer (along the  $c$  axis) is shortened in the calculation ( $2.84\text{ \AA}$  vs  $2.93\text{ \AA}$ ) ; the distance between the phenyl rings in adjacent molecules (parallel to the  $ac$  plane) is also shortened of almost  $0.2\text{ \AA}$ , from  $3.82\text{ \AA}$  vs  $3.64\text{ \AA}$ .

The spectrum is accurate in shape, with clear potential assignments of all the computational peaks to experimental features. In the region  $50\text{ cm}^{-1}$  to  $115\text{ cm}^{-1}$  the calculated absorption peaks are shifted about  $15\text{ cm}^{-1}$  to higher frequency: this is probably the effect of the contraction of the unit cell volume.

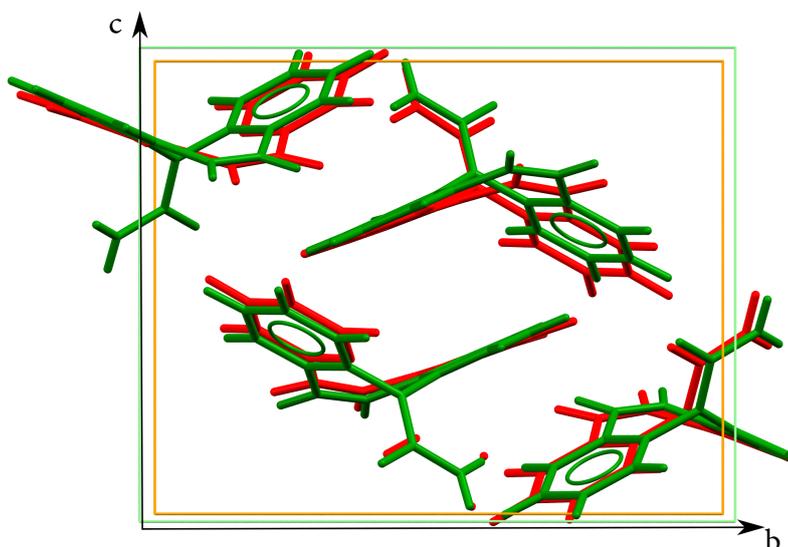


Figure 7.24. Superimposition of the experimental unit cell of carbamazepine form III (in green) with the CASTEP optimised geometry (in red), as seen through the crystallographic  $a$  axis. The unit cell axes are in the same colour of the structure

## 7.10 SELECTED COMPUTED SPECTRA

Due to the difficulty in the synthesis of some of the polymorphs (paracetamol hydrates, for example, are only stable for minutes in air) we were not able to obtain an experimental spectrum for all the systems listed in chapter 6. In this section we will briefly comment on the calculations and what we predict (within the limits of the method, highlighted by the comparisons with the experimental spectra in the previous sections) to be the results of such measurements.

	Crystal parameters variation		
	Exp. (293 K)	PBE-D	Change /%
$a / \text{\AA}$	7.53	7.38	-1.99
$b / \text{\AA}$	11.15	10.65	-4.48
$c / \text{\AA}$	13.91	13.26	-4.67
$\beta / ^\circ$	92.86	92.21	-0.69
Vol. / $\text{\AA}^3$	1165.34	1042.23	-10.55

Table 7.15. Change in the lattice parameters for the CASTEP geometry optimisation of carbamazepine form III, dispersion corrected PBE. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$ . The initial structure parameters are taken from CMBZPN10

### 7.10.1 DMACRYS calculations for paracetamol form II

The DMACRYS calculations and CASTEP calculations were performed using the HXACAN08 crystal structure at 123 K. The minimisation left the unit cell parameters and volume almost unchanged ( $F$  is less than 18 for both forcefields, see Table 7.16).

The simulated absorption spectra are dominated by a high frequency absorption, around  $140\text{ cm}^{-1}$ . The overall spectrum is therefore very flat in the experimental range from  $0\text{ cm}^{-1}$  to  $110\text{ cm}^{-1}$ ; therefore, according to these calculations, a higher ratio API/polyethylene might be necessary to appreciate any absorption peak (reported in logarithmic scale in Figure 7.26).

### 7.10.2 CASTEP calculation for paracetamol form II

We performed a dispersion corrected CASTEP calculation, starting from the HXACAN08 crystal structure. The larger contractions are in the directions of the hydrogen bonds (crystallographic plane  $bc$ ), which are decreased from respectively  $2.94\text{ \AA}$  ( $\text{NH} \cdots \text{O}$ ) and  $2.71\text{ \AA}$  ( $\text{OH} \cdots \text{O}$ ) to  $2.81\text{ \AA}$  and  $2.61\text{ \AA}$ .

The calculated spectrum is very similar in shape to the DMACRYS spectra, with a strong absorption feature at  $117\text{ cm}^{-1}$  (Figure 7.25), but there is a shift of almost  $20\text{ cm}^{-1}$  compared with the DMACRYS computed spectra, which could be due to the unit cell volume.

### 7.10.3 DMACRYS calculations for paracetamol form III

We did not calculate a CASTEP phonon spectrum of this system. The minimisation of the unit cell with DMACRYS changes slightly the unit cell axes, but leaves the total volume almost unchanged.

As we discussed in chapter 6.1, form III is structurally very similar to form II; not surprisingly, the calculated spectrum (Figure 7.25) is very similar to the calculated form II spectrum (Figure 7.27) with an absorption around  $135\text{ cm}^{-1}$  much stronger than the others. It is still possible to tell the differences with the calculated spectra of form II, particularly with the intensity of the peaks at  $120\text{ cm}^{-1}$ .

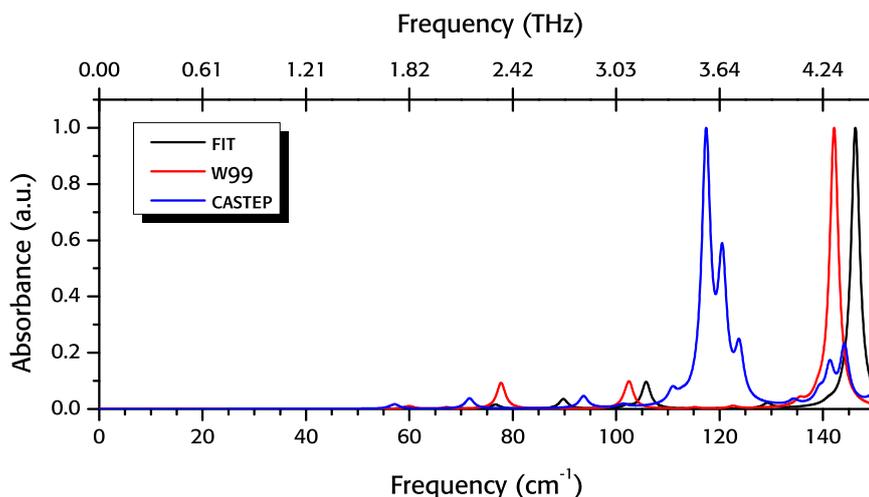


Figure 7.25. Calculated spectrum of paracetamol form II with DMACRYS, w99 and FIT forcefields, and CASTEP, PBE/DFT-D in the region  $0\text{ cm}^{-1}$  to  $150\text{ cm}^{-1}$

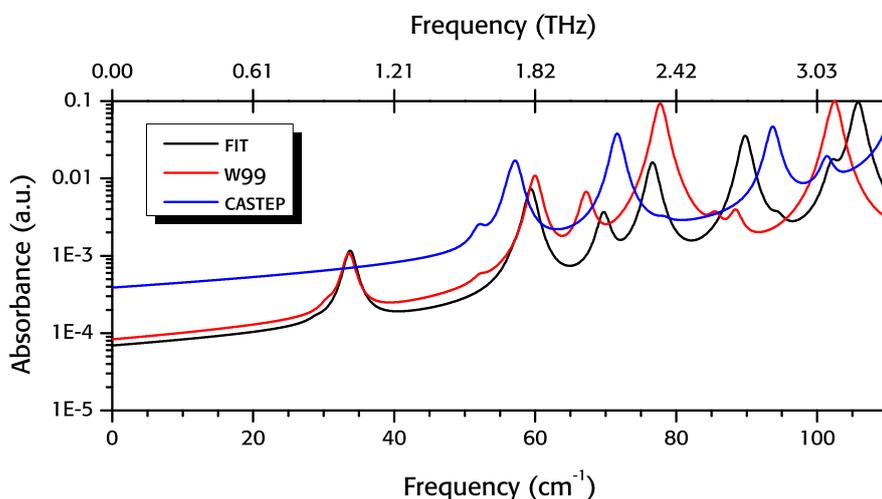


Figure 7.26. Calculated spectrum of paracetamol form II with DMACRYS, w99 and FIT forcefields, and CASTEP, PBE/DFT-D in the region  $0\text{ cm}^{-1}$  to  $110\text{ cm}^{-1}$ . The absorbance is in logarithmic scale

	Crystal parameters variation				
	Exp. (123 K)	FIT	Change/%	w99	Change/%
$a / \text{\AA}$	17.16	17.37	0.23	17.28	0.71
$b / \text{\AA}$	11.77	11.98	1.73	11.93	1.37
$c / \text{\AA}$	7.21	7.42	2.90	7.41	2.87
Vol. / $\text{\AA}^3$	1458.01	1545.37	5.99	1531.31	5.02
$F$ value		17.46		13.39	

Table 7.16. Variation of the lattice parameters and volume for the DMACRYS geometry optimisation of paracetamol form II. Symmetry constraints: unit cell angles  $\alpha = \beta = \gamma = 90^\circ$

	Experimental (123 K)	Crystal parameters variation	
		PBE-D	Change /%
$a / \text{\AA}$	17.16	17.06	-0.58
$b / \text{\AA}$	11.77	11.45	-2.71
$c / \text{\AA}$	7.21	6.85	-4.99
Volume / $\text{\AA}^3$	1458.01	1338.73	-8.23

Table 7.17. Change in the lattice parameters for the CASTEP geometry optimisation of paracetamol form II, dispersion corrected PBE. Symmetry constraints: unit cell angles  $\alpha = \beta = \gamma = 90^\circ$ . The initial structure parameters are taken from HXACAN08

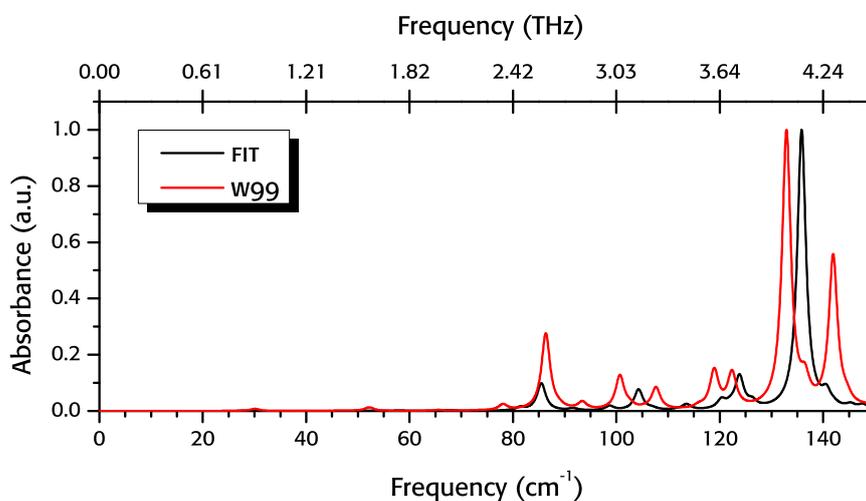


Figure 7.27. Calculated spectrum of paracetamol form III with DMACRYS, w99 and FIT forcefields in the region  $0 \text{ cm}^{-1}$  to  $150 \text{ cm}^{-1}$

	Exp. (293 K)	Crystal parameters variation			
		FIT	Change/%	w99	Change/%
$a / \text{\AA}$	11.83	12.17	2.83	12.20	3.14
$b / \text{\AA}$	8.56	8.64	0.97	8.59	0.36
$c / \text{\AA}$	14.82	14.42	-2.64	14.37	-3.01
Vol. / $\text{\AA}^3$	1501.41	1517.82	1.09	1507.38	0.39
$F$ value		38.52		42.04	

Table 7.18. Variation of the lattice parameters and volume for the DMACRYS geometry optimisation of paracetamol form III. Symmetry constraints: unit cell angles  $\alpha = \beta = \gamma = 90^\circ$

	Crystal parameters variation				
	Exp. (150 K)	FIT	Change/%	w99	Change/%
$a / \text{\AA}$	4.50	4.55	1.13	4.46	-0.03
$b / \text{\AA}$	10.53	10.78	2.36	11.02	4.57
$c / \text{\AA}$	17.05	17.37	1.91	17.26	1.26
$\beta / \text{\AA}$	96.34	94.726	2.90	94.14	-2.33
Vol. / $\text{\AA}^3$	804.17	850.93	5.81	847.60	5.40
$F$ value		26.82		47.64	

Table 7.19. Variation of the lattice parameters and volume for the DMACRYS geometry optimisation of paracetamol monohydrate. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$

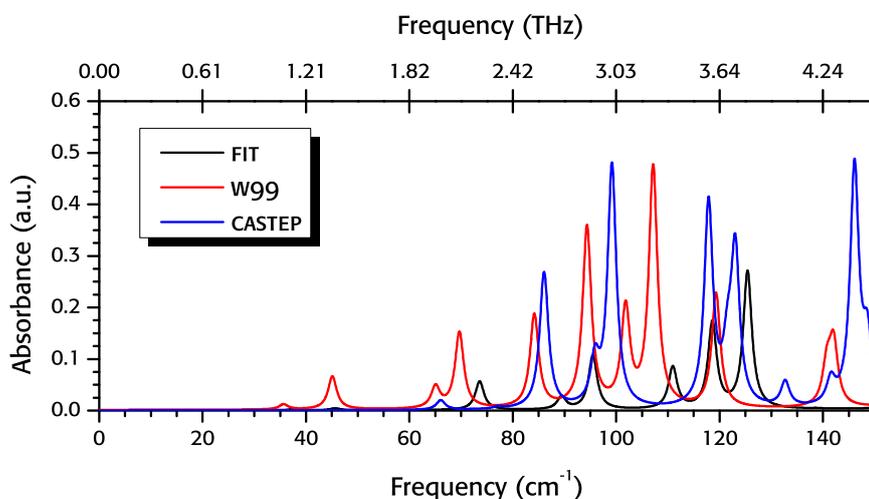


Figure 7.28. Calculated spectrum of paracetamol monohydrate with DMACRYS, w99 and FIT forcefields, and CASTEP, PBE potential

#### 7.10.4 Phonon calculations for paracetamol monohydrate

We performed both a PBE CASTEP 4.1 calculation and DMACRYS calculations (FIT and w99) on the system, taking as a starting point of the minimisation the only crystal structure in the CSD, HUMJEE.

In contrast to nitrofurantoin hydrate and theophylline hydrate, the most visible change in the crystal structure (half of the  $F$  factors in Table 7.19) is in the unit cell axes, with the positions and rotations of the water molecules being relatively unaffected.

There are a lot of similarities between the w99 and the CASTEP spectrum, both in the distribution and intensities of the absorption peaks; the FIT determination are quite different from the other two, especially for the intensity of the peaks.

	Crystal parameters variation				
	Exp. (150 K)	FIT	Change/%	w99	Change/%
$a / \text{\AA}$	7.33	7.35	0.28	7.33	0.06
$b / \text{\AA}$	12.59	12.68	0.78	12.65	0.50
$c / \text{\AA}$	22.63	22.74	0.48	22.59	-0.19
Vol. / $\text{\AA}^3$	2089.6	2122.3	1.6	2097.5	0.4
$F$ value		15.97		6.04	

Table 7.20. Variation of the lattice parameters and volume for the DMACRYS geometry optimisation of paracetamol trihydrate. Symmetry constraints: unit cell angles  $\alpha = \beta = \gamma = 90^\circ$

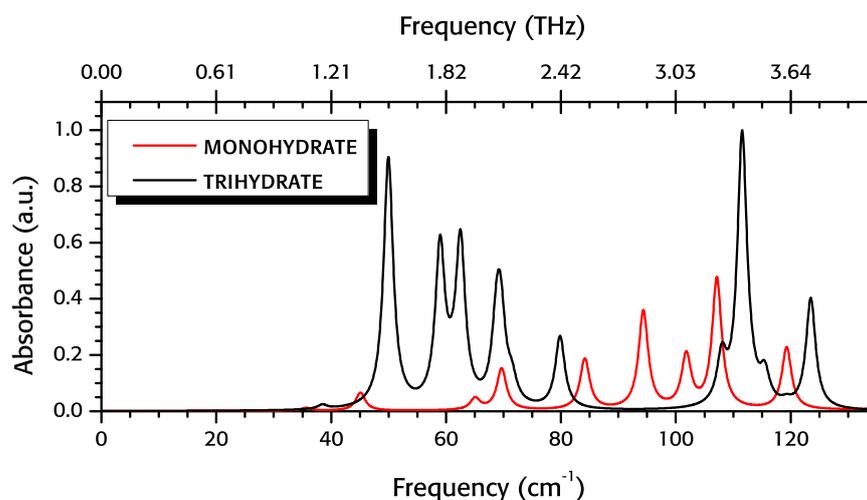


Figure 7.29. Comparison between the monohydrate and trihydrate of paracetamol with DMACRYS, w99

#### 7.10.5 Phonon calculations for paracetamol trihydrate

The minimisation of the crystal structure of paracetamol trihydrate have a surprisingly low  $F$  value (as low as 6 in the FIT calculation, Table 7.20), considering the size and complexity of the system, and the difficult determination of the hydrogen position of the waters. This might be due to the low temperature determination (150 K). By far the biggest contribution to  $F$  is due to the rotation of one of the three water molecules ( $9^\circ$  and  $4^\circ$  respectively for w99 and FIT).

We report only the comparison with the monohydrate spectrum of the w99 calculation, to show how the spectra for the two systems are very different from each other, most certainly allowing for an easy determination with an experimental measurement.

### 7.10.6 Phonon calculations for nitrofurantoin form $\alpha$

Due to a problem with the terahertz spectrometer we were not able to measure the spectrum of nitrofurantoin form  $\alpha$ , despite having the experimental sample, obtained by Bhavnita Patel. A room temperature measurement is available in literature (Figure 7.31, adapted from [209]). We can see three clear peaks at  $40\text{ cm}^{-1}$ ,  $81\text{ cm}^{-1}$  and  $99\text{ cm}^{-1}$ .

Nitrofurantoin  $\alpha$  is the smallest system we analysed, both in term of unit cell volume and molecules per unit cell (two). Of the 12 resulting lattice modes of the DMACRYS calculations, only three are IR-active, which is consistent with the number of peaks in the experimental spectrum; however, the intensities do not agree well with the experimental determinations (see Figure 7.30). In particular, the peak at  $105\text{ cm}^{-1}$  is very weak, and it is almost invisible in the FIT spectrum.

We performed two calculations with CASTEP, with and without dispersion correction. The resulting spectrum is reported in Figure 7.32. We can notice how there is a general agreement in shape, but the absorption features are shifted to higher frequencies: in the dispersion corrected calculation the lowest feature is  $20\text{ cm}^{-1}$  higher than its experimental counterpart, while the doublet at  $80\text{ cm}^{-1}$  is shifted  $10\text{ cm}^{-1}$ . Even more problematic is the shift in the non dispersion corrected calculation, with all the features shifted by  $40\text{ cm}^{-1}$  (without taking into account the spurious peak at  $60\text{ cm}^{-1}$ ).

The two CASTEP calculations have of course the same number of peaks, but their distribution is different: the lowest two active modes have almost the same frequency, while the next four are shifted in frequency by almost  $20\text{ cm}^{-1}$ .

### 7.10.7 Phonon calculations for carbamazepine anhydrous forms

The minimisation process of the anhydrous forms of carbamazepine with DMACRYS<sup>3</sup> does not considerably affect the crystal structure, as already noted by Day [60], with the  $F$  values ranging from a minimum of 1.58 (form II, FIT determination) to 12.4 (form IV, w99 determination). We therefore do not summarise the structural changes in the energy minimisation of each structure in a table.

<sup>3</sup> Excluding form V, which we did not analyse, due to its crystal structure being published only in August 2011 in the CSD

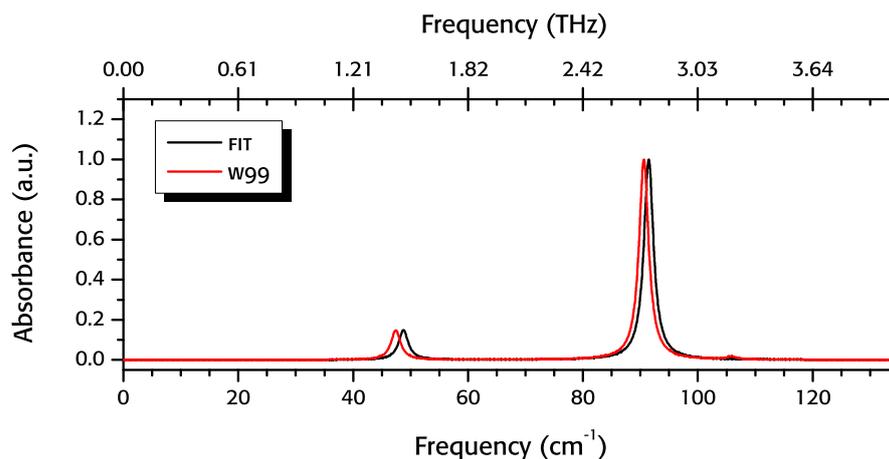


Figure 7.30. Calculated spectrum of nitrofurantoin  $\alpha$  with DMACRYS, FIT and w99 forcefield

	Exp. (293 K)	Crystal parameters variation			
		FIT	Change/%	w99	Change/%
$a / \text{\AA}$	6.77	6.94	2.46	6.92	2.27
$b / \text{\AA}$	7.79	8.07	3.60	8.07	3.58
$c / \text{\AA}$	9.80	9.74	-0.63	9.67	-1.29
$\alpha / \text{\AA}$	106.68	108.14	1.37	108.36	1.57
$\beta / \text{\AA}$	104.09	106.27	2.10	105.26	1.13
$\gamma / \text{\AA}$	92.29	92.748	0.49	91.61	-0.67
Vol. / $\text{\AA}^3$	477.56	492.57	3.14	491.95	3.01
$F$ value		33.55		30.13	

Table 7.21. Variation of the lattice parameters and volume for the DMACRYS geometry optimisation of nitrofurantoin form  $\alpha$

	Experimental (123 K)	Crystal parameters variation	
		PBE-D	Change %
$a / \text{\AA}$	6.77	6.77	0.02
$b / \text{\AA}$	7.79	7.41	-4.87
$c / \text{\AA}$	9.80	9.71	-0.91
$\alpha / \text{\AA}$	106.68	107.61	0.87
$\beta / \text{\AA}$	104.09	102.95	-1.85
$\gamma / \text{\AA}$	92.29	90.72	-1.70
Volume / $\text{\AA}^3$	477.56	451.51	-5.45

Table 7.22. Change in the lattice parameters for the CASTEP geometry optimisation of nitrofurantoin form  $\alpha$ , dispersion corrected PBE

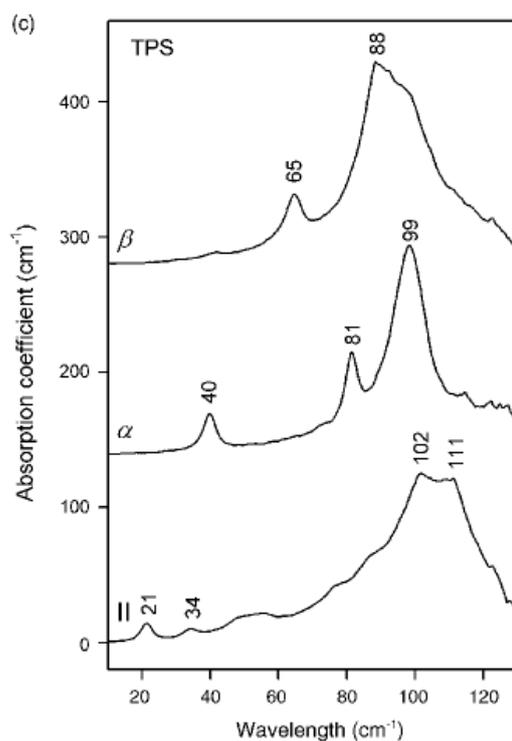


Figure 7.31. Experimental spectrum at room temperature of nitrofurantoin  $\alpha$  (middle in the graph) and its comparison with the spectra of nitrofurantoin  $\beta$  (top) and hydrate II (bottom. Adapted from [209]

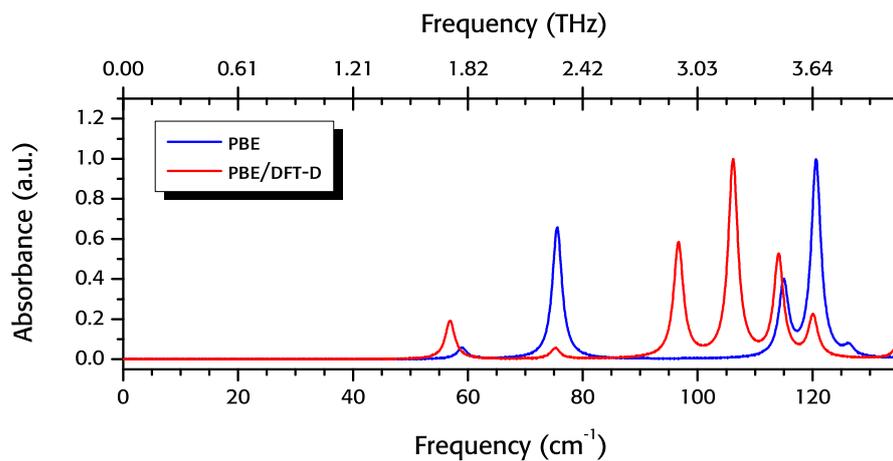


Figure 7.32. Calculated spectrum of nitrofurantoin  $\alpha$  with CASTEP, PBE and dispersion corrected PBE

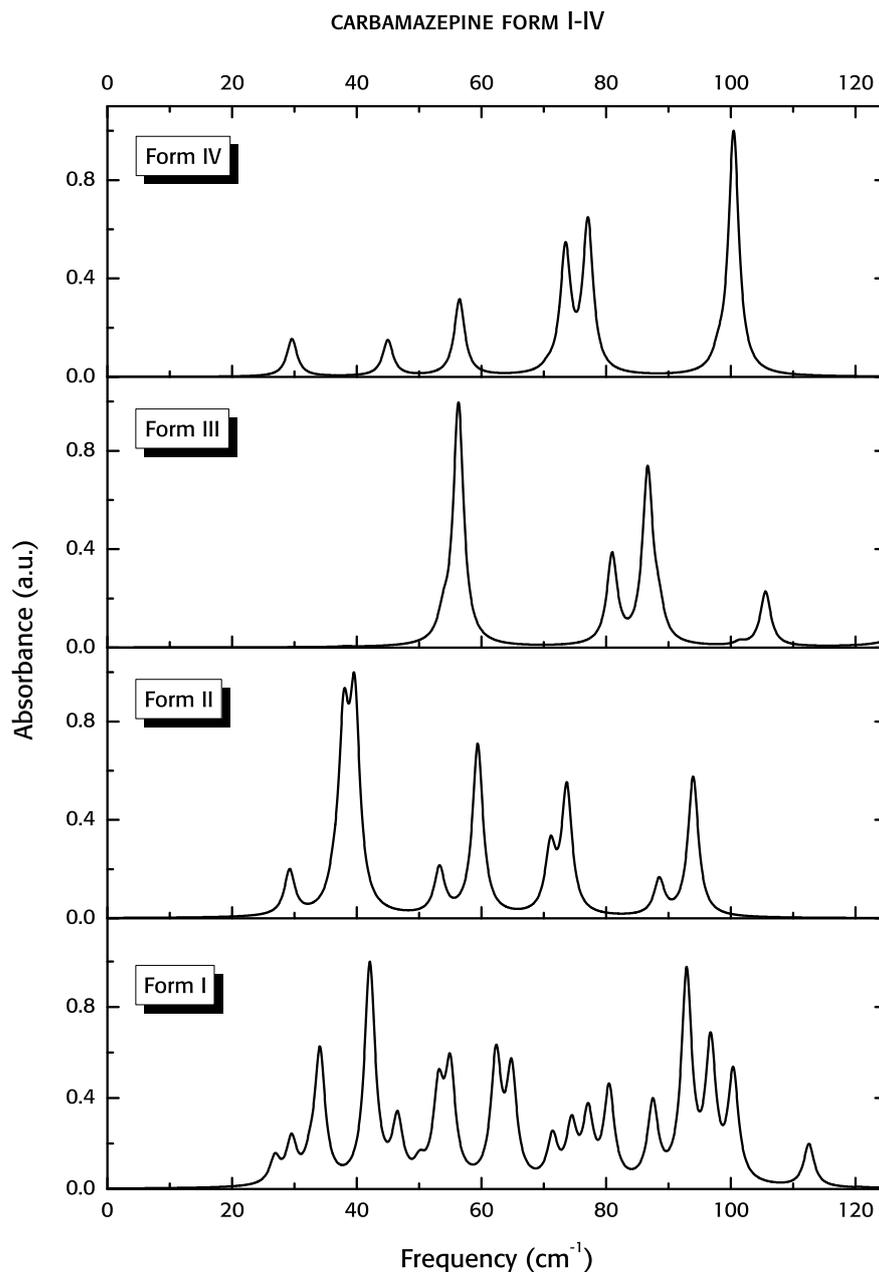


Figure 7.33. Calculated spectra of form I-IV of carbamazepine calculated with DMACRYS, w99 forcedief

We report the calculated spectra using w99 in Figure 7.33 to show the difference in the spectrum between the carbamazepine polymorphs. We can notice how each crystal form can be easily distinguished from the others by one or more characteristic absorptions in the spectrum:

## 7.11 ANALYSIS OF THE DMACRYS EIGENVECTORS

The calculated spectra presented in this section are plotted as sum of Lorentzian functions centred in the absorption frequency, with their height proportional to the square of the variation of the unit cell dipole due to the molecular displacements. The differences in the spectra using different forcefields (FIT or w99) are due to the change in absorption frequencies and the change of the eigenvectors (different dipole variation), since the electrostatic model used to describe the molecule is the same.

In Figure 7.34 we report the values of the scalar product of each vibrational eigenvector for the anhydrous crystal systems we have studied. With the exception of the highest frequency absorption modes of benzoic acid (frequency higher than  $115\text{ cm}^{-1}$  in modes 16–24, see Figure 7.34a) there is no substantial variation of the vibrational modes: the changes in the shape of the spectra between FIT and w99 are therefore to be attributed solely to the change in absorption frequencies.

The major visual difference in anhydrous theophylline, with a missing feature in the FIT calculation, is due the fact that the two strongest absorptions are almost superimposable ( $53.67\text{ cm}^{-1}$  and  $54.3\text{ cm}^{-1}$ ) and not distinguishable in our spectrum (see Table 7.5). As a consequence, not only there seem to be a missing peak, but all the other absorptions appear to be weaker (the highest point in the absorption spectra is normalised to 1).

The eigenvector comparison of the hydrate systems is discussed in chapter 9 (see page 180). There is less agreement between the two forcefields, especially for theophylline and nitrofurantoin (Figures 9.7a and 9.7d), and for these systems we can indeed notice differences in the absorption intensities of the calculated spectra.

## 7.12 CONCLUSIONS

In this chapter we reported the comparison between the experimental and calculated spectra, and the predictions for the spectra of some the forms we could not measure.

An unavoidable step in the calculation of the phonon spectrum with lattice dynamics is to reach an equilibrium position for the atomic posi-

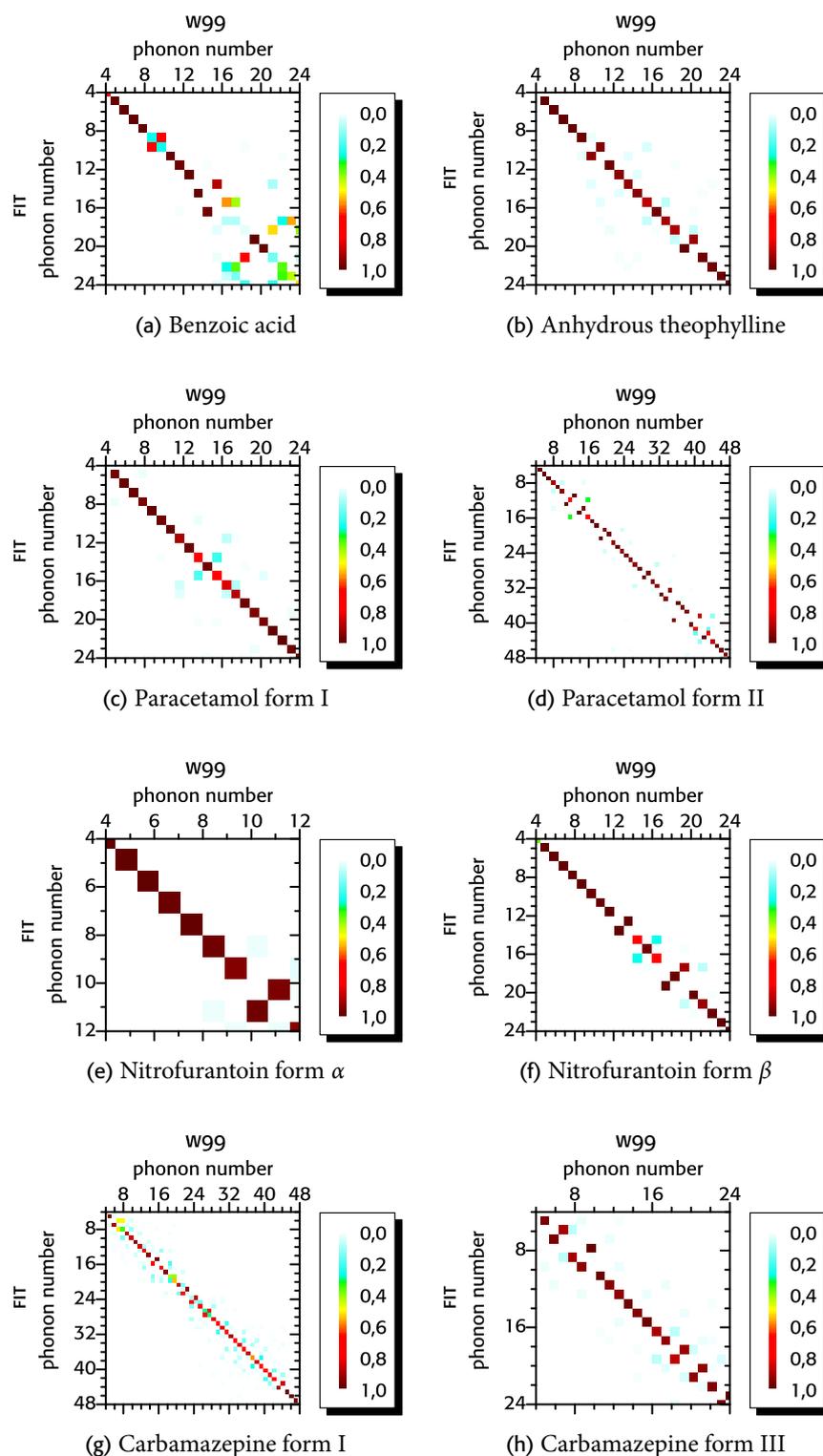


Figure 7.34. Comparison of the DMACRYS eigenvectors using FIT and w99 force-field: the colour represents the square of the magnitude of the projections of each vector upon each other

w99 forcefield		FIT forcefield	
Freq. (cm <sup>-1</sup> )	Abs. (a.u)	Freq. (cm <sup>-1</sup> )	Abs. (a.u)
24.81	0.06	23.09	0.07
26.47	0.00	25.23	0.00
29.03	0.04	27.52	0.03
54.13	1.00	53.68	1.00
62.68	0.51	54.31	0.49
62.83	0.00	57.03	0.00
66.95	0.01	61.73	0.02
72.48	0.00	68.69	0.03
74.14	0.29	73.50	0.32
83.56	0.18	75.46	0.12
90.51	0.09	81.31	0.08
110.55	0.05	97.78	0.10
112.04	0.11	98.94	0.01
124.09	0.11	106.01	0.10
175.44	0.00	175.72	0.00

Table 7.23. Absorption frequencies and intensities calculated of the IR-active DMACRYS phonons, FIT and w99 forcefield

tions, accomplished with an energy minimisation. In general, the relaxation of the unit cell axes was avoided for non dispersion-corrected DFT calculations, as it leads to unphysical expansion of the unit cell volume: Neumann and Perrin found an average 20 % expansion for a set of 25 molecular crystals DFT [210].

However, the CASTEP implementation of the dispersion correction we used is affected by the opposite problem, as it implements an apparent overcorrection: in our calculations we have an average *contraction* of 7.4 %. We can expect a contraction due to thermal expansion in the crystal, since the calculations are theoretically at 0 K and the experimental determinations are in the range 100 K to 295 K. Anyway the contractions are much larger than those predicted by the thermal expansion coefficients: from analysis of volume dependence of the crystal unit cell, it is less than 2 % per 100 K for benzoic acid and paracetamol. Furthermore, other parameterisations (as reported by van de Streek and Neumann for the DFT code VASP [211]) achieve volumes closer to the experimental values.

The volume expansion therefore suggests that the dispersion correction implemented in CASTEP does not take into account correctly the intermolecular forces in the crystals, and the frequency calculation might

be affected as well. The results of the calculations seem to agree with this observation: for the systems where we have experimental comparison available with both PBE and PBE/DFT-D (benzoic acid, paracetamol form I, nitrofurantoin  $\beta$ ) the dispersion correction does not improve the agreement. In the case of paracetamol form I, paradoxically, the excellent agreement we found without taking into account the dispersion for the peaks in the region  $40\text{ cm}^{-1}$  to  $80\text{ cm}^{-1}$  is completely lost by adding the dispersion correction.

The structural changes using DMACRYS were typically very small, especially for the non-hydrate crystals, signifying a well behaved energy minimisation. The agreement with the experimental structure was usually good, especially at lower frequencies: for some of the structures (for example anhydrous theophylline) the level of agreement was better than its CASTEP counterpart. A major limitation of the methods is its intrinsic inability to describe intramolecular interactions, which translate to the inability to correctly identify experimental features with frequency higher than  $90\text{ cm}^{-1}$ , absent or misplaced in most of the spectra.

As a general observation, the FIT forcefield performs significantly worse in the agreement with experimental position and absorption intensity with experimental peaks, as it is visible in almost all of the spectra comparisons.

The position and number of absorption frequencies in the calculations as well as in the measurement confirms the ability of terahertz spectroscopy to detect small changes in the crystal structure: this fact, an advantage in the actual analysis of the materials, can actually backfire in the computation of the spectrum, as an extreme accuracy must be employed in the choice of the starting geometry (especially with DMACRYS, where the molecular conformation, as well as the bond length and torsion angles are fixed) to ensure a correct calculation.

None of the methods we described here – DMACRYS and CASTEP – can provide a reliable calculation of the terahertz spectrum for all of our systems, and the methods cannot, as of now, be fully trusted. It is necessary to improve the methods to achieve a more accurate description of the forces in the systems. There are several possible ways to achieve this:

- Improvement of the dispersion correction in DFT, since the values provided by CASTEP are not accurate for our molecular systems. A promising method has been used by King [212], with a

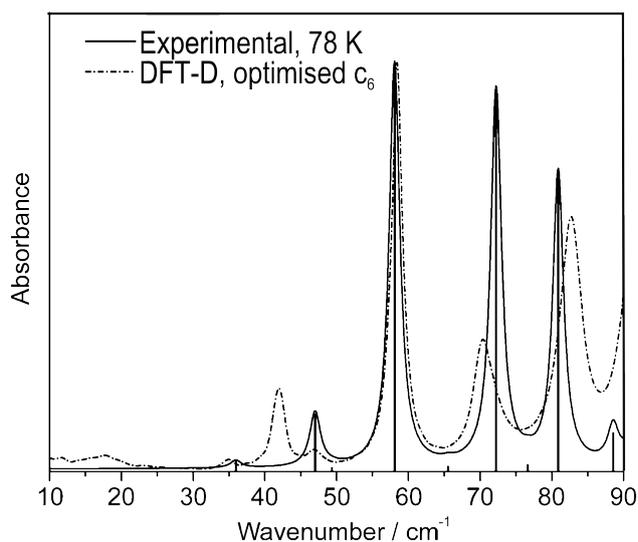


Figure 7.35. Experimental (78 K) and calculated terahertz spectrum of naproxen (Figure 7.36) with PBE/DFT-D with optimised  $c_6$  parameters. Adapted from [212]

fitting of the  $c_6$  dispersion parameters in order to achieve a better agreement of the calculated crystal lattice parameters with the experimental determinations. Such an approach (possible using the solid state DFT package CRYSTAL09 [213, 214], but not with CASTEP) achieved a variation of the unit cell of only 2.23 % when applied to the study of naproxen (Figure 7.36) and a very good agreement between the position of calculated and measured absorption peaks in the terahertz spectra of API (see Figure 7.35);

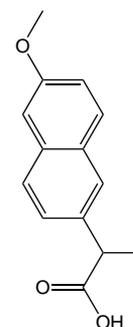


Figure 7.36.

- Calculations using DMACRYS are unreliable for frequency higher than about  $90 \text{ cm}^{-1}$ , limiting its practical applicability. One way to extend the efficacy of the method would be to introduce partial support to intramolecular motions, to achieve a more realistic description of the vibrations in the crystal (see chapter 8);
- The electronic charge density of the molecules in DMACRYS is taken from a GAUSSIAN calculation of the isolated molecule, therefore ignoring the important effect of the polarisation induced by the crystal environment. The implementation of the polarisation is discussed in chapter 10.



## EFFECT OF THE RIGIDITY OF THE SYSTEM ON THE SPECTRUM

---

**I**N THIS CHAPTER we analyse how the the rigidity of a molecule affects its terahertz spectrum. For this purpose we need to consider a working definition of what we mean by “rigid”, to establish a scale on which to compare our systems; the analysis of the amount of rigid body vibration will be performed using the program CASTEP, implementing the tools highlighted in chapter 4.4.

We briefly consider the properties of the calculated CASTEP eigenvectors along the whole spectrum of frequency 0 to  $4000\text{ cm}^{-1}$ , to recognise what distinguishes the lowest energy vibrations from the others, and we apply our analysis tools in the region 0 to  $450\text{ cm}^{-1}$ .

### 8.1 FLEXIBILITY

The flexibility of a system is a measure of how easy it is to perturb the internal conformation of a molecule: it can be investigated by modifying one of the molecular parameters (e.g. rotate a bond or modify a torsion angle) and to assess the difference in energy between the two conformations: a high energy difference means it is more difficult to change the conformation of a molecule, which is therefore more rigid.

Another way to define flexibility, more relevant to our interests, is to consider the normal modes of vibration of an isolated molecule: the eigenvalues of the mass-weighted Hessian correspond to the squared frequency (and consequently to the energy) of a particular normal mode; since all the possible vibrations of a molecule can be obtained by linear combination of its normal modes, there are no possible vibrations of smaller energy than the one of its lowest frequency normal mode.

A ranking of the flexibility of a molecule can therefore be assigned

by considering the lowest molecular frequencies. We should stress that such a calculation would not provide a reasonable guess of where to find vibrational frequencies in the solid state, because – as we mentioned in chapter 3.2.3 – the presence of intermolecular interactions can shift considerably the position of an absorption feature in the condensed state.

The normal modes of the molecules studied here are calculated using the GAUSSIAN 03 program, with the B3LYP/6-31G\*\* level of theory.

## 8.2 SYSTEMS CONSIDERED AND THEIR RIGIDITY RANKING

To analyse the effect of rigidity on the vibrational spectrum we worked with the following systems (see chapter 6 for more information): benzoic acid, theophylline, paracetamol, N-phenyl acetamide derivatives and nitrofurantoin.

The lowest vibrational eigenvectors typically involve rigid displacements of substructures of the molecules, and they are usually easily understandable from energy considerations. This is clearly visible in benzoic acid, our most rigid system: the lowest frequency normal modes – 71.0, 160.2 and 216.3  $\text{cm}^{-1}$  – relate to the movement of the carboxylic and the phenyl group with respect to each other (Figure 8.1a).

The second most rigid molecule, theophylline, consists of two fused rings, with the aromatic five-membered ring undistorted by vibrations up to the 15th normal mode, at 529.2  $\text{cm}^{-1}$ . The two lowest normal modes at 66.1 and 92.7  $\text{cm}^{-1}$  corresponds to methyl group rotations, while the third lowest vibration is a general “flapping” motion of all the functional groups bonded to the six-membered hexagonal ring (Figure 8.1b).

In the paracetamol molecule we can still distinguish between two subgroups: the acetamide group ( $\text{H}_3\text{CCOHN}-$ ) and the hydroxyphenyl: three of the lowest energy vibrations – 50.1, 79.2 and 157.9  $\text{cm}^{-1}$  – involve the relative movement of these groups (Figure 8.1c), while the lowest (at 42.8  $\text{cm}^{-1}$ ) involves the rotation of the methyl group.

In our least rigid molecule, nitrofurantoin, a -CHN- bridge connects two five-membered rings, only one of them with double bonds; unsurprisingly, the lowest energy features (24.4, 46.6, 50.6 and 126.9  $\text{cm}^{-1}$ ) involve a distortion of this bridge (Figure 8.1f).

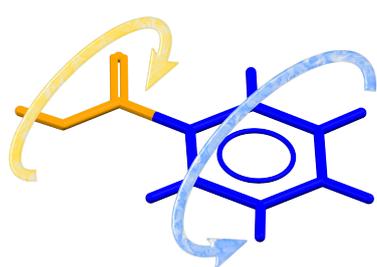
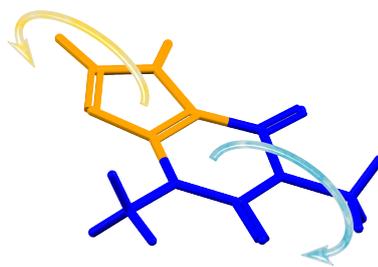
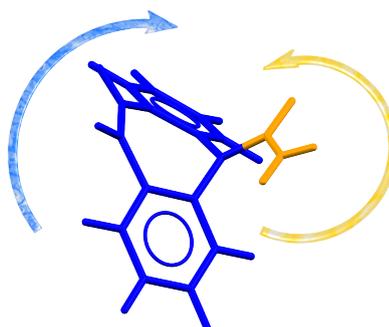
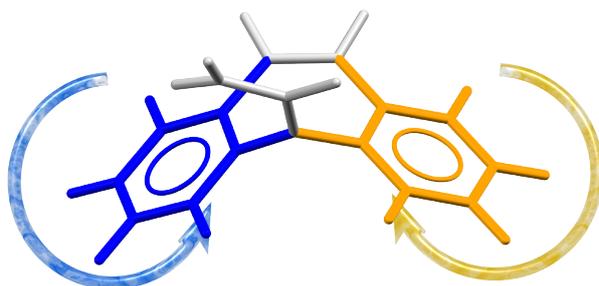
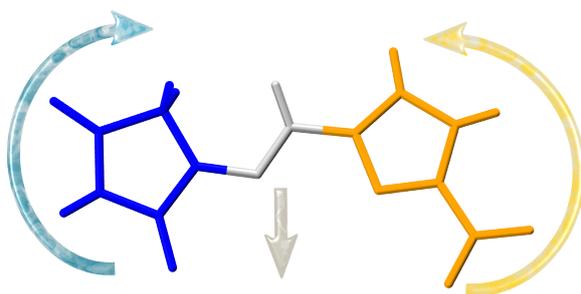
(a) Benzoic acid,  $\omega = 71.0 \text{ cm}^{-1}$ (b) Theophylline,  $\omega = 96.8 \text{ cm}^{-1}$ (c) Paracetamol,  $\omega = 50.1 \text{ cm}^{-1}$ (d) Carbamazepine,  $\omega = 55.9 \text{ cm}^{-1}$ (e) Carbamazepine,  $\omega = 65.9 \text{ cm}^{-1}$ (f) Nitrofurantoin,  $\omega = 46.6 \text{ cm}^{-1}$ 

Figure 8.1. Representative low-energy normal modes of vibration of the molecules studied in this chapter. The molecular fragments move according to the direction of the arrow of the same colour

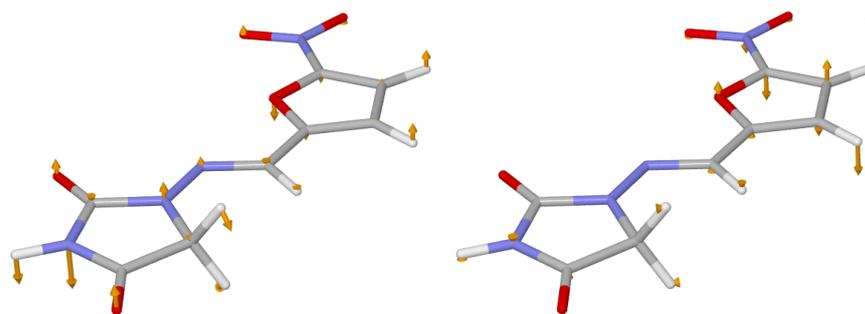


Figure 8.2. Internal vibration of the nitrofurantoin molecule calculated using B3LYP/6-31G\*\* GAUSSIAN. Left, the vibrational mode at  $144.6\text{ cm}^{-1}$ ; right, the vibrational mode at  $596.0\text{ cm}^{-1}$

Carbamazepine, the biggest molecule we studied with 30 atoms, but lighter than nitrofurantoin (236 u vs 238 u), has four different normal modes in the range from 0 to  $100\text{ cm}^{-1}$ , but they are not as low energy as those of nitrofurantoin. The lowest vibration – the rigid rotation of the carboxamide tail,  $\text{O}=\text{CNH}_2$ , see Figure 8.1d – has a frequency of  $56\text{ cm}^{-1}$ . The second lowest vibration is the “flapping” motion of the two phenyl rings at  $66\text{ cm}^{-1}$  (Figure 8.1f), while the other two vibrations are again an asymmetric flapping of the phenyl rings ( $88\text{ cm}^{-1}$ ) and a rotation of the carboxamide group ( $83\text{ cm}^{-1}$ ).

The presence of rings is not a guarantee of rigidity unless steric effects or multiple bonds constrain the vibrations: this is shown for example in nitrofurantoin: one of the rings – with double bonds reinforcing the structure – is unaffected by vibrations up to  $596.0\text{ cm}^{-1}$ , while the other shows distortion as low as  $144.6\text{ cm}^{-1}$  (see Figure 8.2).

Based on the calculated frequencies of the normal modes, we would expect some of the internal modes to mix with the lattice modes at terahertz frequencies for all of our systems. This would not be the case for more rigid systems: for example, as we mentioned in chapter 3 the lowest energy vibration of benzene is the distortion of the very stable aromatic ring, occurring at  $466.7\text{ cm}^{-1}$ .

At the other end of the flexibility spectrum, taking into account intramolecular distortions can be fundamental to correctly assign vibrations in the solid state: in macromolecular systems such as the ricin protein [215] the lowest vibrational frequency is as low as  $3.8\text{ cm}^{-1}$ , and the first 20 modes are below  $10\text{ cm}^{-1}$ .

The internal vibrations can contribute to more than one lattice mode, due to their inevitable mixing happening in the crystal. However, we

can expect the mixing between two modes to be not significant if they are well separated in energy.

There is more than one phonon for each of the internal vibrations, depending on the symmetry group of the crystal. For example, since most of the crystal structures analysed here have four molecules in the unit cell, we can expect four  $\mathbf{k} = 0$  phonons for each of the internal vibrations, one per molecule. The corresponding vibrations in the crystal will correspond to symmetry-adapted combinations of the individual molecular internal vibrations.

### 8.3 CASTEP SIMULATIONS

To understand more easily the distribution of rigid-molecule vibrations in the spectrum we have to consider all of the calculated CASTEP eigenvectors. All of the non-solvated systems studied in this chapter exhibit the same characteristic behaviour as that illustrated for paracetamol form I in Figure 8.3, where the rigid body component of each eigenvector has been analysed using the method described in section 4.4.

There is a sharp drop in the rigid body contribution as the frequency increases. All of the vibrations with wavenumbers higher than  $350\text{ cm}^{-1}$  are almost fully internal, so much that we can appreciate the differences between modes only on a logarithmic scale (see Figure 8.4).

We can notice a few regions of interest: one in the range from 0 to  $150\text{ cm}^{-1}$ , where the rigid body motion is predominant; a transition region between  $150\text{ cm}^{-1}$  and  $350\text{ cm}^{-1}$  (see the inset to Figure 8.3), where internal and external contributions mix and a higher energy region of the spectrum, with no external contribution to the vibrational modes at all.

There are many vibrational eigenvectors: their exact number is  $3N - 3$ ,  $N$  being the number of atoms in the unit cell, and therefore we have hundreds of points (80 atoms and 237 modes for paracetamol form I) in Figure 8.3 and 8.4. There are  $3N_H$  modes ( $N_H$  is the number of hydrogen in the unit cells) well separated from the others, with frequencies higher than  $2500\text{ cm}^{-1}$ ; these are all symmetry adapted stretching or bending of bonds to hydrogen atoms. The modes in the range from  $\approx 500$  to  $2000\text{ cm}^{-1}$  are internal molecular vibrations, that do not in-

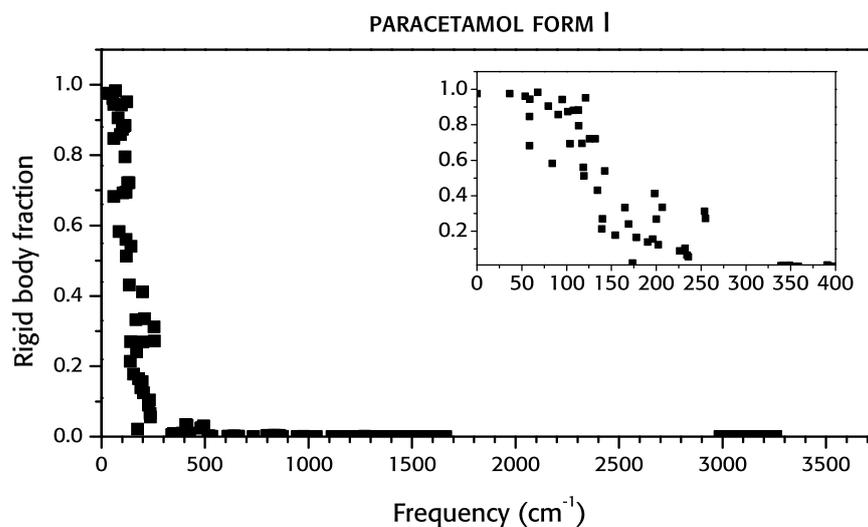


Figure 8.3. Rigid body contribution to the lattice modes calculated with CASTEP, using the dispersion corrected PBE, for paracetamol form I. The rigid body contribution on a linear scale. The small inset graph highlights the region from 0 to 400  $\text{cm}^{-1}$

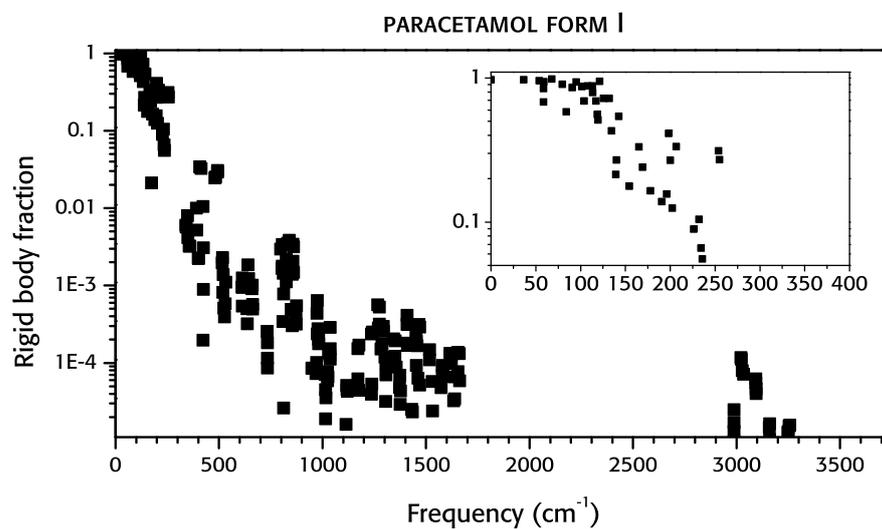


Figure 8.4. Rigid body contribution to the lattice modes calculated with CASTEP, using the dispersion corrected PBE, for paracetamol form I. The rigid body contribution on a logarithmic scale. The small inset graph, the region from 0 to 400  $\text{cm}^{-1}$

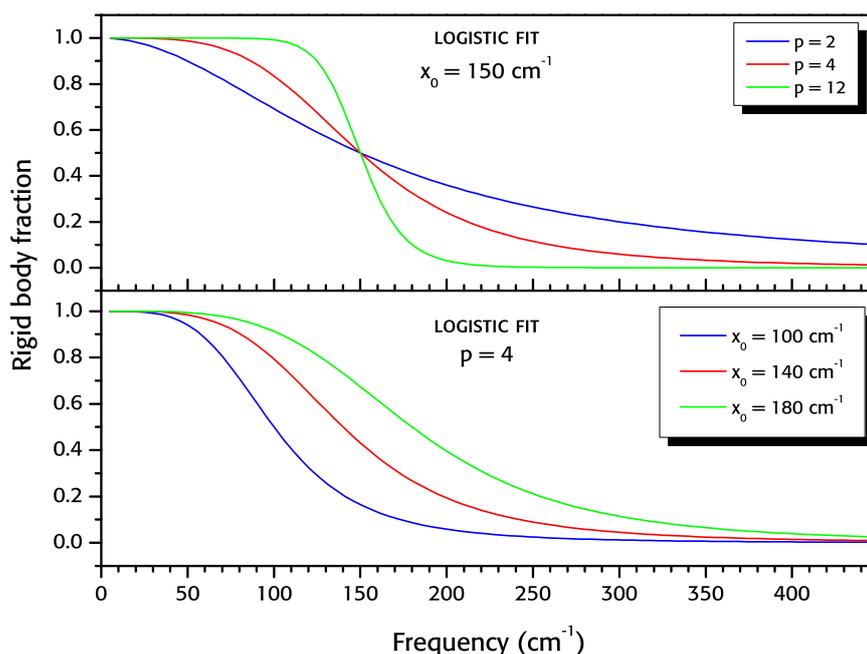


Figure 8.5. Effect of the variation of the parameters  $p$  (top) and  $x_0$  (bottom) on the logistic function in Equation 8.1

volve displacement of the centre of mass of the molecule or rigid rotation. These internal vibrations are not relevant to our discussion, and they will not be included in the subsequent discussion and graphs.

We introduce three characteristic frequencies that can be used to characterise our systems:

1.  $\omega_{\text{in}}$ , the frequency of the lowest energy mode with less than 75 % rigid molecule contribution, which we define as the first frequency where internal contributions become significant;
2.  $\omega_{\text{nrig}}$ , the frequency of the highest energy vibration with greater than 50 % rigid molecule contribution. We interpret this as the end of the rigid body section and the start of the intermediate region of the spectrum (as defined above);
3.  $\omega_{\text{high}}$ , the frequency of the highest energy vibration that has a rigid-molecule contribution of more than 10 %. We define this as the frequency that separates modes where external vibration is absent (only intramolecular motions) from the others.

Furthermore, we find that the rigid body contribution as a function of

frequency can be approximated by a sigmoidal function, with the rigid-body region separated from an internal high frequency region by an intermediate region with descending importance of rigid-body motions. We tried to fit our data using a 2-parameter logistic curve of the form

$$f_L(x) = \frac{1}{1 + (x/x_0)^p} \quad (8.1)$$

The two parameters are related to the steepness of the function  $p$  and the position of the halfway point  $x_0 : f_L(x_0) = 0.5$ .

We can see the effects of the parameters in Figure 8.5. The increase of the  $p$  parameter (top graph) reduces the width of the intermediate area, while keeping position of the halfway point constant<sup>1</sup>; the variation of  $x_0$  with  $p$  constant widens the low frequency, high rigidity region as well as the intermediate region.

The extent of the intermediate zone is related to both the parameters: the interval where the function has a value between 0.1 ( $x_{10}$ ) and 0.9 ( $x_{90}$ ) is

$$x_{10} - x_{90} = x_0 \left( 9^{1/p} - \frac{1}{9^{1/p}} \right) \quad (8.2)$$

We also evaluate the goodness of the fit, by the Pearson reduced  $\chi^2$  test and the residual sum of squares  $\zeta$  (sum of the deviations from the fitted values).

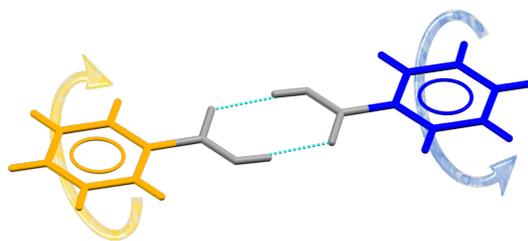
In the next sections we analyse in detail our molecular systems.

### 8.3.1 Benzoic acid

We ran two CASTEP calculations on benzoic acid, using the PBE functional, with and without dispersion correction. With both simulations the system appears to be remarkably rigid, with only four of the first 28 eigenvectors displaying partial non-rigid behaviour (Figure 8.6).

Sixteen modes show a rigidity higher than 90 %, a high value considering that ideally at most 21 modes ( $6Z - 3$  degrees of freedom, for  $Z = 4$  molecules in the unit cell) can be perfectly rigid. Most of the non-rigidity of these vibrations appears to be connected with the hydrogen bond bridge, which retain its linearity as much as possible when the dimers are distorted (Figure 8.6). In some lattice modes, this requires a twisting of the C–C connecting the acid group to the phenyl ring.

<sup>1</sup> The limit for  $p \rightarrow +\infty$  is a Heaviside-type step function centred at  $x_0$



**Figure 8.6.** Schematic representation of the vibration of a benzoic acid dimer in mode 17,  $120.89\text{ cm}^{-1}$ , calculated with dispersion corrected PBE. The molecular fragment moves according to the colour of the arrow. The grey hydrogen bonding bridge keeps its linearity

The lowest frequency internal vibrations are very close in frequency for the two CASTEP calculations: four sets of doublets can be found at  $120\text{ cm}^{-1}$  ( $110\text{ cm}^{-1}$  for the non-dispersion corrected calculation),  $150\text{ cm}^{-1}$ ,  $190\text{ cm}^{-1}$  and  $210\text{ cm}^{-1}$ . They are symmetry adapted vibrations related to the lowest frequency vibration in the isolated molecule ( $71.0\text{ cm}^{-1}$ ), which is therefore shifted by almost  $40\text{ cm}^{-1}$  for the lowest energy lattice mode, to the upper limit of the bandwidth that we can access experimentally with our terahertz spectrometer.

The frequencies  $\omega_{\text{in}}$ ,  $\omega_{\text{nrig}}$  and  $\omega_{\text{high}}$  (defined in the previous section) are summarised in Table 8.1 and the parameters of the curve fit are reported in Table 8.2. We can notice how the characteristic frequencies of the dispersion corrected calculation are approximately  $10\text{ cm}^{-1}$  higher than the non-dispersion corrected calculation. This is a sum of two effects: a shift to higher frequencies of some of the modes (most notably, the first two internal vibrations) and a larger rigid body fraction of most of the vibrations in the range from  $110$  to  $200\text{ cm}^{-1}$ .

### 8.3.2 Theophylline

Theophylline turned out to be a particularly difficult system to simulate with CASTEP without dispersion correction, as we ended up with imaginary eigenfrequencies in four independent calculations using different basis set cutoffs ( $800\text{ eV}$  and  $1200\text{ eV}$ ) and with different constraints on the space group symmetry; the imaginary frequencies tells us that the final structure is not a real energy minimum; we also conclude that the energy surface is very flat, since the geometry optimisations obeyed very tight convergence requirements.

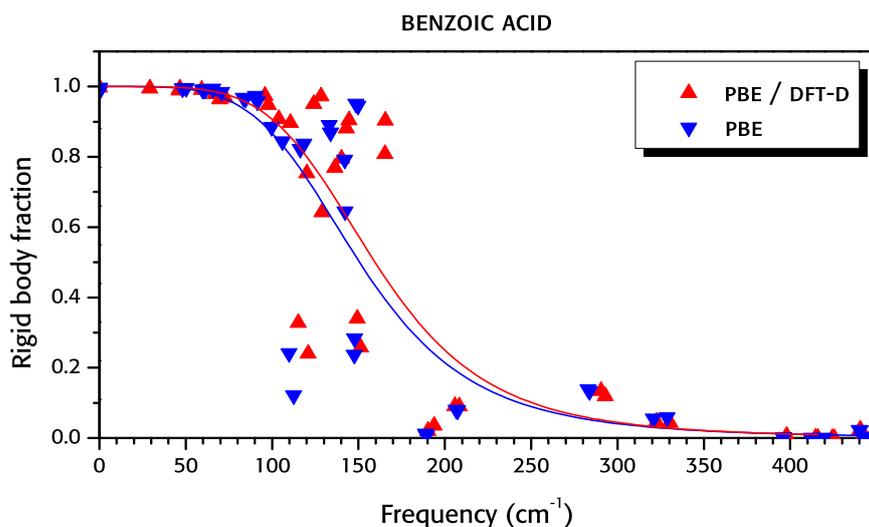


Figure 8.7. Rigid body contribution to the phonon modes of benzoic acid in the region from 0 to 450  $\text{cm}^{-1}$  as calculated with CASTEP, PBE (in blue) and dispersion corrected PBE (in red). The lines (in the same colour as the data) represent the best fit of the logistic function (Equation 8.1)

Type	Method	# mode	Rigid body fraction	$\omega$ ( $\text{cm}^{-1}$ )
$\omega_{\text{in}}$	PBE	14 of 177	0.24	109.9
	PBE/DFT-D	12 of 177	0.52	115.1
$\omega_{\text{nrig}}$	PBE	25 of 177	0.95	149.7
	PBE/DFT-D	24 of 177	0.91	165.4
$\omega_{\text{high}}$	PBE	31 of 177	0.13	283.8
	PBE/DFT-D	31 of 177	0.11	294.5

Table 8.1. Characteristic frequencies of the optical phonons of benzoic acid from a CASTEP calculation, PBE and dispersion corrected PBE

Method	$x_0$ ( $\text{cm}^{-1}$ )	$p$	$\chi^2$	$\varsigma$
PBE	150.61	4.57	0.039	1.655
PBE/DFT-D	159.53	4.86	0.042	1.493

Table 8.2. Characteristic parameters of the logistic fit of the rigid body fraction in the phonon eigenvectors of benzoic acid from a CASTEP calculation, dispersion corrected PBE

The harmonic vibration  $\omega$  relates to a periodic displacement of all the atoms in time proportional to  $\cos(\omega t)$ ; when  $\omega$  is imaginary, by the relation

$$\cos i\omega t = \frac{1}{2} (e^{\omega t} + e^{-\omega t}) \quad (8.3)$$

we do not get a periodic displacement anymore, but an acceleration in one direction, highlighting that the structure is not stable to displacement along the corresponding eigenvector.

The inclusion of the dispersion correction to the PBE functional provided us with a complete set of positive eigenvalues.

This system proved to be remarkably rigid: as depicted in Figure 8.8, the first 18 eigenvectors, completely spanning the 0 to 100  $\text{cm}^{-1}$  interval, have a rigid fraction greater than 90 %. The first partially internal vibration mode is therefore outside the terahertz region ( $\omega_{\text{nrig}} = 159.5 \text{ cm}^{-1}$ , Table 8.3), and marks the start of the transition region.

The functional fit to the data has an  $x_0$  value of 127.05, much smaller than the corresponding value for benzoic acid, as expected from the relative rigidity of the two molecules. However, the  $p$  parameter is higher: this is related to the fact that the internal modes are on the edge of the rigidity region, instead of internal modes having frequencies in the middle of the rigid region.

Similar to the benzoic acid calculation, the frequency of the lower mode of vibration in the crystal is shifted to higher frequency than in the isolated molecule by approximately 40  $\text{cm}^{-1}$ .

### 8.3.3 Paracetamol

For paracetamol we have data available from several CASTEP calculations, two for each of the two polymorphs. Figure 8.9 reports the calculated rigidity of the phonon displacements in the two polymorphs using dispersion corrected PBE.

The density of points is higher for form II as the number of phonons is larger due to the greater number of molecules in the unit cell (8, instead of the 4 of form I).

The lowest internal vibration does not correspond to the methyl group rotation (which has a low contribution is most of the non-completely rigid modes), but to the torsion about the hydroxyl group and of the acetamide relative to the phenyl group (Figure 8.1c).

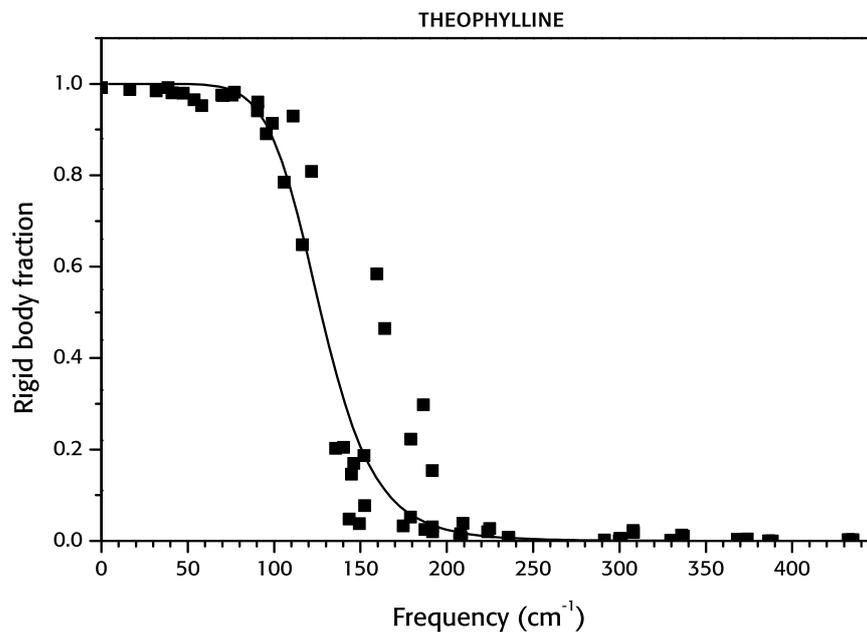


Figure 8.8. Rigid body contribution to the phonon modes of theophylline in the region from 0 to 450  $\text{cm}^{-1}$  as calculated with CASTEP, dispersion corrected PBE. The line represents the best fit of the logistic function (Equation 8.1)

Type	# mode (of 249)	Rigid body fraction	$\omega$ ( $\text{cm}^{-1}$ )
$\omega_{\text{in}}$	19	0.78	105.7
$\omega_{\text{nrig}}$	31	0.58	159.5
$\omega_{\text{high}}$	39	0.15	191.6

Table 8.3. Characteristic frequencies of the optical phonons of theophylline from a CASTEP calculation, dispersion corrected PBE

Method	$x_0$ ( $\text{cm}^{-1}$ )	$p$	$\chi^2$	$\varsigma$
PBE/DFT-D	127.05	8.13	0.014	0.67

Table 8.4. Characteristic parameters of the logistic fit of the rigid body fraction in the phonon eigenvectors of theophylline from a CASTEP calculation, dispersion corrected PBE

*Dispersion corrected calculations*

We notice a substantial similarity in the lowest frequency of the internal-only eigenvectors at approximately  $375\text{ cm}^{-1}$  for the dispersion corrected calculation of the two paracetamol polymorphs (Figure 8.9). In the rigid-body region two eigenvectors of form I (the points indicated by arrows in Figure 8.9) clearly do not fit with the general trend of the others; they are not as rigid as expected.

The logistic curves do not differ much up to  $120\text{ cm}^{-1}$ , but start to spread more visibly at higher frequency, highlighting the visibly higher number of rigid body phonon modes in form II; in form I the intermediate region is less extended, as suggested by the higher value of the exponent  $p$ . The fit of form II data is characterised by a high residual sum of squares  $\zeta$ , since there is a wide number of points that are significantly far from the curve; form I has a lower  $\zeta$ , as less points are distant from the fit.

One of the main things to observe is the persistence of rigid body motion at high frequencies ( $150$  to  $200\text{ cm}^{-1}$ ) in the form II eigenvectors.

*Non-dispersion corrected calculations*

The results of the calculations without dispersion correction, graphically expressed in Figure 8.10, confirm the general trend observed in the dispersion corrected case. One noticeable difference is the absence of the low frequency internal vibrations for form I. Instead, the general trend for these calculations is that the internal vibrations for the two forms can be found almost at comparable frequencies, less dependent on the crystal form than in the DFT-D calculations. This highlights the importance of the dispersion correction in modelling the differences in crystal packing between polymorphs in molecular systems

The logistic fits show two very different curves: the values of  $x_0$  for form I is much lower than in the dispersion corrected calculation ( $109.62\text{ cm}^{-1}$  instead of  $137.49\text{ cm}^{-1}$ ), and  $p$  is almost halved. The change in form II are not so dramatic ( $6\text{ cm}^{-1}$  for  $x_0$ ,  $0.25$  for  $p$ ).

As in the dispersion corrected calculation, the distribution is characterised by a high residual sum of squares  $\zeta$ .

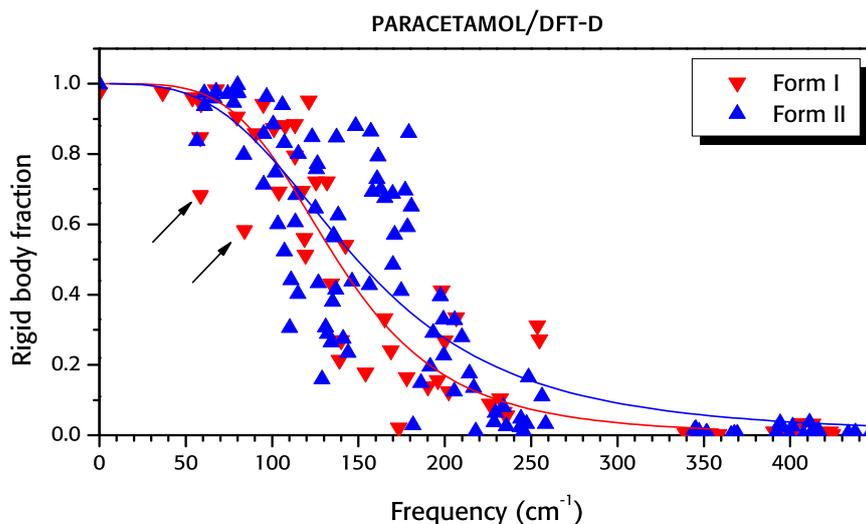


Figure 8.9. Rigid body contribution to the phonon modes of paracetamol form I (red) and paracetamol form II (blue) in the region from 0 to 450  $\text{cm}^{-1}$  as calculated with CASTEP, dispersion corrected PBE. The arrows indicate the low-frequency internal modes noted on page 147. The logistic curves are of the same colour of the data

Type	Form	# mode	Rigid body fraction	$\omega$ ( $\text{cm}^{-1}$ )
$\omega_{\text{in}}$	I	6 of 237	0.68	58.68
	II	13 of 477	0.71	95.01
$\omega_{\text{nrig}}$	I	27 of 237	0.54	142.47
	II	61 of 477	0.65	180.66
$\omega_{\text{high}}$	I	44 of 237	0.27	254.80
	II	84 of 477	0.11	256.25

Table 8.5. Characteristic frequencies of the optical phonons of paracetamol form I and II from CASTEP calculations, dispersion corrected PBE

Form	Method	$x_0$ ( $\text{cm}^{-1}$ )	$p$	$\chi^2$	$\varsigma$
Form I	PBE/DFT-D	137.49	4.23	0.020	0.927
Form II	PBE/DFT-D	148.97	3.27	0.023	3.34

Table 8.6. Characteristic parameters of the logistic fit of the rigid body fraction in the phonon eigenvectors of paracetamol form I and II calculated with CASTEP, dispersion corrected PBE

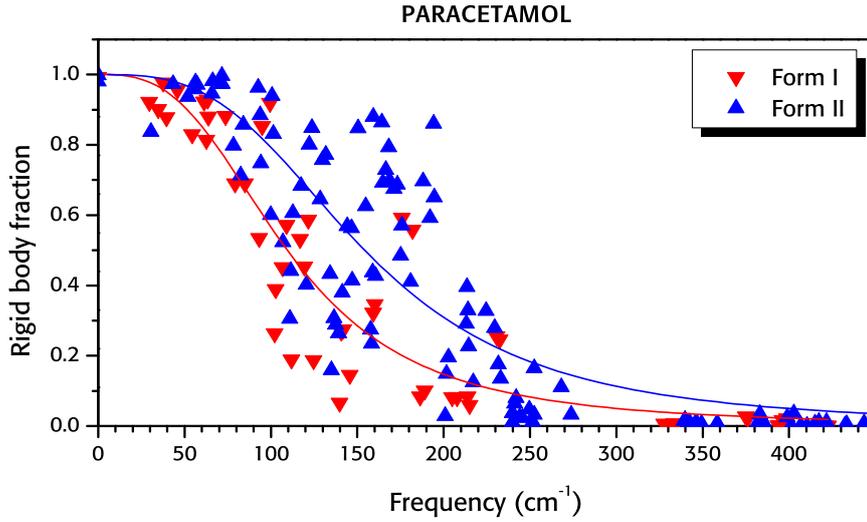


Figure 8.10. Rigid body contribution to the phonon modes of paracetamol form I (red) and paracetamol form II (blue) in the region from 0 to 450  $\text{cm}^{-1}$  as calculated with CASTEP, PBE functional. The colours of the curves correspond to the data points

Type	Form	# mode	Rigid body fraction	$\omega$ ( $\text{cm}^{-1}$ )
$\omega_{\text{in}}$	I	12 of 237	0.68	79.21
	II	13 of 477	0.71	82.58
$\omega_{\text{nrig}}$	I	33 of 237	0.56	181.92
	II	61 of 477	0.65	194.70
$\omega_{\text{high}}$	I	41 of 237	0.25	232.35
	II	83 of 477	0.16	252.50

Table 8.7. Characteristic frequencies of the optical phonons of paracetamol form I and II from CASTEP calculations, PBE functional

Form	Method	$x_0$ ( $\text{cm}^{-1}$ )	$p$	$\chi^2$	$\varsigma$
Form I	PBE	109.62	2.91	0.018	1.010
Form II	PBE	154.27	3.09	0.024	3.372

Table 8.8. Characteristic parameters of the logistic fit of the rigid body fraction in the phonon eigenvectors of paracetamol form I and II, from CASTEP calculations, PBE functional

### *Form II high frequency rigid-body fraction*

The two calculations of form II show the same trend of rigidity up to almost  $200\text{ cm}^{-1}$ ; this can probably be attributed to the ordering in parallel sheets in this polymorph, where some of the vibrations that take place in the plane of the sheets are less sterically hindered than similar modes in form I. This might reduce the coupling between inter- and intramolecular motions.

#### 8.3.4 Nitrofurantoin

The non-dispersion corrected CASTEP calculations of nitrofurantoin were among the first performed, and for these calculations we had chosen a lower energy cutoff. We simulated the stable form of anhydrous nitrofurantoin – form  $\beta$  – with an energy cutoff of 950 eV using CASTEP 4.4, and with dispersion corrected PBE, CASTEP 5.5 at a cutoff of 1200 eV. The metastable form  $\alpha$  was simulated without dispersion correction, with an energy cutoff of 800 eV using CASTEP 5.0, as well as with dispersion correction using CASTEP 5.5 with an energy cutoff of 1200 eV.

The results are summarised in Figure 8.11 and Table 8.9 for the non dispersion corrected calculations. Once again, the difference in the density of the phonons between the two forms is due to the different number of molecules in the unit cell (2 in form  $\alpha$ , 4 in form  $\beta$ ). There are several intramolecular vibrational modes around  $75\text{ cm}^{-1}$  for both of the polymorphs.

From the characteristic frequencies, the molecules in nitrofurantoin  $\alpha$  behave as the least rigid molecules among all the crystal structures we have studied. There are only three eigenvectors that exhibit a rigidity  $> 80\%$ . The other lattice modes have a considerable degree of internal contribution.

The range of rigidity is again limited up to  $110\text{ cm}^{-1}$ , while for the intermediate region we have a big difference for the two forms: a low value of  $\omega_{\text{nrig}}$  ( $120\text{ cm}^{-1}$ ) for the  $\beta$  form, but almost  $150\text{ cm}^{-1}$  for the  $\alpha$  form.

The logistic fits show that the  $\alpha$  form has a better fit than the  $\beta$  form, with half the residual sum of squares  $\varsigma$  with respect to form  $\beta$ . This is due to the almost equal amount of low-rigidity and high-rigidity modes between  $75$  and  $150\text{ cm}^{-1}$  that have to be interpolated by the fit in form

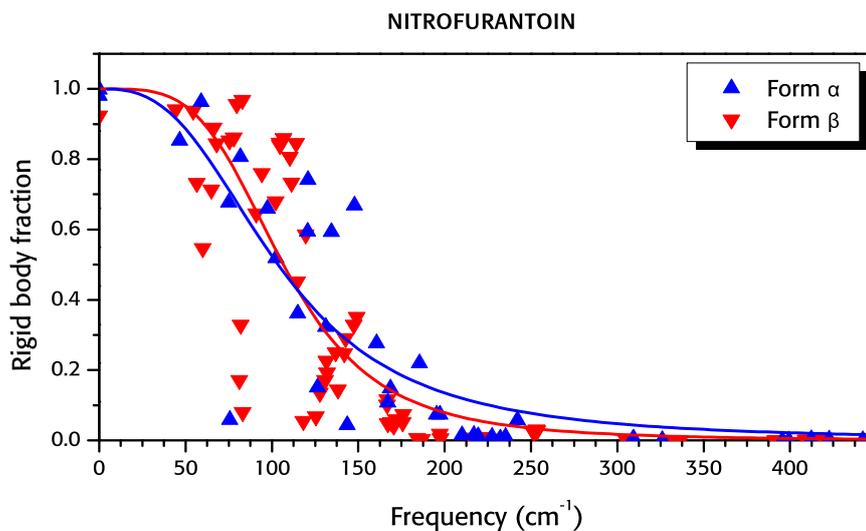


Figure 8.11. Rigid body contribution to the phonon modes of nitrofurantoin  $\alpha$  (blue) and  $\beta$  (red) in the region from 0 to 450  $\text{cm}^{-1}$ , as calculated with CASTEP, PBE functional. The logistic curves are of the same colour of the data

Type	Form	# Mode	Rigid body fraction	Frequency ( $\text{cm}^{-1}$ )
$\omega_{\text{in}}$	$\alpha$	3 of 135	0.68	75.06
	$\beta$	6 of 276	0.73	56.60
$\omega_{\text{nrig}}$	$\alpha$	15 of 135	0.66	147.86
	$\beta$	29 of 276	0.58	119.63
$\omega_{\text{high}}$	$\alpha$	19 of 135	0.22	185.41
	$\beta$	42 of 276	0.12	166.88

Table 8.9. Characteristic frequencies of the optical phonons of nitrofurantoin form  $\alpha$  and  $\beta$  from a CASTEP calculation, PBE functional

Form	Method	$x_0$ ( $\text{cm}^{-1}$ )	$p$	$\chi^2$	$\varsigma$
Form $\alpha$	PBE	103.56	2.82	0.003	1.07
Form $\beta$	PBE	106.68	3.91	0.003	2.40

Table 8.10. Characteristic parameters of the logistic fit of the rigid body fraction in the phonon eigenvectors of nitrofurantoin form  $\alpha$  and  $\beta$  from a CASTEP calculation, PBE functional

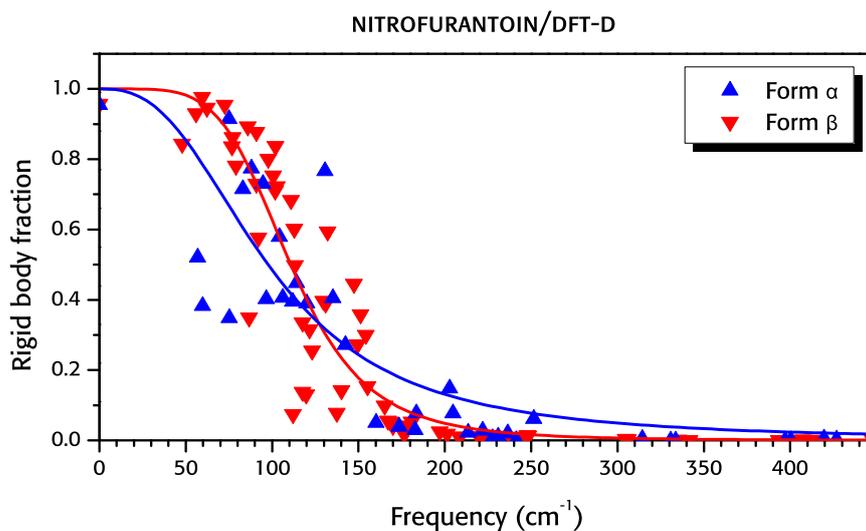


Figure 8.12. Rigid body contribution to the phonon modes of nitrofurantoin  $\alpha$  (blue) and  $\beta$  (red) in the region from 0 to 450  $\text{cm}^{-1}$ , as calculated with CASTEP, dispersion corrected PBE. The colours of the curves correspond to the data points

Type	Form	# mode	Rigid body fraction	Frequency ( $\text{cm}^{-1}$ )
$\omega_{\text{in}}$	$\alpha$	4 of 138	0.52	56.91
	$\beta$	13 of 276	0.34	86.62
$\omega_{\text{nrig}}$	$\alpha$	17 of 138	0.76	130.53
	$\beta$	33 of 276	0.59	132.13
$\omega_{\text{high}}$	$\alpha$	24 of 138	0.15	202.83
	$\beta$	40 of 276	0.15	155.23

Table 8.11. Characteristic frequencies of the optical phonons of nitrofurantoin form  $\alpha$  and  $\beta$  from CASTEP calculations, dispersion corrected PBE

Form	Method	$x_0$ ( $\text{cm}^{-1}$ )	$p$	$\chi^2$	$\varsigma$
Form $\alpha$	PBE	97.68	2.63	0.02	0.86
Form $\beta$	PBE	110.79	5.04	0.01	1.09

Table 8.12. Characteristic parameters of the logistic fit of the rigid body fraction in the phonon eigenvectors of nitrofurantoin form  $\alpha$  and  $\beta$  from CASTEP calculations, dispersion corrected PBE

$\beta$ .

The analysis of nitrofurantoin with dispersion correction gives a similar trend in both polymorphs for frequencies higher than  $150\text{ cm}^{-1}$ , but quite different in the lower energy region ( $0$  to  $100\text{ cm}^{-1}$ ). Form  $\alpha$  displays a more pronounced rigid-body behaviour, with only one of the modes being internal, at  $70\text{ cm}^{-1}$  (Figure 8.12). Conversely, form  $\beta$  displays four low energy rigid-body modes between  $55\text{ cm}^{-1}$  and  $85\text{ cm}^{-1}$ , whereas the non-dispersion corrected had only one.

The parameters of the dispersion-corrected DFT calculation are shown in Table 8.12: the result is that the residual sum of squares  $\zeta$  is much reduced (1.09 instead of 2.40). The dispersion corrected calculation of form  $\alpha$  is the only one among the systems studied to have an  $x_0$  value under  $100\text{ cm}^{-1}$ ; at the same time, the exponent  $p$  is low, and the intermediate region is very wide.

### 8.3.5 Carbamazepine

We only performed the dispersion corrected CASTEP calculation on the smallest carbamazepine polymorph, form III (the most stable form). The system appears to be fairly rigid (see Figure 8.13), with all of the modes in the range from  $0$  to  $85\text{ cm}^{-1}$  except for one displaying a rigid body fraction greater than  $0.8$ ; the logistic fit is very good, with the lowest deviation from the calculated values of all the CASTEP simulations ( $\zeta = 0.05$ ), due to the very smooth transition from the rigid body to the internal vibration region.

The parameters of the logistic fit,  $x_0$  and  $p$ , (Table 8.14) have very similar values to those for nitrofurantoin  $\beta$ , dispersion corrected calculation, with the high value of  $p$  suggesting a fast transition from the rigid body region to the internal vibration region. The similarity in behaviour with nitrofurantoin  $\beta$  is confirmed by the values of  $\omega_{\text{nrig}}$  and  $\omega_{\text{high}}$ .

### 8.3.6 Flexibility assessment – conclusions

We have analysed a series of molecular systems with increasing degree of flexibility, and tried to correlate molecular properties with the calculated rigid-molecule contribution to the eigenvectors in the terahertz region.

The frequency  $\omega_{\text{high}}$  provides information about the position of purely

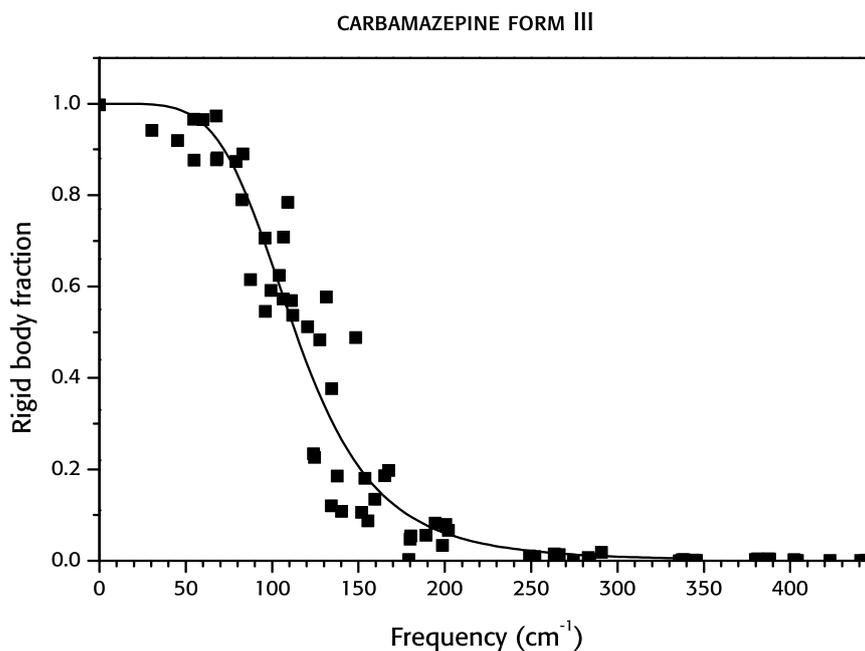


Figure 8.13. Rigid body contribution to the phonon modes of carbamazepine form III in the region from 0 to 450  $\text{cm}^{-1}$ , as calculated with CASTEP, dispersion corrected PBE

Type	# mode (of 357)	Rigid body fraction	$\omega$ ( $\text{cm}^{-1}$ )
$\omega_{\text{in}}$	15	0.61	87.49
$\omega_{\text{nrig}}$	29	0.58	131.45
$\omega_{\text{high}}$	40	0.19	167.39

Table 8.13. Characteristic frequencies of the optical phonons of carbamazepine form III from a CASTEP calculation, dispersion corrected PBE

Form	Method	$x_0$ ( $\text{cm}^{-1}$ )	$p$	$\chi^2$	$\varsigma$
III	PBE/DFT-D	113.52	4.83	0.006	0.48

Table 8.14. Characteristic parameters of the logistic fit of the rigid body fraction in the phonon eigenvectors of carbamazepine form III from a CASTEP calculation, dispersion corrected PBE

System	First internal vibration (cm <sup>-1</sup> )	$\omega_{\text{in}}$ (cm <sup>-1</sup> )	$x_0$ (cm <sup>-1</sup> )
Benzoic acid	71.0	115.1	159.5
Theophylline	66.1	105.7	127.1
Paracetamol I	50.1	58.6	139.5
Paracetamol II	50.1	95.1	149.0
Nitrofurantoin $\alpha$	24.4	56.9	103.6
Nitrofurantoin $\beta$	24.4	86.6	106.7
Carbamazepine III	55.9	87.5	113.5

Table 8.15. Characteristic frequencies of the optical phonons of the systems studied with CASTEP, dispersion corrected PBE

intramolecular vibrational modes. This frequency varies from almost 300 cm<sup>-1</sup> for benzoic acid, to 155 cm<sup>-1</sup> of nitrofurantoin. The frequencies of the lattice modes with a high fraction of rigid body rotation and translation are more sensitive to differences between polymorphs and are distributed over an even smaller range of frequencies. The upper bound of the rigid-molecule vibrations ( $\omega_{\text{nrig}}$ ) varies from 120 cm<sup>-1</sup> (benzoic acid) to 190 cm<sup>-1</sup> (nitrofurantoin form  $\alpha$ ).

We introduced the frequency of the lowest normal mode of the isolated molecule to rank the flexibility of our molecular systems: as reported in Table 8.15, we found that the frequency of the first internal mode in the crystal,  $\omega_{\text{in}}$ , is always higher than the corresponding mode of the isolated molecule. The increment is of about 30 cm<sup>-1</sup> for most systems, excluding the dispersion corrected paracetamol form I (8.6 cm<sup>-1</sup>, but 35 cm<sup>-1</sup> in the non-dispersion corrected calculation).

The logistic fits of the rigid-body fraction of the phonon modes presented in this chapter roughly follow the trend of the characteristic frequencies we just discussed. The frequency  $x_0$ , which characterises the midpoint of the transition region from rigid-molecule modes to internal vibrations, varies from 97 to 159 cm<sup>-1</sup>. Theophylline has one of the lowest values, as well as a high exponent  $p$ , making it the system with the highest separation between internal and external modes.

Based on this limited set of data, we can assume that the lowest vibrational frequency in the isolated molecule can be used as a “rule of thumb” parameter to estimate whether or not one should expect to find internal vibrations within the region accessible by our experimental technique: and – as a consequence – whether or not using a rigid-body method, such as DMACRYS, is justified to simulate vibrations in this

region.

#### *Effect of the dispersion correction*

We can analyse the effect of the dispersion interactions for the calculations where we have performed both PBE and PBE/DFT-D CASTEP calculations: we have this data for benzoic acid, paracetamol and nitrofurantoin. We analyse how different the two descriptions are with the two potentials by projecting the eigenvectors from the PBE calculations onto those from the PBE/DFT-D calculation, in a similar way to how we compared the DMACRYS/ and CASTEP eigenvectors (as described starting from chapter 4.4 to the following sections). The graphical comparison is reported in Figure 8.14: for all the systems the main deviation from the diagonal is observed both in the low energy region of the spectrum, and for the highest energy vibrations. The latter differences in the highest energy vibrations are of no particular interest to us, as they only represent different mixing of the hydrogen atoms vibrations, far from the frequencies of interest of our spectroscopic measurements.

Dispersion does not seem to have a significant effect on the vibrational eigenvectors of benzoic acid (Figures 8.14a and 8.14b): the most visible effect is the higher number of rigid-body phonons at higher frequency (in the region from 120 to 170  $\text{cm}^{-1}$ , see the deviation from the fitted curves in Figure 8.7). There is significant overlap of each DFT-D eigenvector with only one or two eigenvectors from the non-dispersion corrected calculation (Figure 8.14b). This shows that there is little mixing of modes between the two calculations.

As mentioned before, paracetamol form I exhibits a dramatic variation between DFT and DFT-D in the distribution of the points in the graphs (Figures 8.14c and 8.14d), with important changes in the degree of mixing between internal and external modes across the spectrum (Figures 8.10 and 8.9): this fact greatly affects both the fit parameters  $x_0$  and  $p$ . However, the eigenvectors are essentially the same in the two calculations, apart from a reordering of their energies (Figure 8.14c).

In the calculation of paracetamol form II we noticed again that the main effect is that some of the high frequency, high rigidity vibrational modes are shifted to higher frequencies. However, in this case the eigenvectors from the two calculations are not the same: the eigenvectors exhibit strong mixing and large deviation from the diagonal (Figure 8.14d).

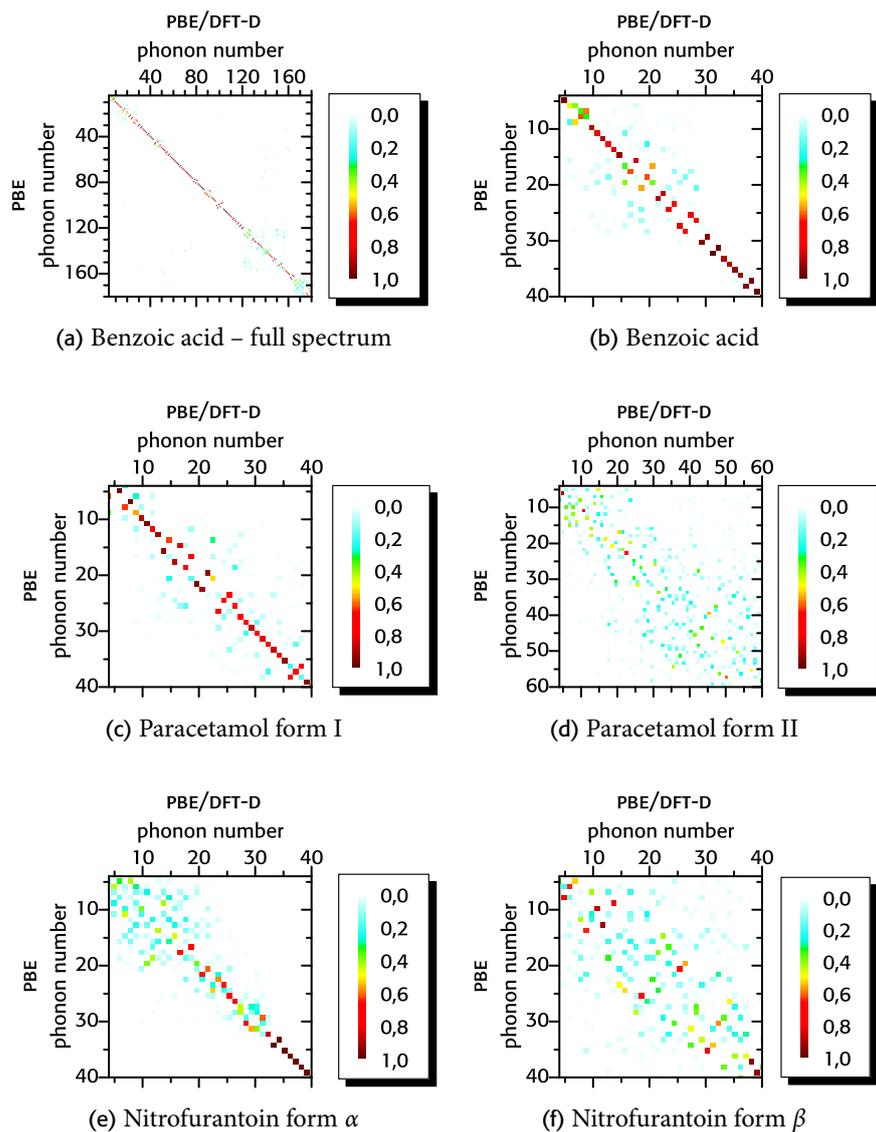


Figure 8.14. Comparison of the CASTEP eigenvectors, PBE functional with and without dispersion correction: the colour represents the square of the magnitude of the projections of each vector upon each other for the two calculations

For paracetamol form II a detailed comparison of the eigenvectors is necessary to understand the impact of the dispersion correction, which was not evident from the frequency distribution of the rigid-molecule contributions

In nitrofurantoin form  $\alpha$  there is a big change in the rigid body fraction for some of the lowest energy eigenvectors. There are only approximately 20 phonon vectors with a rigid body contribution greater than 0.1 ( $\omega_{\text{high}}$ , Tables 8.9 and 8.11). It is these modes that have a rigid-

molecule component that is most strongly influenced by the inclusion of dispersion correction in the DFT calculation (Figure 8.14e).

The eigenvectors of the dispersion corrected calculation of nitrofurantoin form  $\beta$  display more rigidity at lower frequency, as shown by the values of  $x_0$  and  $p$ ; the position of the lowest internal vibrations ( $\omega_{in}$ ) is found at higher frequency ( $56.6 \text{ cm}^{-1}$  vs  $86.6 \text{ cm}^{-1}$ ), and it displays a better agreement with the logistic fit, which demonstrates a smoother transition between rigid-molecule and internal modes.

## 8.4 EIGENVECTOR AGREEMENT BETWEEN CASTEP AND DMACRYS

In this section we analyse the agreement between the lowest energy eigenvectors calculated with DMACRYS and CASTEP.

As described in section 4.4 (Equation 4.4), we developed a method to compare the agreement between the two computational methods by expressing the rigid molecule contribution to the CASTEP eigenvectors  $|c_i^{\text{rigid}}\rangle$  as a linear combination of DMACRYS eigenvectors  $|d_j\rangle$ :

$$|c_i^{\text{rigid}}\rangle = \sum_j a_j |d_j\rangle \quad (8.4)$$

We can understand the mechanism of mixing by considering a model two level system, with two non-degenerate vibrational eigenvectors  $|v_1\rangle$  and  $|v_2\rangle$  and eigenvalues  $\varepsilon_1$  and  $\varepsilon_2$ . In the eigenvector basis set the vibrational Hamiltonian  $\widehat{H}_0$  is the matrix

$$\widehat{H}_0 = \begin{pmatrix} \varepsilon_1 & 0 \\ 0 & \varepsilon_2 \end{pmatrix} \quad (8.5)$$

which is diagonal due to the orthonormality of the eigenvectors. We can now introduce a perturbation to account for the differences in the treatment of forces between the two methods, in the form of a matrix  $\widehat{W}$ ,

$$\widehat{W} = \begin{pmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{pmatrix}. \quad (8.6)$$

The diagonal elements of  $\widehat{W}$  act as a shift of the eigenvalues of  $|v_1\rangle$  and

$|v_2\rangle$ , but they cannot modify the eigenvectors: we will now consider a simpler matrix by considering  $W_{11} = W_{22} = 0$  in the preceding formulas. The eigenvalues  $\varepsilon_{\pm}$  of the Hamiltonian  $\widehat{H}_1 = \widehat{H}_0 + \widehat{W}$  can be found by diagonalisation of the matrix  $\widehat{H}_1$ ,

$$\widehat{H}_1 = \begin{pmatrix} \varepsilon_1 & W_{12} \\ W_{12} & \varepsilon_2 \end{pmatrix}, \quad (8.7)$$

and they take the form

$$\varepsilon_{\pm} = \frac{\varepsilon_1 + \varepsilon_2}{2} \pm \sqrt{\left(\frac{\varepsilon_1 - \varepsilon_2}{2}\right)^2 + W_{12}^2} \quad (8.8)$$

The eigenvectors  $|v_+\rangle$  and  $|v_-\rangle$  are

$$\begin{aligned} |v_-\rangle &= \cos \frac{\theta}{2} |v_1\rangle + \sin \frac{\theta}{2} |v_2\rangle \\ |v_+\rangle &= -\sin \frac{\theta}{2} |v_1\rangle + \cos \frac{\theta}{2} |v_2\rangle \end{aligned} \quad (8.9)$$

where

$$\tan \theta = \frac{2W_{12}}{\varepsilon_1 - \varepsilon_2} \quad (8.10)$$

In this case, depending on the strength of the parameter  $\theta$ , we have coupling and intermixing of the two eigenvectors. If the interaction is small we can expect the eigenvectors to be unchanged.

The effect of the off-diagonal matrix elements is also an increased separation between the energy levels: this is true also if the two eigenvectors are degenerate in energy.

We can see this kind of effect in the calculations of benzoic acid in Figure 8.15: there is a strong 1:1 match between the first 12 eigenvectors for DMACRYS and the non-dispersion corrected CASTEP calculation, except for eigenvectors 5 and 6 of the CASTEP result, which are nearly perfectly rigid but do not correspond to a single DMACRYS vector.

If we look at the details of the DMACRYS contribution, we see that for these two calculated CASTEP phonons (frequency 44 and 48  $\text{cm}^{-1}$ ) the only significant contribution comes from two DMACRYS eigenvectors, respectively at 25 and 48  $\text{cm}^{-1}$ . We can therefore hypothesise that there is an interaction that is not properly taken into consideration by one of the two methods: this is confirmed by the split in energy between the

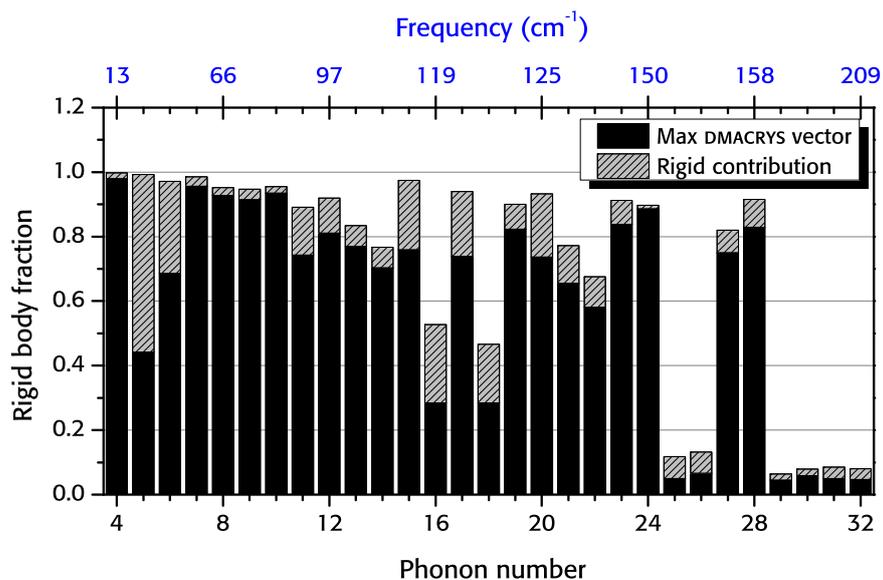


Figure 8.15. Rigid body contribution to the first 32 benzoic acid phonon eigenvectors calculated by CASTEP using the PBE functional (excluding the 3 acoustic phonons). The solid portion of each bar shows the contribution from the most representative DMACRYS eigenvector

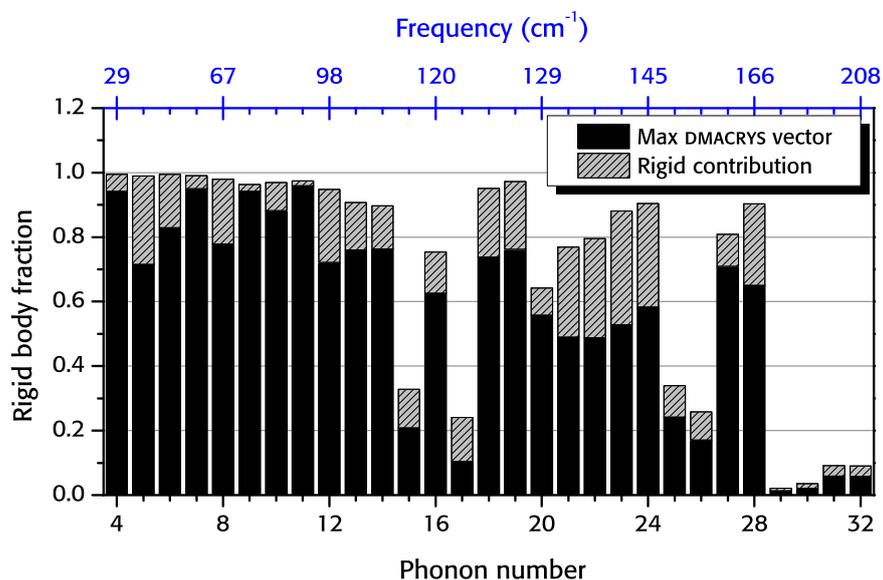


Figure 8.16. Rigid body contribution to the first 32 benzoic acid phonon eigenvectors calculated by CASTEP using the dispersion corrected PBE functional (excluding the 3 acoustic phonons). The solid portion of each bar shows the contribution from the most representative DMACRYS eigenvector

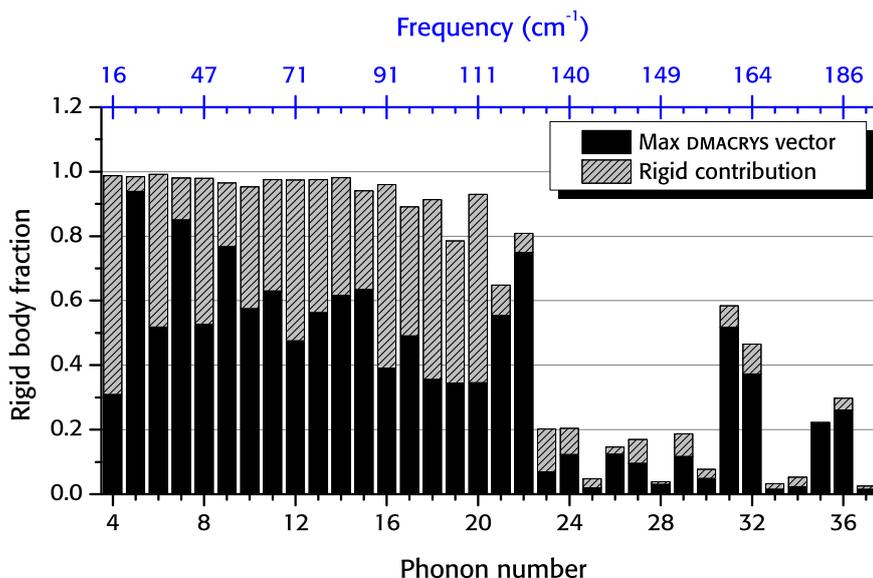


Figure 8.17. Rigid body contribution to the first 37 theophylline phonon eigenvectors calculated by CASTEP using the dispersion corrected PBE functional (excluding the 3 acoustic phonons). The solid portion of each bar shows the contribution from the most representative DMACRYS eigenvector

phonons in DMACRYS, while they are almost degenerate in CASTEP.

This discrepancy between the two methods is much less evident when dispersion correction is included in the calculation, which shows a greater consistency between the two methods for all the modes reported (Figure 8.16). These results suggest that the differences in the two modes described above were due to the lack of dispersion interactions in the pure DFT CASTEP calculation.

Another effect must be surely the intermixing between rigid-molecule phonons and internal vibrations, which becomes important if the intermolecular forces in the solid have comparable strength to the intramolecular interactions.

It is more difficult to find the same level of agreement as seen for benzoic acid in the other systems studied. The comparison between the DFT-D CASTEP calculation of theophylline with its DMACRYS counterpart (see Figure 8.17) is not particularly good; while the DFT-D eigenvectors have a large rigid-molecule component up to the 22nd eigenvector, few of these correspond in a 1:1 manner to a calculated DMACRYS eigenvector. A problem with this CASTEP calculation is the imperfect separation between acoustic and optical phonons, which is an indicator of

problems in the convergence of the DFT-D calculation. As an example, the first optical phonon, with a frequency of  $16.33 \text{ cm}^{-1}$ , has a staggering 30 % contribution from one of the DMACRYS acoustic phonons, which makes it the most representative rigid-body contribution!

This suggests that there is a mixing of the acoustic modes with the optical modes in the CASTEP calculation, which may be due to inadequate  $k$ -points sampling or too small a basis set, resulting in imperfect translational invariance of the crystal energy. However, a second calculation with increased  $k$ -point sampling ( $0.02 \text{ \AA}^{-1}$  instead of  $0.05 \text{ \AA}^{-1}$  density sampling, with an increase from 12 to 60  $k$ -points) gives the same results, with a maximum variation of  $0.2 \text{ cm}^{-1}$  in the phonon frequencies, and almost identical DMACRYS-vectors decomposition.

Another important factor is that the intermixing outlined in Equation 8.4 is possible only between eigenvectors belonging to the same symmetry group. The case of paracetamol form II (Figure 8.19) shows how symmetry affects the intermixing of modes; the phonons belong to one of the 8 irreducible representations of the crystal's space group symmetry ( $Pbca$ ). The phonons belonging to each representation are not equally spaced in energy: for example, the two lowest energy DMACRYS phonons belonging to the group  $A_u$  have frequencies  $21.39 \text{ cm}^{-1}$  and  $71.14 \text{ cm}^{-1}$  respectively, while the lowest  $B_{2g}$  frequencies are  $53.74 \text{ cm}^{-1}$  and  $57.41 \text{ cm}^{-1}$ .

The mixing is greatest between eigenvectors of similar energy (Equation 8.10), since our two computational methods are not likely to disagree significantly on the evaluation of forces. So while there is more than 97 % agreement in the well separated  $A_u$  CASTEP modes (number 8 and 14 in Figure 8.19), modes  $B_{2g}$  are intermixing with one another (number 11 and 18 in Figure 8.19). The same kind of behaviour is also present at higher frequencies.

At the same time, a good agreement between the computational methods, along with separation in energy, is seen in the overall mixing between the modes, with low energy DMACRYS vectors mixing only with low energy CASTEP vectors and vice versa. This is reported again for paracetamol form II,  $A_g$  representation, in Table 8.16.

It can be noted that the values reported are not only expansion coefficients of CASTEP vectors in terms of DMACRYS vectors, but also vice versa; so, while the well-separated DMACRYS phonon 27 is a major component of CASTEP vector 32, vector 27 itself (Table 8.16) is a sum of sev-

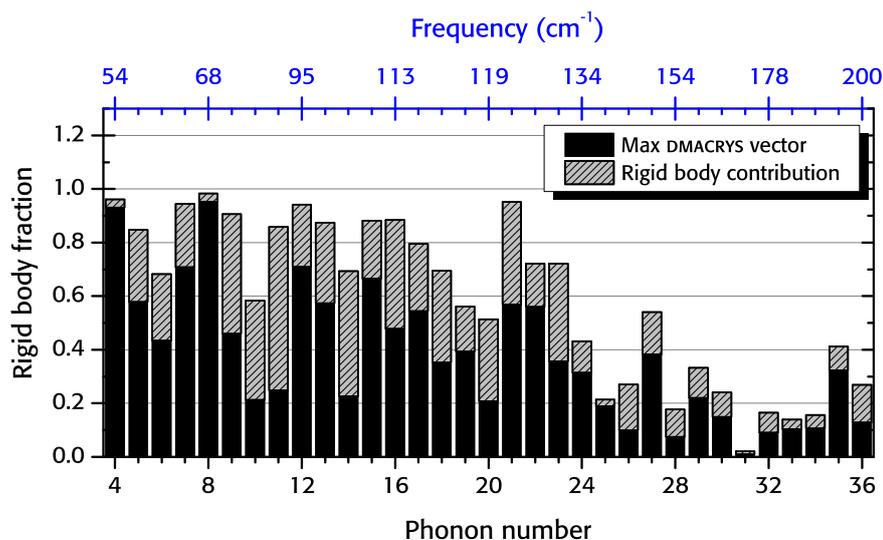


Figure 8.18. Rigid body contribution to the first 36 paracetamol form I phonon eigenvectors calculated by CASTEP using the dispersion corrected PBE functional (excluding the 3 acoustic phonons). The solid portion of each bar shows the contribution from the most representative DMCARYS eigenvector

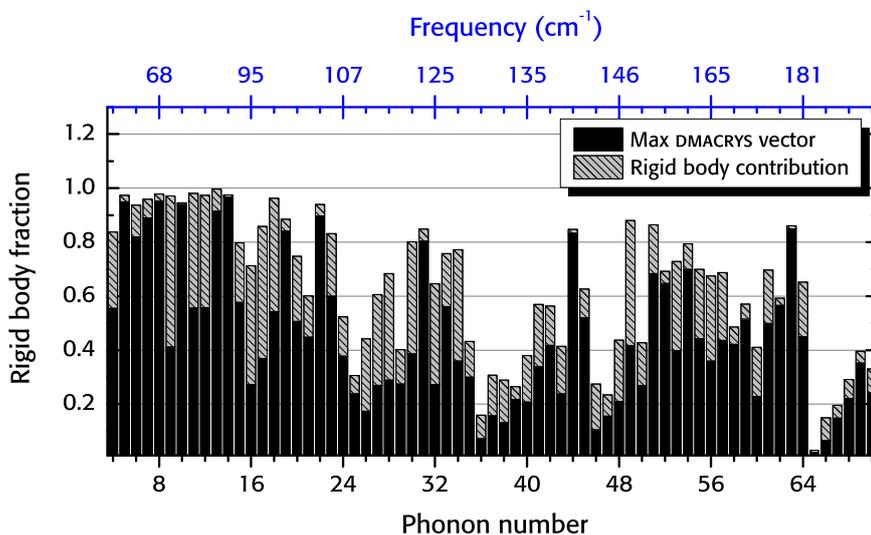


Figure 8.19. Rigid body contribution to the first 70 paracetamol form II phonon eigenvectors calculated by CASTEP using the dispersion corrected PBE functional (excluding the 3 acoustic phonons). The solid portion of each bar shows the contribution from the most representative DMCARYS eigenvector

		DMACRYS phonon mode					
		7	9	13	27	41	42
CASTEP phonon mode	4	0.56	0.00	0.19	0.06	0.01	0.02
	9	0.18	0.41	0.37	0.00	0.00	0.01
	12	0.08	0.56	0.32	0.01	0.00	0.00
	24	0.08	0.01	0.00	0.38	0.00	0.05
	32	0.05	0.00	0.09	0.20	0.03	0.27
	43	0.04	0.00	0.01	0.11	0.24	0.01
	56	0.00	0.00	0.01	0.08	0.22	0.36
	60	0.01	0.00	0.00	0.11	0.23	0.06
73	0.00	0.00	0.01	0.05	0.07	0.20	

Table 8.16. Squared coefficient of expansion of CASTEP vector in term of DMACRYS vectors,  $A_g$  representation phonon modes for paracetamol form II

eral CASTEP vectors of similar energies.

In paracetamol form I (Figure 8.18) the situation is more complex. There are a few modes with a contribution greater than 80 %, and for many of the eigenvectors there is a clearly predominant contribution, but other DMACRYS eigenvectors are important as well. The same behaviour is present in the non-dispersion corrected calculation.

Due to its flexibility, nitrofurantoin is the system expected to show the smallest agreement between CASTEP and DMACRYS among those studied. However, this appears to be true only for the  $\alpha$  form. Once again, the apparent scarceness of phonons in the  $\alpha$  form is due to its different space group symmetry ( $P-1$ , with just two molecules in the unit cell, compared with four in form  $\beta$ , symmetry  $P2_1/n$ ).

The most evident feature in the results of the dispersion corrected calculation on nitrofurantoin  $\alpha$  (Figure 8.22) is the lack of completely rigid eigenvectors in form  $\alpha$ ; only one mode shows greater than 80 % contribution for rigid-molecule degrees of freedom. In contrast, in form  $\beta$  (Figure 8.23) many of the low frequency modes show a fairly rigid-molecule behaviour, despite a very similar arrangement of molecules, in almost equally spaced molecular sheets.

It is difficult to understand why there should be such a big difference in the behaviour of the two similar polymorphs: one major difference is in the arrangement of the hydrogen bonding network within the molecular layers, and in the slightly different stacking of the molecules (see Figure 8.20).

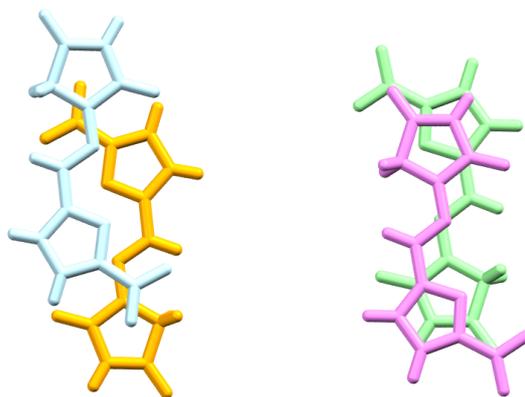


Figure 8.20. Molecular stacking of two nitrofurantoin molecules in form  $\alpha$  (right) and  $\beta$  (left)

The main problem with a rationalisation of the molecular vibrations is the marked difference between the eigenvectors resulting from the DFT and DFT-D CASTEP calculations, which is much more pronounced than for other molecules studied. This difference is due to the different evaluation of the dispersion interactions in the crystal, and the layered structure of the system.

The results of the DFT-D calculation show that the structures prefer to keep the molecular stacking as it is during the lowest vibrations, sacrificing instead the internal rigidity. However, in the non-dispersion corrected calculation the lowest vibrations consist of an almost rigid sliding of the molecular layers upon each other. We can see how in the lowest energy phonon in the non-dispersion corrected calculation (Figure 8.21a) the molecules rigidly slide upon each other, and all the molecules within a plane moves in the same direction. In the dispersion corrected calculation (Figure 8.21b) the two five-membered rings in each nitrofurantoin rotate in opposite directions: stacked rings rotate in the same direction to maintain their parallel orientation.

For carbamazepine (Figure 8.24) the agreement between DMACRYS and CASTEP is very good, despite the size of the molecule: most of the lowest frequency phonon vectors have a predominant DMACRYS mode (more than 75 % of contribution).

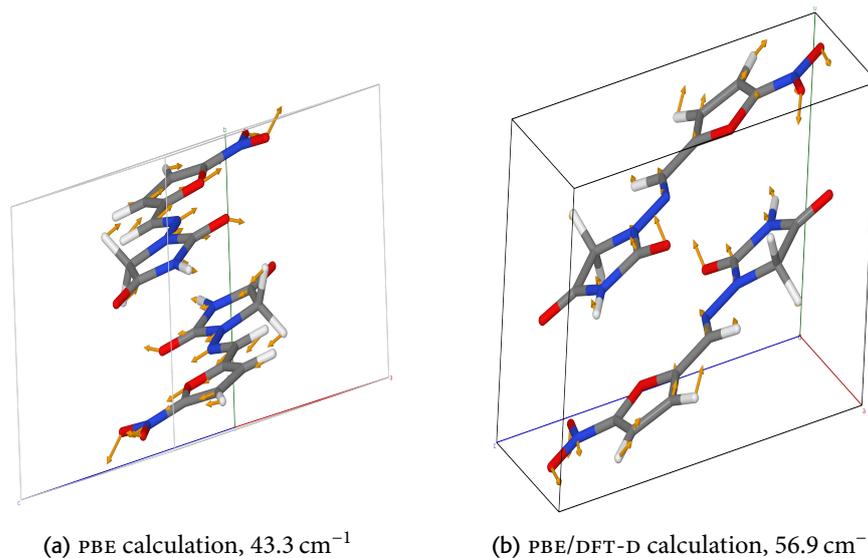


Figure 8.21. Unit cell view of the lowest optical mode of nitrofurantoin  $\alpha$  calculated with CASTEP. On the left, a rigid-molecule sliding of two molecular planes each other; on the right (dispersion corrected calculation), the internal vibrations of the two molecules. The point of view is slightly different for clarity

#### 8.4.1 Eigenvectors agreement – conclusions

We can take into account the general agreement between a DMACRYS and a CASTEP phonon calculation by summing up all the most important rigid contributions for each mode. We can get a comparable parameter among our calculations if we normalise this sum by all of the rigid body contributions, which depends on the number of molecules  $Z$  in the unit cell as  $6Z - 3$ .

This number cannot give us a precise picture, as it does not take into account the coupling between single vibrational modes. However, we can use this as a rough estimate of the CASTEP/DMACRYS agreement across the set of molecules; the results are presented in Table 8.17.

We can see how in benzoic acid we observe the highest agreement between the two computational methods. For paracetamol there is a very good agreement as well (slightly less for the dispersion corrected calculation of form I); a similar level is also found for the substituted acetanilides (see chapter 11), which are molecularly very similar to paracetamol. The value drops significantly for the non-dispersion corrected calculations of nitrofurantoin, as discussed before. The agreement between the methods seems to be dependent on the rigidity of the mo-

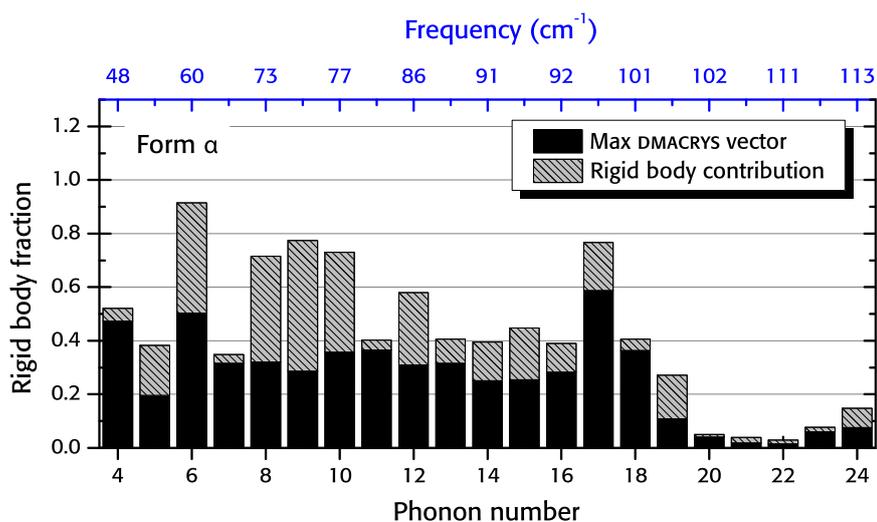


Figure 8.22. Rigid body contribution to the first 24 nitrofurantoin  $\alpha$  phonon eigenvectors calculated by CASTEP using the dispersion corrected PBE functional (excluding the 3 acoustic phonons). The solid portion of each bar shows the contribution from the most representative DMACRYS eigenvector

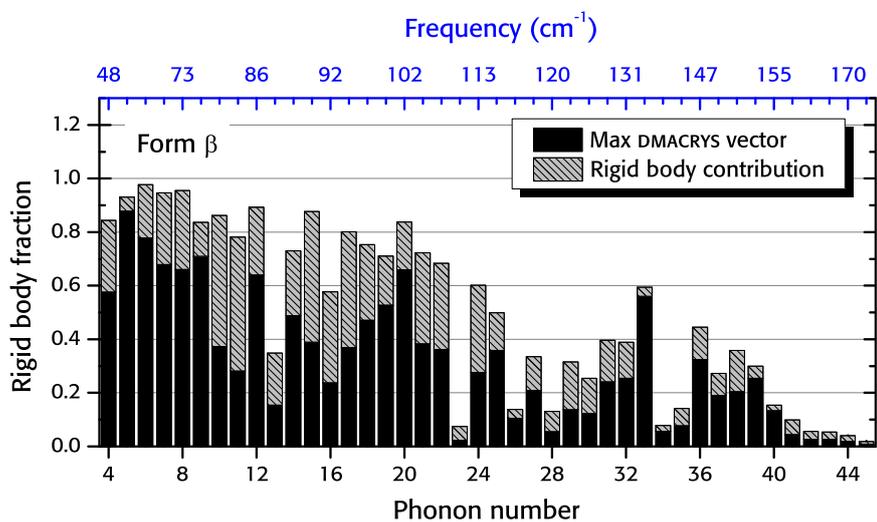


Figure 8.23. Rigid body contribution to the first 45 nitrofurantoin  $\beta$  phonon eigenvectors calculated by CASTEP using the dispersion corrected PBE functional (excluding the 3 acoustic phonons). The solid portion of each bar shows the contribution from the most representative DMACRYS eigenvector

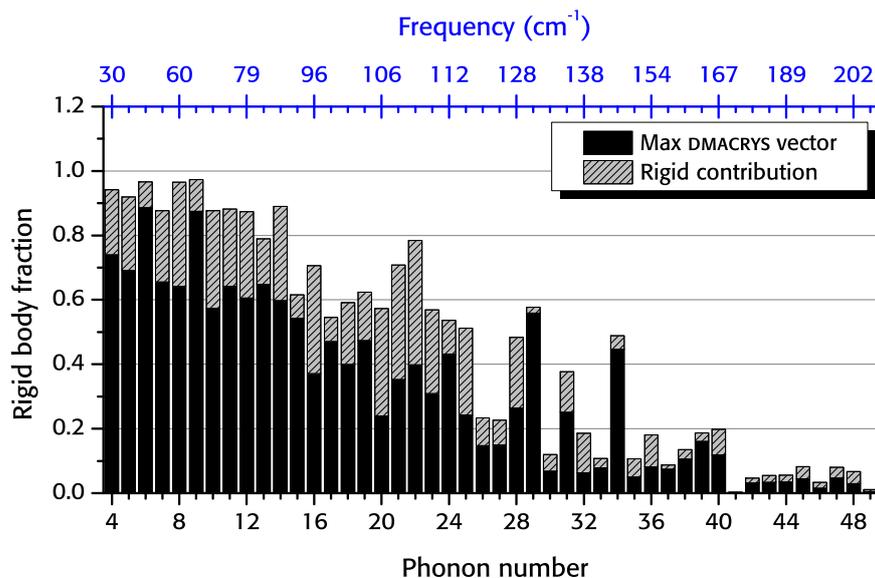


Figure 8.24. Rigid body contribution to the first 49 carbamazepine form III phonon eigenvectors calculated by CASTEP using the dispersion corrected PBE functional (excluding the 3 acoustic phonons). The solid portion of each bar shows the contribution from the most representative DMACRYS eigenvector

lecule, rather than its size, as exemplified by the good agreement in the case of the carbamazepine form III, the crystal with the larger molecule among our systems.

## 8.5 CONCLUSIONS

We have analysed the results from the CASTEP phonon calculations to assess the importance of molecular rigid-body vibrations in the eigenvectors across the whole spectrum, and to determine whether it is correct to use a rigid-molecule approach for the calculations of the terahertz spectrum.

For every system we tested we obtained an approximately sigmoidal behaviour of the rigid body contribution to the total vibrational eigenvectors: the vibrations are only intramolecular if the frequency is higher than a clearly identifiable threshold, that varies between molecular crystals (we used a 10 % maximum rigid-molecule contribution to define  $\omega_{\text{high}}$  for each system), that varies from system to system and spans from  $150 \text{ cm}^{-1}$  for nitrofurantoin to almost  $300 \text{ cm}^{-1}$  for benzoic acid, decreasing with the size of the system and depending slightly on its crystal

Crystal system	# DMACRYS	DFT-D vectors sum		PBE vectors sum	
	opt. modes	total	%	total	%
Benzoic acid	21	16.85	80.25	17.19	81.90
Theophylline	21	<sup>a</sup>	<sup>a</sup>	13.39	63.78 <sup>b</sup>
Paracetamol I	21	15.06	71.72	14.51	69.10
Paracetamol II	45	32.78	72.84	32.85	73.01
Nitrofurantoin $\alpha$	9	5.91	65.76	6.45	71.74
Nitrofurantoin $\beta$	21	13.95	66.45	14.33	68.26
Carbamazepine III	21	<sup>c</sup>	<sup>c</sup>	15.51	73.88

<sup>a</sup> Imaginary frequencies in the calculation

<sup>b</sup> Acoustic phonons are not fully translational

<sup>c</sup> Calculation was not performed without dispersion correction

Table 8.17. Sum of the most representative DMACRYS to each CASTEP calculated eigenvector for seven molecular crystals (number of rigid-molecule optical modes for each system reported). The total sum and the average percentile contribution per mode is reported

form.

The main limitation of a rigid molecule approach is that we neglect internal vibrations due to molecular flexibility. We have considered the position of the lowest normal modes of vibration of the isolated molecules we studied, and the lowest frequency internal modes within the periodic DFT approach. The internal vibrations in the solid state are shifted to higher frequencies (Table 8.15), so that in our most rigid systems (benzoic acid, theophylline) the absorptions within the terahertz range are all lattice modes; for the other systems only 1-2 of the lowest frequency intramolecular normal modes are seen in the terahertz range.

In the second part of the chapter we analysed the similarities between the DMACRYS and CASTEP eigenvectors. As before, we found a good agreement for our smallest systems, and an increase in the mixing for our more flexible systems, in particular nitrofurantoin form  $\alpha$ . In the case of nitrofurantoin, and for more flexible molecules the DMACRYS approach might not be method of choice, since these calculations ignore the influence that internal vibrations will have on the vibrational spectrum in the frequency range on interest for terahertz spectroscopy.

A possible solution for our calculations would be to model the mixing between these internal vibrations and the lattice modes: this will eventually be implemented in DMACRYS, by including the evaluation of the second derivative of the crystal energy with respect to the conformational

change in the molecule, and the inclusion in the dynamical matrix of intramolecular degrees of freedom (Equation 3.4). This would allow us to selectively include only the flexibility relevant to the terahertz region, without formally including all the intermolecular degrees of freedom force constant, with small changes to the DMACRYS code.

## THE ROLE OF WATER IN THE LATTICE DYNAMICS OF CRYSTALLINE HYDRATES

---

**I**N THIS CHAPTER we discuss the effect of water inside the crystalline structure on the lattice modes of molecular crystals. The study and the monitoring of the formation of the hydrates in molecular crystals is important for pharmaceutical purposes: as a result of water inclusion, the crystal unit cell is not the same and, as with polymorphs, the physical properties of the material are affected. More precisely, the solubility of the hydrate form is reduced, as can be its dissolution rate. All such changes ultimately affect the bioavailability of the drug molecule. For these reasons, pharmaceutical drug molecules are usually marketed as neat crystals [216, 217]. The variation of the mechanical properties of the material that could be caused by hydrate formation can be a problem for the process of manufacturing: these changes might seriously affect the tablettability [218] and the processability of the drug [219].

Spectroscopy is not only useful as a form of fast characterisation of a material [220], but also to analyse the kinetics and mechanisms of the interconversion between forms, to better monitor their formation and disaggregation [57, 221].

For these reasons, it is sensible to study and understand the vibrational properties of water in the solid state, and its effect on the vibrational modes of the active ingredient in the crystal (API).

In the following sections we analyse the lattice mode eigenvectors of the hydrate crystal forms described in chapter 6: theophylline monohydrate; paracetamol monohydrate and trihydrate (the stable forms at standard conditions); nitrofurantoin form I and II; carbamazepine dihydrate. We examine these eigenvectors to look for similar features in the absorptions of all the hydrate structures studied and a systematic description of the hydrate spectra.

### 9.0.1 Peculiarity of the water molecule

The water molecule is in many ways different from the API molecules we have studied up to now. It is evidently a more “compact” and lighter molecule: it has a molar mass of 18 Da, much lower than our smallest system, benzoic acid (122 Da). We also expect water to behave as a more rigid molecule than the larger organic molecules: the lowest internal vibrational frequency calculated with GAUSSIANO3, B3LYP/6-31G\*\* is  $1711\text{ cm}^{-1}$ .

The rigidity of water suggests that we should not see any of its internal mode of vibrations in the lattice mode region, and all the water vibrations in the region 0 to  $1000\text{ cm}^{-1}$  will therefore be rigid translational and rotational motions. We investigate the amount of rigid-body contribution in the eigenvectors resulting from the CASTEP calculations following the procedure outlined in chapter 4.4, and we look for characteristic vibrations in the spectrum.

The mass of a molecule is a very important parameter in the determination of the absorption frequencies in the terahertz spectrum: as it is shown in Equation 3.4, page 31, the vibrational frequencies are inversely proportional to the square root of the molecular masses. The ratio of the masses between the API and water (Table 9.1) varies between 8.4 and 13.2.

The difference between water and the API molecules is even more evident in the analysis of the moment of inertia of the molecule around the axis of rotation, that takes the role of the mass in Equation 3.4 whenever rotations – instead of translations – are involved. These moments are calculated using the moment of inertia tensor  $\mathcal{I}$ , whose components are defined as

$$\mathcal{I}_{ij} = \sum_k^{\text{atoms}} m_k (r_{xk}^2 + r_{yk}^2 + r_{zk}^2 \delta_{ij} - r_{ik} r_{jk}), \quad (9.1)$$

where we used the atomic mass  $m_k$ , the  $i$ th Cartesian components of the position of the  $k$ th atom  $r_{ik}$  with respect to the axis of rotation, and the Kronecker delta  $\delta_{ij}$ .

For water these values are very low for geometrical reasons, since the centre of mass is very close to the only heavy atom, oxygen, as can be noticed in Table 9.1.

As a consequence (from these considerations only, and without taking into account the differences in the force constants) we should expect

a factor of approximately 30 between the absorption frequencies of the heavier molecules compared to water for the modes composed only of rotations, and a factor 4 between the absorption frequencies of the heavier molecules compared to water for the modes composed of translations only. Since the vibrations in the real hydrate crystal are composed of a concerted movement of the water and the API, we can assume that we will see (at least for some of the phonons) a mixing of features, with water movements having a varying degree of relevance to the overall crystal vibrations. This is discussed in the later sections in the context of the CASTEP and DMACRYS eigenvector analysis.

## 9.1 CASTEP CALCULATIONS

CASTEP calculations were performed for the hydrate forms of theophylline, paracetamol (monohydrate only) and nitrofurantoin (forms I and II); DMACRYS calculations were performed for each of the hydrates presented in chapter 6, with the FIT and w99 forcefields.

Unfortunately the CASTEP calculation of theophylline hydrate and polymorph II of nitrofurantoin hydrate (CSD reference code HAXBUD), while apparently finding a minimum in the geometry optimisation step, displayed imaginary vibrational frequencies. This highlights the possibility that the final crystal structures for these systems are not real minima, but rather saddle points for the energy of the crystal structure. We will not discuss these calculations, because of the possible inaccuracies of the calculated forces and frequencies. Furthermore, one of the CASTEP

Molecule	Mass (Da)	Principal moments of inertia (DaÅ <sup>2</sup> )		
Theophylline	180.16	985.49	567.09	425.19
Paracetamol	151.16	1055.71	923.06	137.72
Nitrofurantoin	238.15	3151.56	2780.68	374.11
Carbamazepine	236.26	1603.45	1310.01	740.56
Water	18.02	1.56	1.03	0.53

**Table 9.1.** Molecular masses and moments of inertia of the molecules in the hydrate systems studied using the X-ray determined molecular geometries. For paracetamol this values are taken from the monohydrate (HUMJEE), for nitrofurantoin from form II (HAXBUD). The water geometry is taken from HUMJEE

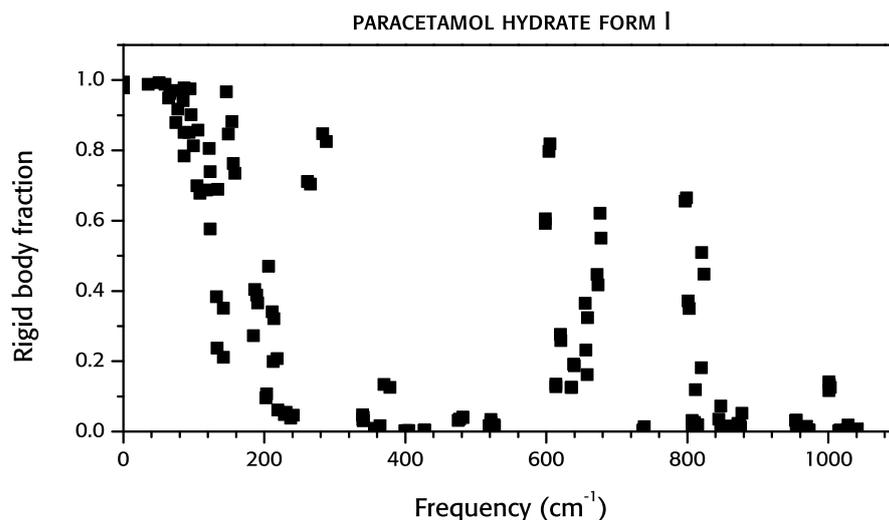


Figure 9.1. Rigid body contribution to the vibrational modes as calculated with CASTEP 5, PBE functional, for paracetamol monohydrate in the range 0 to 1250  $\text{cm}^{-1}$

acoustic phonons in nitrofurantoin hydrate form II does have a considerable fraction of internal contribution (22 %), indicating some problems with the calculation.

The CASTEP calculation for paracetamol monohydrate was among the first we completed and we did not use dispersion correction. The rigid body fraction of the eigenvectors calculated with CASTEP 5, PBE functional is reported in Figure 9.1. Similar to the results of the neat paracetamol calculations, a rigid body region is visible in the region 0 to 150  $\text{cm}^{-1}$ . However, there are some new rigid body features around 250  $\text{cm}^{-1}$  and, more evidently, a set of lattice modes with large rigid molecule contributions in the range 600 to 800  $\text{cm}^{-1}$  where none of the non-hydrate systems have any rigid molecule modes.

The distribution of rigid molecule vibrations in the low frequency region for the hydrate is more similar to that seen for neat paracetamol form I than form II: this is probably related to the similarity between the corresponding crystal structures (see section 6.1). A comparison of the calculated spectra of paracetamol form I and II with the hydrate form in the frequency range 0 to 450  $\text{cm}^{-1}$  highlights the presence of three other highly rigid modes near 160  $\text{cm}^{-1}$  (circled in green in Figure 9.2); despite the similarity in frequency to the high rigidity modes in form II, these are unrelated to one another.

The same kind of behaviour for the rigidity of the phonons can be

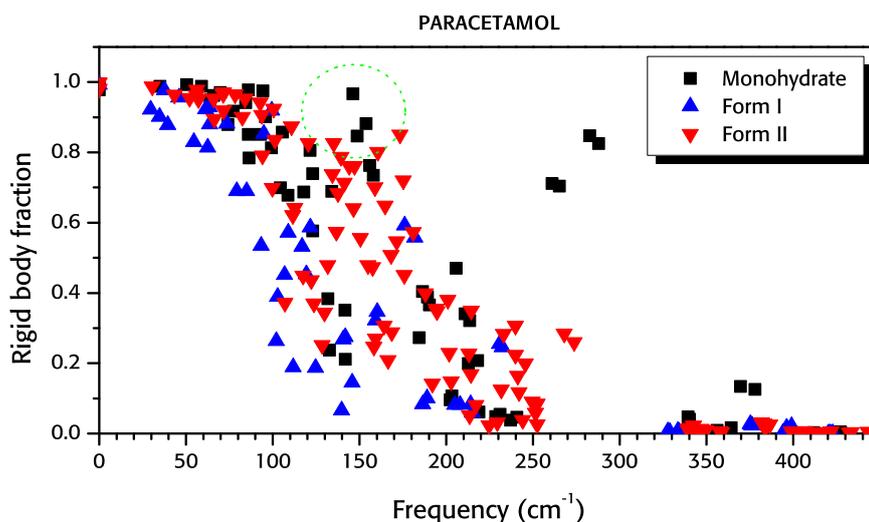


Figure 9.2. Comparison of the rigid body features in the neat and monohydrate forms of paracetamol in the range 0 to  $450\text{ cm}^{-1}$  calculated with CASTEP 5, PBE functional. Form I data points are blue, form II data points are red and the monohydrate data points are in black. Anomalous high frequency, high rigidity lattice modes in the hydrate (see text) are circled in green

seen in the case of nitrofurantoin hydrate form I, shown in Figure 9.3: none of the four CASTEP calculations of the nitrofurantoin neat forms have any rigid body modes (higher than  $\approx 60\%$ ) at a frequency higher than  $120\text{ cm}^{-1}$ , while in the hydrate there are two modes between  $150$  to  $200\text{ cm}^{-1}$  (drawn in red), four modes around  $300\text{ cm}^{-1}$  (drawn in green) and four others (the most rigid of the modes we just mentioned, drawn in blue) at  $550$  to  $700\text{ cm}^{-1}$ ; other modes with varying degrees of rigidity can be seen in the range from  $200$  to  $700\text{ cm}^{-1}$ .

The position of these high rigidity modes can be understood by considering the effect of the mass and moment of inertia of water described in the previous section. We can get a better understanding of the different types of motion in the crystal by removing the amount due to water movements from the total rigid body fraction. In Figure 9.4, we see the effective contribution of only the paracetamol molecules in each vibration, which can be more easily compared to similar values in the anhydrous crystal.

In the acoustic phonons, water and paracetamol molecules exhibit equal displacement (hence an equal  $50\%$  contribution to the total). The contribution of rigid paracetamol motions is significant in the region  $0$  to  $100\text{ cm}^{-1}$ , it then steadily declines until  $200\text{ cm}^{-1}$ , and from this fre-

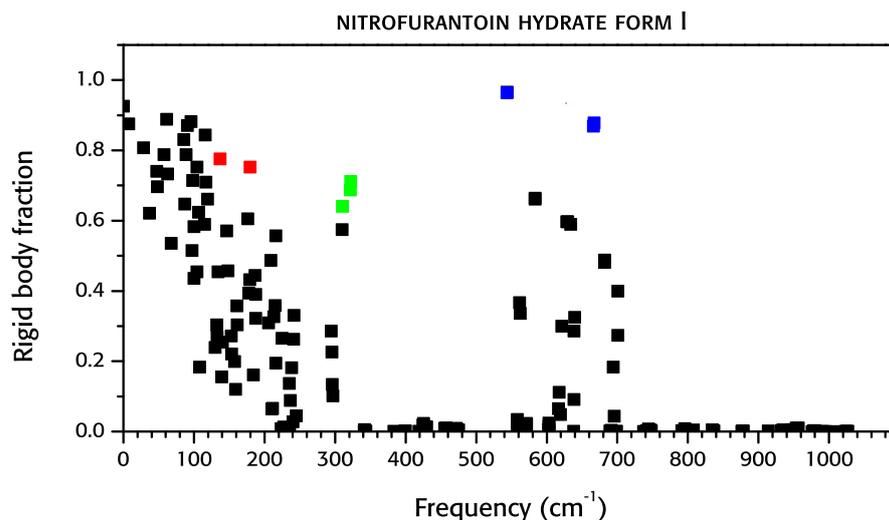


Figure 9.3. Rigid body contribution as calculated with CASTEP 5.5, dispersion corrected PBE, for nitrofurantoin hydrate form I in the range 0 to 1100  $\text{cm}^{-1}$ . Anomalously high frequency, high rigidity lattice modes are coloured according to their frequency (see text)

quency onward becomes negligible, following the same trend as the calculation for the neat crystal structures.

As in the previous chapter, we analysed the resulting rigid body distribution in energy by fitting it with a logistic function. This time the maximum of the function was chosen to be 0.5, since the translational component amount of the acoustic phonons is equally divided between the water and API molecules (the stoichiometry for both hydrates is 1:1). The parameters of the logistic fit (Table 9.2) appear to be slightly higher than for their neat crystal counterparts, probably due to the effect of water. The  $x_0$  value for paracetamol is almost the same as in nitrofurantoin, but the curve is steeper for paracetamol (the  $p$  parameter governs the extent of the intermediate rigidity region). The distribution of nitrofurantoin rigid body fraction appears to be more erratic, as can be seen both from the graph and the  $\zeta$  sum of residues.

Within the limit of only two systems analysed (but with good agreement with the other neat system studied) we can conclude that all the rigid body contributions at higher frequencies, which are absent in the spectra of the anhydrous crystals, are associated only with the motion of the water molecules. We analyse these motions further in the results of DMACRYs calculations, which have been completed successfully for a wider range of hydrate crystal structures.

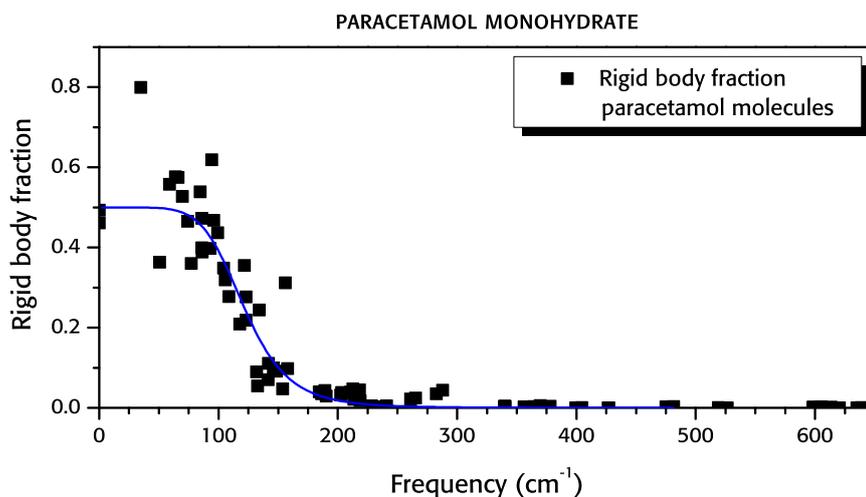


Figure 9.4. Rigid body contribution of the paracetamol molecules to the vibrational modes (excluding water) as calculated with CASTEP, dispersion corrected PBE, for paracetamol hydrate in the region 0 to  $650\text{ cm}^{-1}$ . The best fit for a logistic curve is displayed

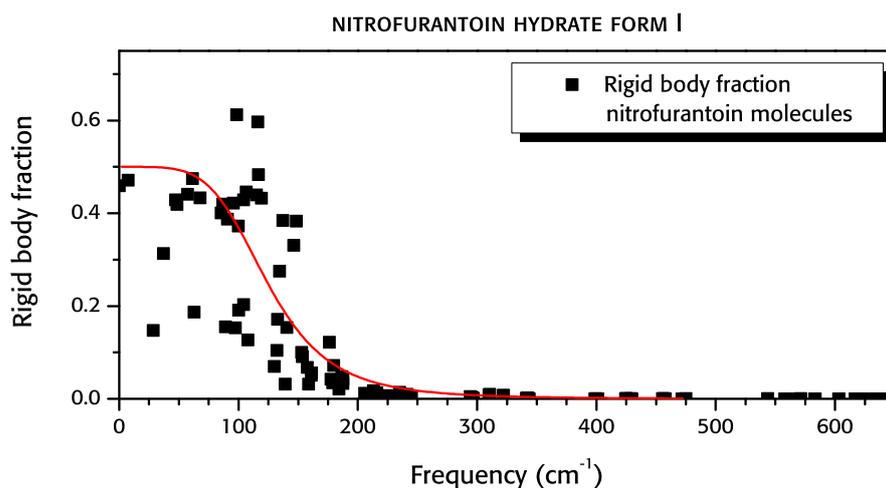


Figure 9.5. Rigid body contribution of the nitrofurantoin molecules to the vibrational modes (excluding water) as calculated with CASTEP, dispersion corrected PBE, for nitrofurantoin hydrate form I in the region 0 to  $650\text{ cm}^{-1}$ . The best fit for a logistic curve is displayed

Molecule	$x_0\text{ (cm}^{-1}\text{)}$	$p$	$\chi^2$	$\varsigma$
Paracetamol	121.59	6.55	0.004	0.32
Nitrofurantoin	124.33	4.73	0.010	0.95

Table 9.2. Characterising parameters of the logistic fit of the rigid body fraction associated with the API in the phonon eigenvectors of paracetamol monohydrate and nitrofurantoin hydrate form I with CASTEP

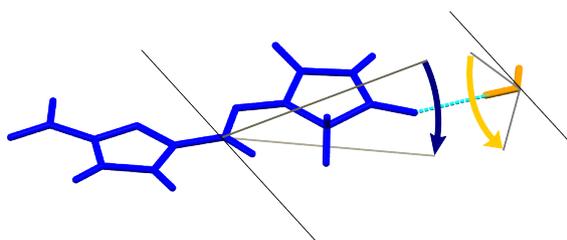


Figure 9.6. A nitrofurantoin molecule hydrogen bonded to water in nitrofurantoin hydrate form I (HAXBUD01). This schematic shows an (exaggerated) angle of rotation for the two molecules if the hydrogen bond distance  $N \cdots H$  (in blue) is to remain unaffected by the rotation. Axes of rotation are in black

## 9.2 DMACRYS CALCULATIONS AND ANALYSIS

The calculations with DMACRYS confirmed the results of our CASTEP data. The eigenvectors were analysed to extract the amount of rotational and translational contribution to each vibration, separately for water and for the API. The translational and rotational vectors (scaled by mass and moment of inertia to represent the displacements) are given in the output of DMACRYS and converted to orthogonal coordinates by RUDOLPH. The use of the rotation vectors as they are in the comparison of the displacement magnitude can give misleading results in the case of molecules of very different sizes.

Consider, for example, the arrangement of two molecules in nitrofurantoin hydrate form I (Figure 9.6). The peripheral atoms of the molecules generally display a similar displacement of their the phonon modes, since this is where the hydrogen bond interactions happen. However, for a compact molecule like water, a much greater angle of rotation is required than for nitrofurantoin, where a small rotation has a greater effect far from its centre.

As a consequence, if we keep the molecular rotations in DMACRYS as the descriptor we end up with an over-representation of the motion of water compared to the one we observe in reality. To deal with this problem we applied a slightly different approach in our description: for each molecule we considered an average over all the magnitudes of the rotational vectors' displacements.

For all of our simulated systems we found that the lattice modes frequencies can be classified into different regions:

REGION 1, the region of the low energy eigenvectors, where each vibrational mode is comprised of a weighted mixture of all the molecules in the unit cell. This is true up to the rigidity region in DMACRYS, that differs for each molecule;

REGION 3, the highest energy region, where vibrational modes are comprised solely of water movements. These vibrational modes can effectively be seen as as free vibrations of the water molecules in the field of the frozen heavier molecules. The  $3Z_w$  highest energy vibrations (where  $Z_w$  is the number of water molecules in the unit cell) are characterised principally by rotational movement, and they are well separated in energy from the rest of the phonons;

REGION 2 is an region intermediate in frequency between the previous two. In this region, the main interaction probed is the water-API hydrogen bond stretching; the modes in this region can be easily differentiated from the others, and in most of the systems are well separated in energy as well.

The frequency ranges of these regions are summarised of all systems studied here in Table 9.3.

We expect to see a clear frequency separation in the phonon frequencies between modes prevalently composed of water motion and those that involve heavier organic molecules. We have observed in all of our systems that whenever motion of the water molecules makes a significant contribution to the eigenvectors at lower frequencies (region 1), it does not perturb significantly the hydrogen bond between API and water: the water behaves as “rigid connector” between the API molecules, mediating the vibration with a bending of its hydrogen bond, rather than a stretch (that is, changing the angle in the hydrogen bond rather than its length). Vibrational modes exemplifying this effect are reported in the later sections.

The main perturbation of the hydrogen bonds involving water appear in the intermediate region 2, when the API still makes a partial (although small) contribution to the total vibrational eigenvector .

No appreciable variation in the composition of the eigenvectors was found by changing from the w99 to the FIT forcefield, as almost all of the modes show a 1:1 agreement, as can be seen in the projection of the eigenvectors of the w99 calculation onto the FIT eigenvectors (see

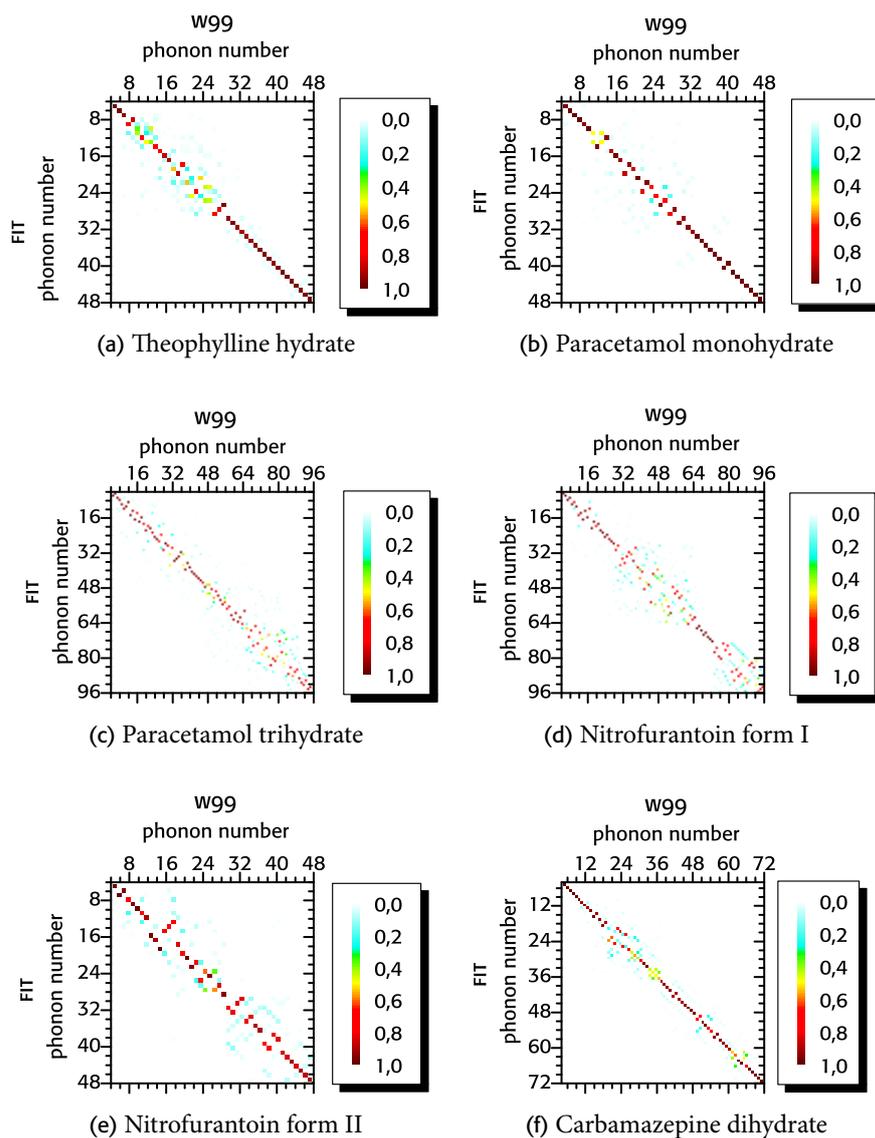


Figure 9.7. Direct product of DMACRYS eigenvectors using FIT and w99 force-field: the colour represents the square of the magnitude of the projections of each vector upon each other

Figure 9.7). For some of the modes the effect of the potential was to swap their relative energy, as happens for nitrofurantoin form II, modes 12–17 (Figure 9.7e).

The most sensitive parts of the results appear to be the frequency of the water-dominated normal modes in region 3, which are found at considerably higher frequencies in the w99 than with FIT calculations (Table 9.3).

There is also a clear difference in the highest frequency vibration between the systems where water molecules are contained in pockets

System	Refcode	Hydrate type	Forcefield	Frequency / cm <sup>-1</sup>		
				Region 1	Region 2	Region 3
Theophylline monohydrate	THEOPH	Channel	FIT	18-114	159-172	234-774
			W99	20-105	146-175	207-979
Paracetamol monohydrate	HUMJEE	Pocket	FIT	35-125	134-220	281-594
			W99	35-119	133-202	386-728
Paracetamol trihydrate	XOMWOL	Channel	FIT	22-126	147-204	226-737
			W99	21-126	129-217	234-962
Nitrofurantoin form I	HAXBUD	Pocket	FIT	16-127	183-191	323-442
			W99	18-133	179-187	405-546
Nitrofurantoin form II	HAXBUD01	Pocket	FIT	22-118	158-217	276-470
			W99	24-113	184-216	454-657
Carbamazepine dihydrate	FEFNTO2	Channel	FIT	13-120	158-204	227-823
			W99	26-130	158-219	239-1127

Table 9.3. Extent of the three vibrational regions defined at page 178 for all of the hydrate crystal forms studied, with the w99 and FIT forcefields. The csd refcode and the type of hydrate (channel or pocket) is reported

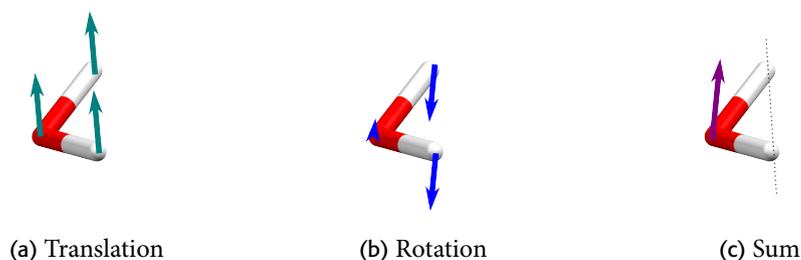


Figure 9.8. Decomposition of a rotation around an arbitrary axis (dotted in the Figure on the right) as centre of mass rotations and translation

(paracetamol monohydrate and nitrofurantoin hydrates) and the channel hydrates. The vibration of the water molecules in the pocket hydrates can be seen as happening in an enclosed field of frozen molecules. However, in structures where the interactions between water molecules are possible (the channels in the theophylline and carbamazepine crystals, and the more complex reticular interactions of waters within the paracetamol trihydrate crystal) the concerted movement of water molecules results in higher vibrational frequencies, around  $1000\text{ cm}^{-1}$  with the w99 forcefield, compared to the frequencies for the pocket hydrates, which predicts vibrations only up to  $800\text{ cm}^{-1}$  for all systems (see Table 9.3).

The separation into different spectral regions is difficult to spot if we look only at the breakdown of contributions listed in the tables, because the decomposition of the eigenvectors does not convey the relative movement of molecules and the way that hydrogen bonds are distorted. For example, the high amount of water contribution in the deconstruction of the phonon vibrations in theophylline hydrate modes 27 and 28 (Figure 9.9), could mislead the reader into thinking that these are region 2 (or 3) modes; however, we classify these as region 1 modes because the relative motion of water consists of bending, and not stretching, of the hydrogen bonds. Their classification as form 1 modes is confirmed by their low frequency ( $105\text{ cm}^{-1}$ ).

Another problem is that the rotations are expressed in the terms of centre of mass frame, and this can generate unfortunate artifacts due to the mutual cancellation of rotational and translational contributions to atomic displacements, as depicted in Figure 9.8. Rotations with respect to other axes can be obtained (approximately) as a sum of molecular translations and rotations. Such an effect is clearly visible, for example, in the lattice modes of nitrofurantoin hydrate form I (Figure 9.17) at

low frequency. In modes 9–17, since the axis of rotation of the water passes through its hydrogens, vibration and rotation of water cancel out for hydrogen: for this reason water is overrepresented, since the oxygen is the only atom moving.

We will try to disentangle problematic assignments as we discuss each specific system.

### 9.2.1 Theophylline hydrate

For the calculation of the phonon spectrum of theophylline hydrate, it was necessary to relax the symmetry of the unit cell to allow the formation of physically sensible hydrogen bonds in the water channel (see section 7.3.2). The disorder of the hydrogen atoms in the methyl group, despite shifting the positions of some of the lattice mode frequencies, does not significantly affect the eigenvectors. While we will only consider the eigenvectors from one of the symmetry-reduced unit cells in this section (structure 2 in Table 7.7), the discussion is valid for all the other structures as well.

Theophylline hydrate is characterised by a narrower frequency window for the rigid body vibrations of the API compared to the anhydrous form: the highest mode for the w99 calculations with an API contribution greater than 10 % has a frequency of  $104 \text{ cm}^{-1}$  (mode 27 in Figure 9.9), compared to  $174 \text{ cm}^{-1}$  of the highest frequency mode in the neat form.

This difference shows the considerable effect that the molecular environment can have on the lattice mode frequencies, and the differences in intermolecular forces in each arrangement. For instance, we can infer that the energy involved in the distortion of the molecular chains in the anhydrous form is higher than the energy probed in the distortion of the dimers in the hydrate (mode 26 in Figure 9.9).

Most of the characteristic features of the vibrations in theophylline hydrate are due to the channel distribution of the water molecules inside the crystal. The interaction between water molecules within the channel is very strong: in the first 14 eigenvectors (the lower energy modes within region 1) these hydrogen bonds are not perturbed, and the displacement vectors of neighbouring  $\text{O} \cdots \text{H}$  remain parallel. This effect is noticeable in region 2 as well: the four modes in this region, which are

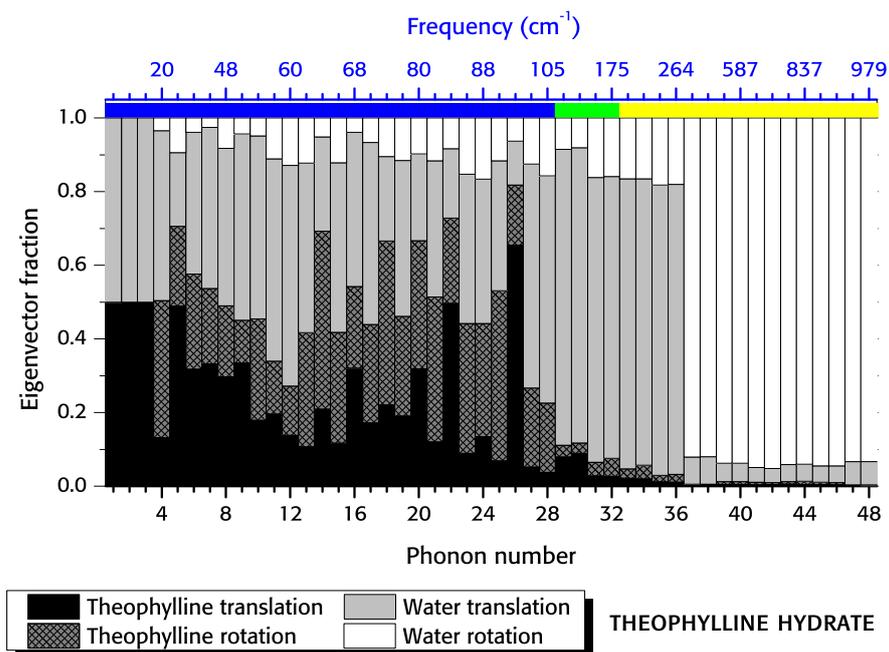


Figure 9.9. Breakdown of the contributions to the eigenvectors of the symmetry-relaxed theophylline hydrate (THEOPH) calculated for the 48 phonon modes with DMACRYS, w99 force field. Regions 1, 2, 3 are respectively in blue, green and yellow on the top bar

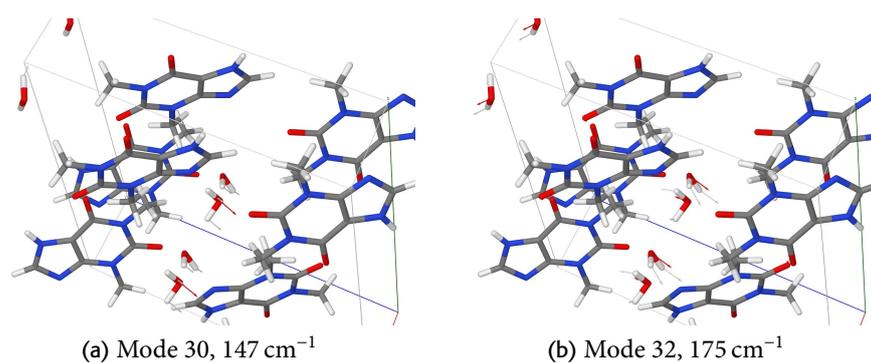


Figure 9.10. Vibrational motions of the water in the symmetry-relaxed theophylline hydrate in region 2 calculated with DMACRYS, w99 force field; on the left, no inter-water hydrogen bond is stretched

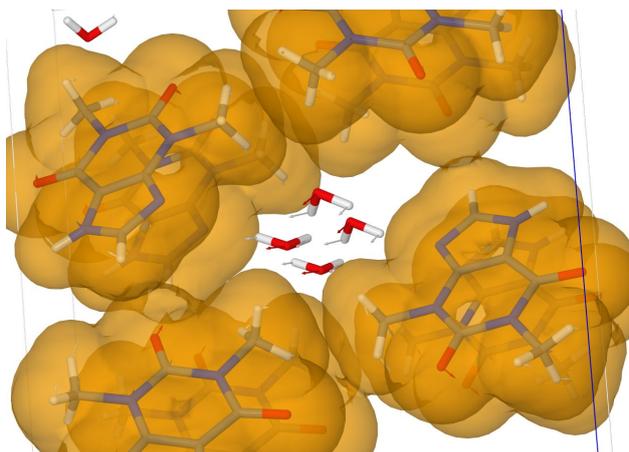


Figure 9.11. Vibrational motions of the water molecules in the symmetry-relaxed theophylline hydrate, mode 12, calculated with DMACRYS, w99 forcefield. The theophylline molecules are displayed within their van der Waals surfaces, to highlight the water channel. The displacement vectors are coloured the same as the atoms

symmetry adapted stretches of the water-theophylline hydrogen bond, have very different frequencies ( $147\text{ cm}^{-1}$  vs  $175\text{ cm}^{-1}$ ) depending on whether they do include stretching of inter-water hydrogen bonds (Figure 9.10).

Another interesting feature of the water molecules inside the channel is their cooperative behaviour: modes with a large translational water contribution (especially modes 9, 11, 12, 17) effectively behave as a collective, bigger “macromolecule” moving through the channel and interacting with the API (Figure 9.11). The fact that this vibrational movement is not hindered by the presence of the API might be the reason behind the low frequency of these water-dominated modes compared to those in region 3.

### 9.2.2 Paracetamol monohydrate

The deconstruction of the phonon displacements for paracetamol monohydrate is reported in Figure 9.14. All but one of the possible hydrogen bond sites for water molecules in the structure are filled (see Figure 6.5), with each water forming bonds with 3 paracetamol molecules. All of these interactions are of the type  $\text{OH} \cdots \text{O}$ . The shortest  $\text{O} \cdots \text{H}$  distance is the one involving the oxygen in water with the hydroxyl group ( $1.755\text{ \AA}$ ), with the two involving the hydrogen atoms be-

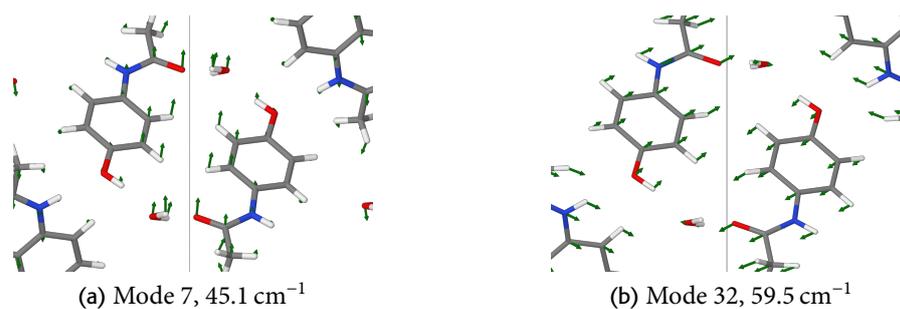


Figure 9.12. Vibrational motions of water in paracetamol monohydrate in region 1 of the spectrum calculated with DMACRYS, w99 forcefield; paracetamol molecules are prevalently rotating, and the water translating; on the right, paracetamol molecules are translating in different directions, and water tries to follow

ing slightly longer (1.831 Å and 1.922 Å) and with almost linear bonds. There is also a strong NH ··· O API-API hydrogen bond, that is stretched in few eigenmodes in the region 95 to 115 cm<sup>-1</sup>.

For this system there is no possibility of modes dominated by water contributions at low frequency (as happens in theophylline monohydrate), since such a movement would invariably perturb the high energy hydrogen bonds in the crystal; for this reason in the lower energy modes water molecules translate along with their hydrogen-bonded paracetamol molecules (mode 7, Figure 9.12a). When such a concerted translation of water and paracetamol is impossible, the water molecules move in such a way to minimise the perturbation of the hydrogen bonds (as in mode 9, Figure 9.12b).

The lower energy modes of the spectrum are dominated by the translational contribution of the API, while the rotational contribution becomes more important for the modes at the high frequency end of region 1. Modes in region 2 and 3 are graphically very recognisable, as they are dominated by translation and rotation of water, respectively.

We can count 12 modes belonging to region 2, a number increased in the comparison with theophylline hydrate, because of the number of hydrogen bonds in the crystal. There is a clear separation in energy between the region 2 modes, depending on the number of hydrogen bonds involving water molecules that are stretched: three ( $\omega = 194$  to 202 cm<sup>-1</sup>), two ( $\omega = 155$  to 165 cm<sup>-1</sup>) or one ( $\omega = 133$  to 142 cm<sup>-1</sup>).

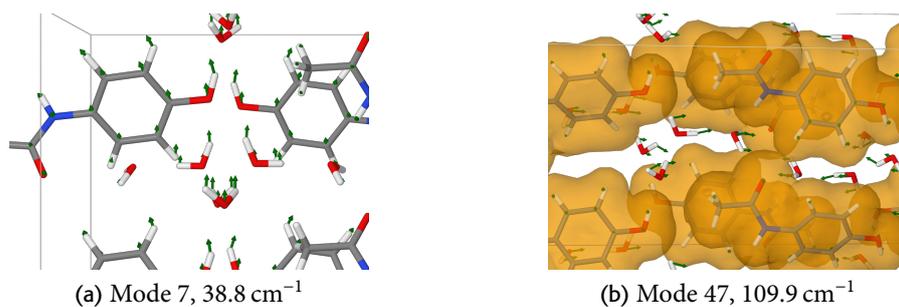


Figure 9.13. Water-dominated modes in paracetamol trihydrate calculated with DMACRYS, w99 forcefield. On the left figure, water movement goes along with paracetamol displacements; on the right, collective movement of water in its channel

### 9.2.3 Paracetamol trihydrate

Paracetamol trihydrate is the biggest system we studied, with 32 molecules in the unit cell (24 water molecules). The water:paracetamol proportion in the displacement is 3:1 for the acoustic phonon modes, as expected from stoichiometric considerations. The mode decomposition, Figure 9.15, is dominated by the predominance of water: we can see that the 72 ( $24 \times 3$ ) highest frequency modes are pure rotational motions of the water molecules and other 56 are dominated by water translations.

Despite the structural differences with paracetamol monohydrate, the deconstruction of the eigenvectors in Figure 9.15 have similar features, with translation being more important at lower frequencies, and rotations more relevant at the high frequency end of region 1.

In contrast to the monohydrate, there are no close contacts between paracetamol molecules, and all the interactions are mediated by water molecules. In particular, one water molecule directly connects three paracetamol molecules (the blue molecules in Figure 6.5, page 81), in an arrangement similar to that in the monohydrate; the other two water molecules are arranged to form a channel.

The collective movement of the water molecules within the channels is visible in most of the modes 44–58, with low paracetamol contribution (frequency range 81 to  $112 \text{ cm}^{-1}$ ). As in the case of theophylline, water molecules move together within the channel (Figure 9.13b), and the number of modes is proportional to the number of symmetry adapted combinations (which is higher than for theophylline hydrate because of the higher number of symmetry elements in the unit cell). In some

other low frequency modes with high water contribution (e.g. modes 4 and 6–8, Figure 9.13a) the translation of the water molecules is highlighted by the fact that they are at the periphery of the rotation of the API molecules, similarly to what we reported in Figure 9.6.

There are several types of mode in region 2, once again according to the number and type of hydrogen bonds probed by the water movements. The lowest energy modes (in the range from 128 to 153  $\text{cm}^{-1}$ ) involve the distortion of the API-water-API structure, with stretching of the hydrogen bond ring (Figure 9.16a); higher energy vibrations involve the other interactions (in magenta, Figure 9.16b) while at the same time distorting other water-water hydrogen bonds (in blue on the same figure). The more the water-water interactions are distorted, the higher the energy of the lattice mode. The highest four vibrations are merely water-water stretching modes, and we classified them as region 3 modes.

In the calculation of the frequencies for the trihydrate (particularly with the w99 forcefield) the energy gaps between the regions are not as sharp as in the other systems, once again because the overwhelming contribution of water makes the transition smoother.

#### 9.2.4 Nitrofurantoin hydrate form I

The breakdown of the eigenvectors of nitrofurantoin hydrate form I is reported in Figure 9.17. The water molecules are coplanar with the nitrofurantoin molecules, and they are located within a pocket formed by three nitrofurantoin molecules, with which they form three hydrogen bonds.

There are eight modes in region 2: the highest frequency four (207 to 216  $\text{cm}^{-1}$ ) stretch all the three hydrogen bonds of water (Figure 9.18b), while the other four lower energy modes (158 to 170  $\text{cm}^{-1}$ ) are a mixture of hydrogen bond stretching and bending.

The distribution of rigid body motion in nitrofurantoin is slightly different than what was found in the paracetamol hydrates: the modes with reduced API contribution are located in the middle of region 1, rather than at its end, despite the absence of water channels that we used to justify similar features in theophylline monohydrate and paracetamol trihydrate.

The water-dominated lattice modes in the range 41 to 88  $\text{cm}^{-1}$  (modes

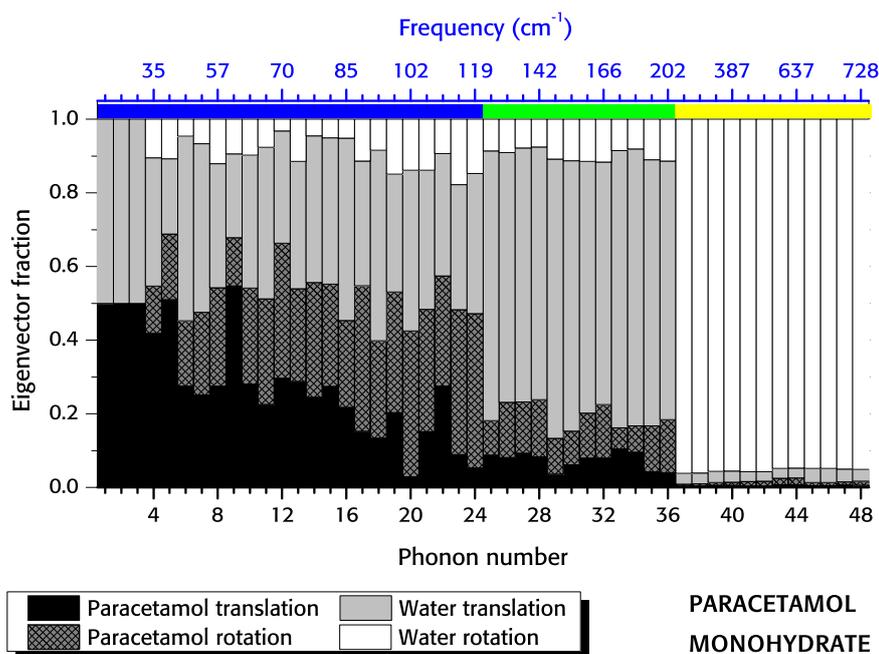


Figure 9.14. Breakdown of the contributions to the eigenvectors of paracetamol monohydrate (HUMJEE) calculated for the 48 phonon modes with DMACRYS, w99 force field. Regions 1, 2, 3 are respectively in blue, green and yellow on the top bar

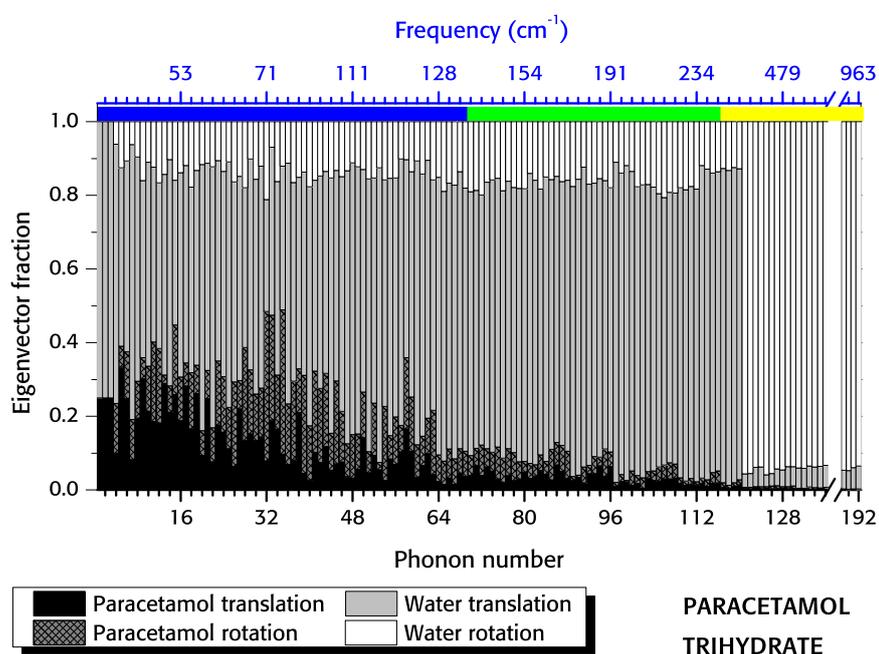


Figure 9.15. Breakdown of the contributions to the eigenvectors of paracetamol trihydrate (XOMWOL) calculated with DMACRYS, w99 forcefield. For clarity, only 16 of the highest 72 phonons are shown, since they all display water contribution > 99.5 %. Regions 1, 2, 3 are respectively in blue, green and yellow on the top bar

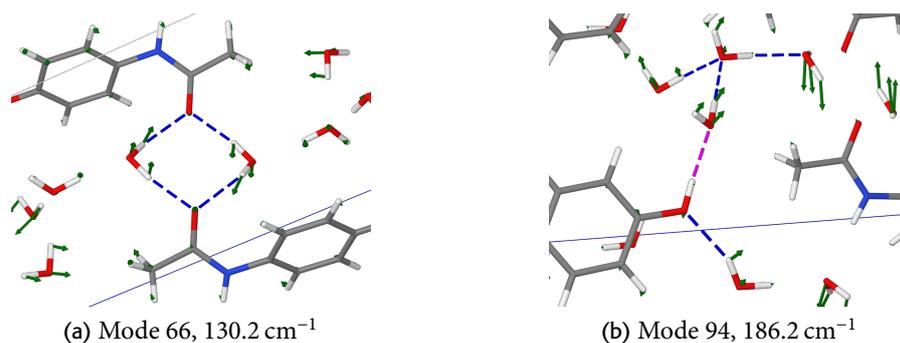


Figure 9.16. Paracetamol trihydrate region 2 modes calculated with DMACRYS, w99 forcefield. On the left, the lowest energy vibration; on the right, an intermediate energy mode

7–19) are an artifact of the decomposition in terms of centre of mass contributions, as we mentioned previously in the chapter. Figure 9.18 shows this clearly: the movement of the water molecules is prevalently rotational, its axis near the position of the hydrogen bonds.

In general, region 1 modes are dominated by the layered packing of the structure: the three lowest frequency involve simple slidings of each molecular plane upon each other and they are therefore very low in energy, since they do not affect any hydrogen bond ( $22.3$  to  $24.6\text{ cm}^{-1}$ ): the higher energy out-of-plane modes start to appear at mode 15 (frequency of  $63.1\text{ cm}^{-1}$ ).

In the highest energy mode in region 1, all of the molecules in a plane form a “wave” movement, involving rotation of neighbouring molecules in alternate directions, with almost no translation involved (Figure 9.18c).

There are eight region 2 modes: only the highest energy four ( $207$  to  $217\text{ cm}^{-1}$ ) involve distortion of the hydrogen bond acceptor in water. Since these modes are well separated from the others in region 2 (which span the energy range from  $158$  to  $180\text{ cm}^{-1}$ ) we can safely assume that the hydrogen bond acceptor in water is the one forming the strongest bond out of the three.

### 9.2.5 Nitrofurantoin hydrate form II

The breakdown of the eigenvectors of nitrofurantoin form II is reported in Figure 9.19. Since form II has double the number of molecules within the unit cell than form I, the number of lattice modes is increased.

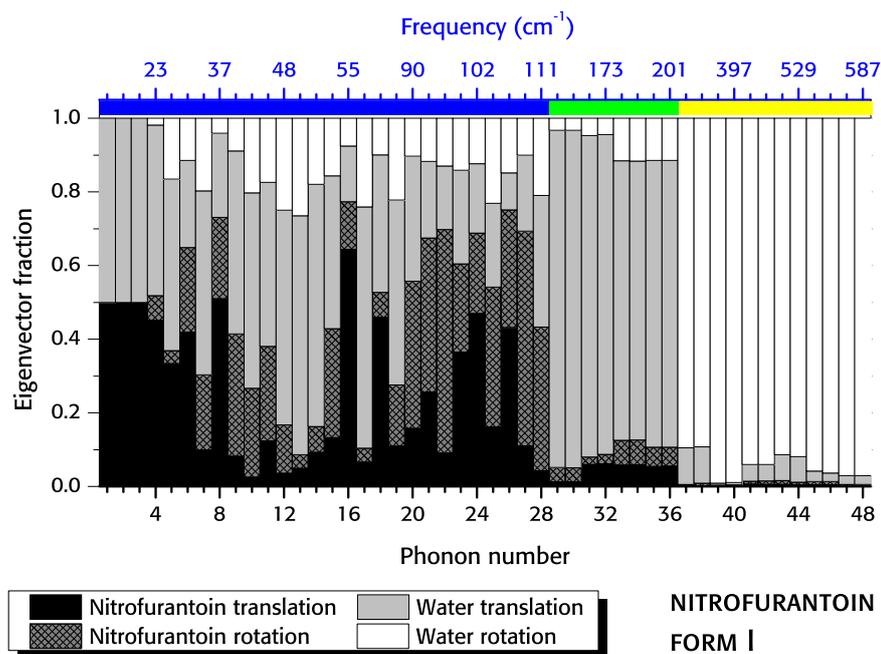


Figure 9.17. Breakdown of the contributions to the eigenvectors of nitrofurantoin hydrate form I (HAXBUD) calculated for all the 48 phonons with DMACRYS, w99 force field. Regions 1, 2, 3 are respectively in blue, green and yellow on the top bar

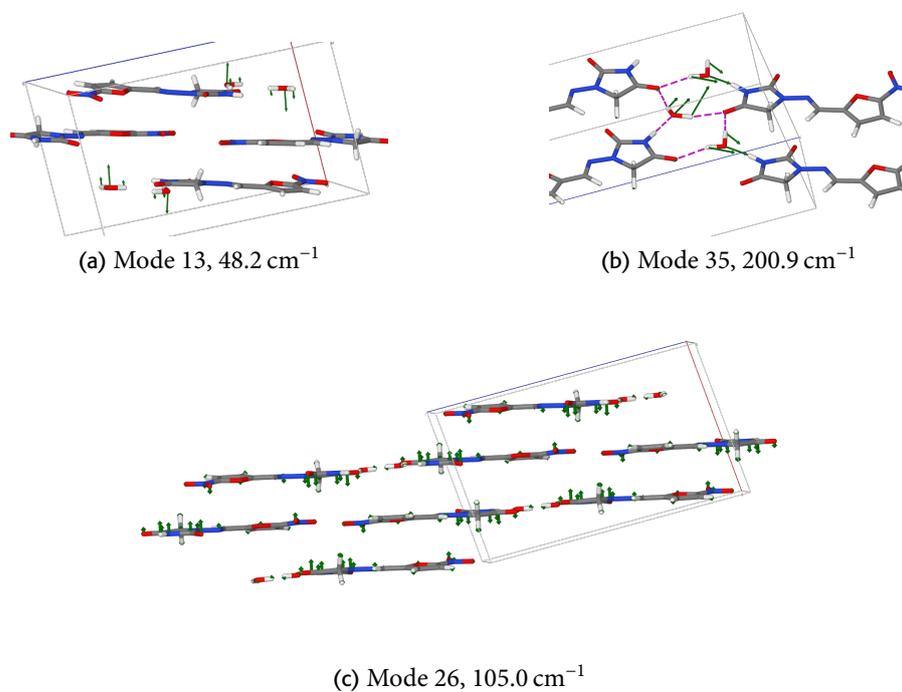


Figure 9.18. Region 1 (a, c) and region 2 (b) modes in nitrofurantoin hydrate form I calculated with DMACRYS, w99 forcefield

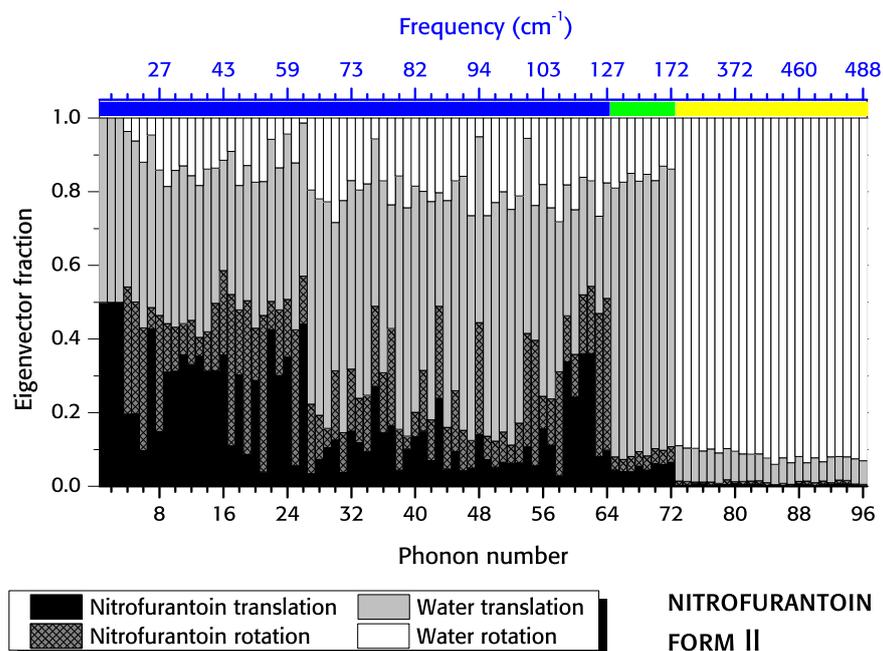


Figure 9.19. Breakdown of the contributions to the eigenvectors of nitrofurantoin hydrate form II (HAXBUD01) calculated for all the 96 phonons with DMACRYS, w99 force field. Regions 1, 2, 3 are respectively in blue, green and yellow on the top bar

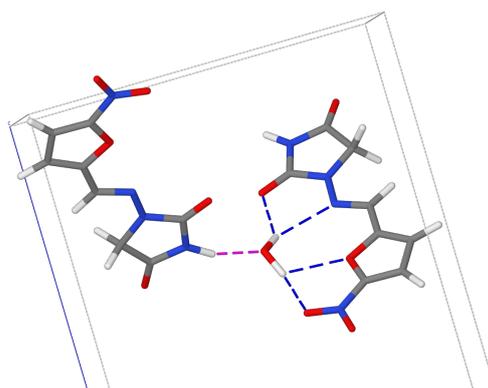


Figure 9.20. Hydrogen bonds for water in nitrofurantoin hydrate form II. In magenta, the hydrogen bond between the water oxygen and nitrofurantoin. In blue, the interaction between the hydrogens in water and the four acceptors in nitrofurantoin. Only part of the unit cell is shown for clarity

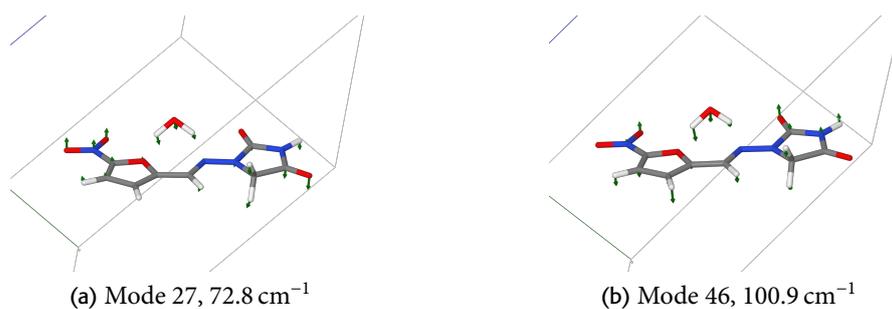


Figure 9.21. Water-nitrofurantoin interaction in mode 27 and 46 for nitrofurantoin hydrate II, calculated with DMACRYS, w99 forcefield

The hydrogen bond network in nitrofurantoin hydrate form II is completely different from form I. Each water molecule acts as a connector between two nitrofurantoin molecules only: while the acceptor in water forms a strong N–H ··· O bond with one of the API molecules, the donors form a complex four-atom interaction with the other nitrofurantoin molecule (see Figure 9.20).

From the analysis of the phonon vibrations the four-atom interaction appears to be the strongest hydrogen bond out of the two: the displacement vectors of the hydrogen atoms in water have the same direction and magnitude as the acceptor atoms of the API up to mode 17 (51.2 cm<sup>-1</sup>), while the oxygen in water has more freedom to move (with an overall rigid body rotational movement around the “fixed” hydrogen atoms).

Most of the modes between 28 and 52 (68 to 104 cm<sup>-1</sup>) are characterised by a high contribution of water to the vibration, with a resulting bending of the four-atom interaction. The energy of the interaction is increased for higher frequency modes, as it can be seen from the different interactions in Figure 9.21a and 9.21b).

The eight region 2 modes involve the stretching of all the hydrogen bonds in water, and they all have roughly the same frequency (165 to 172 cm<sup>-1</sup>).

### 9.2.6 Carbamazepine dihydrate

In carbamazepine dihydrate, the water molecules are organised in four columns along the crystallographic *c* axis, hydrogen bonded to each other to form an infinite chain, and connected to the carbamazepine dimer via the oxygen in the carboxamide group (Figure 9.22).

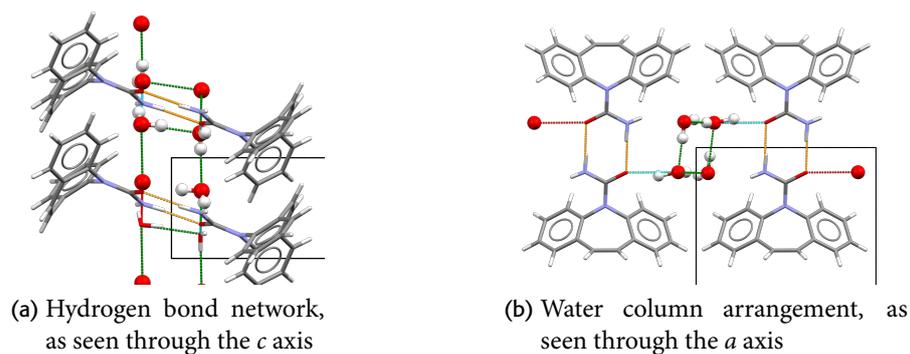


Figure 9.22. Crystal structure of carbamazepine dihydrate (FEFNOTO<sub>2</sub>) highlighting the hydrogen bond in the dimer (orange colour) and between the waters (green)

In the lowest energy lattice vibrations, the hydrogen bonds are not affected, and there is displacement of macro-units connected by hydrogen bonds (see Figure 9.24). As a consequence of the low energy of the interactions, the frequencies of these modes are very low ( $13$  to  $34\text{ cm}^{-1}$ ).

As in the case of the other channel hydrates, most of the lattice modes with high water contribution within region 1 can be interpreted in terms of the relative movements of the water and carbamazepine substructures. In the lowest energy modes, 10–12 and 15 in the frequency range  $38$  to  $68\text{ cm}^{-1}$ , the water columns move as a whole together, while in higher frequency modes (18, 20, Figure 9.26) two water molecule columns move in opposite directions.

While the translational movement of the API is predominant at lower frequencies, rotation becomes important for the higher energy modes in region 1 (see Figure 9.27).

Modes in region 2 are counterintuitive from the point of view of energy compared to the other systems. Mode 35 (Figure 9.28), although having a low frequency ( $118.4\text{ cm}^{-1}$ ) is already stretching the hydrogen bonds of water, while water vibrational modes 41–48 ( $193$  to  $227\text{ cm}^{-1}$ ) belong to region 3, since they stretch the hydrogen bonds within the water columns.

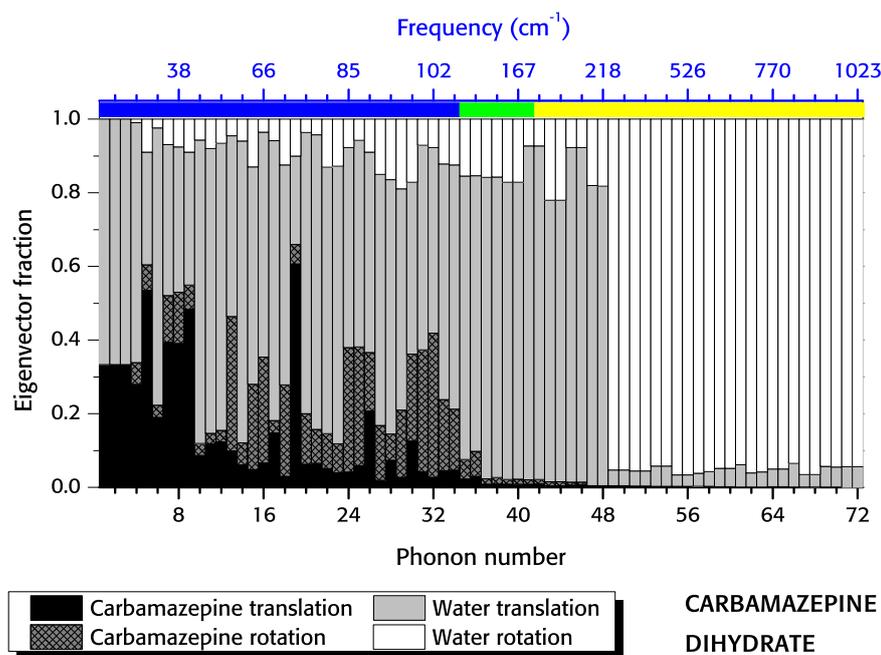


Figure 9.23. Breakdown of the contributions to the eigenvectors of carbamazepine dihydrate (FEFNOTO2) calculated for all the 72 phonons with DMACRYS, w99 force field. Regions 1, 2, 3 are respectively in blue, green and yellow on the top bar

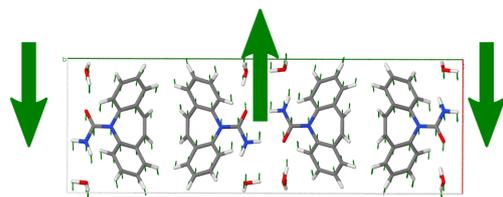


Figure 9.24. Vibration in carbamazepine dihydrate, mode 4 ( $13.4 \text{ cm}^{-1}$ ) calculated with DMACRYS, w99 force field

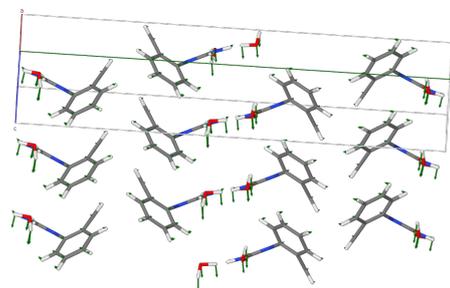


Figure 9.25. Vibration in carbamazepine dihydrate, mode 12 ( $516 \text{ cm}^{-1}$ ) calculated with DMACRYS, w99 force field

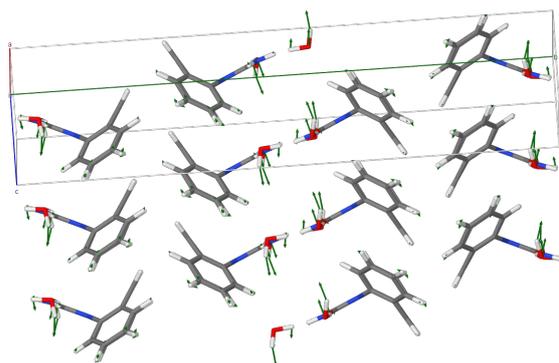


Figure 9.26. Vibration in carbamazepine dihydrate, mode 20 ( $82.2\text{ cm}^{-1}$ ) calculated with DMACRYS, w99 force field

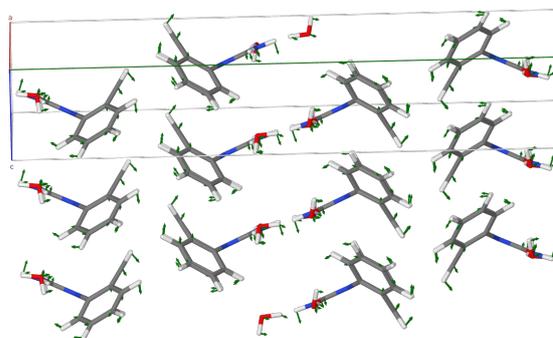


Figure 9.27. Vibration in carbamazepine dihydrate, mode 24 ( $86.0\text{ cm}^{-1}$ ) calculated with DMACRYS, w99 force field

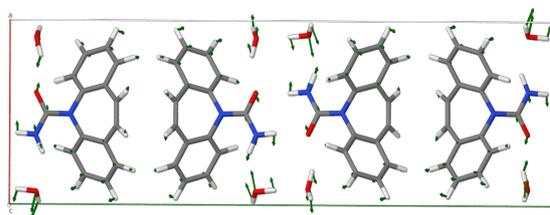


Figure 9.28. Vibration in carbamazepine dihydrate, mode 35 ( $118.4\text{ cm}^{-1}$ ) calculated with DMACRYS, w99 force field

## 9.3 EIGENVECTOR AGREEMENT BETWEEN DMACRYS AND CASTEP

In this section we discuss the agreement between the CASTEP calculations and the DMACRYS eigenvectors for paracetamol monohydrate and nitrofurantoin form II, using the same methods employed in chapter 8.4.

As expected, because of the great separation in energy there is almost no mixing between the DMACRYS phonons of the three regions in the description of the CASTEP features (Figure 9.29a).

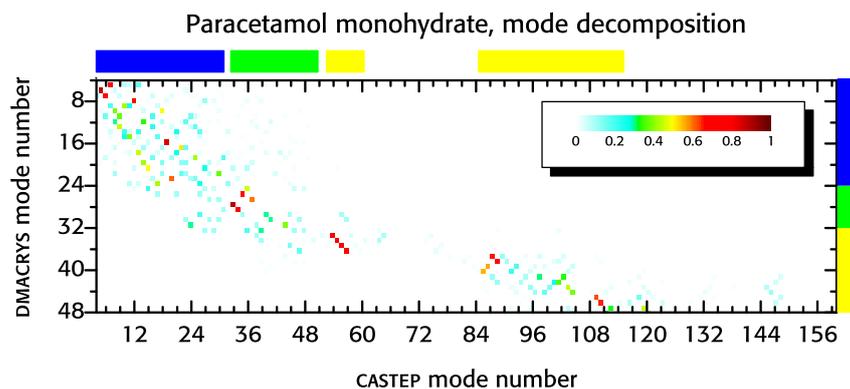
The main difference with the DMACRYS vectors is the presence of internal vibrations: we cannot describe some of the highest vibration as “pure” water molecule vibration in a field of frozen API molecules, because there are internal vibrations with the same frequency, and intermixing is therefore possible. Nonetheless, some of the vibrations can still be described as pure water modes in the nitrofurantoin hydrate calculation (see section 9.3.2).

Another, more complete visual analysis is presented in Figure 9.29, with the projection, for both systems, of all the DMACRYS eigenvectors on the first 160 CASTEP eigenvectors. We can see that the diagonal agreement is not particularly good in region 1, but it improves for region 2 and region 3 modes (Figure 9.29b). The separation between modes belonging to different regions is most evident for the water-rotational lattice modes (the highest 36–48 DMACRYS modes for both systems): even if there is no mode with 100 % agreement, each DMACRYS eigenvector projects onto a small set of CASTEP eigenvectors in a small energy range.

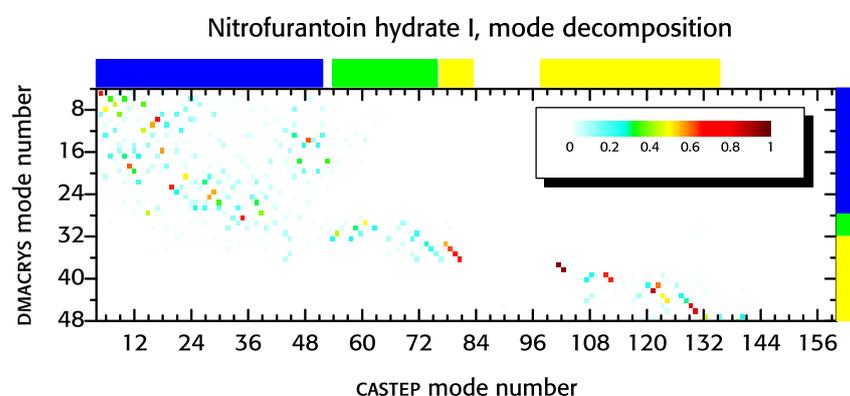
### 9.3.1 Paracetamol monohydrate

The rigidity of CASTEP eigenvectors in the lower part of the spectrum is high (Figure 9.30): the 20th phonon is the first mode showing a rigidity lower than 75 %, with a frequency of  $104.03 \text{ cm}^{-1}$ , a value higher than the equivalent anhydrate forms (where the first mode with less than 75 % rigidity occurs at approximately  $80 \text{ cm}^{-1}$ ).

The agreement between the two methods in the lower frequency part of the spectrum is not very good except for some of the modes, and there is considerable mixing among some of the eigenvectors.



(a) Projection of the DMACRYS eigenvector on the first 160 CASTEP eigenvectors, PBE functional



(b) Projection of the DMACRYS eigenvector on the first 160 CASTEP eigenvectors, PBE/DFT-D functional

Figure 9.29. Projection of the DMACRYS eigenvectors on the CASTEP eigenvector for paracetamol hydrate and nitrofurantoin hydrate form II. The three regions 1–3 are highlighted in blue, green and yellow, respectively

There is a good agreement between CASTEP and DMACRYS eigenvectors for each of the modes in the intermediate and high energy region (modes 32–47 in DMACRYS). There is a good 1:1 match for the high frequency modes involving rotation of the water molecules.

The best agreement between the two methods is between CASTEP modes 53–56 and DMACRYS modes 33–36, in which the highest energy hydrogen bonds are stretched between water and the API.

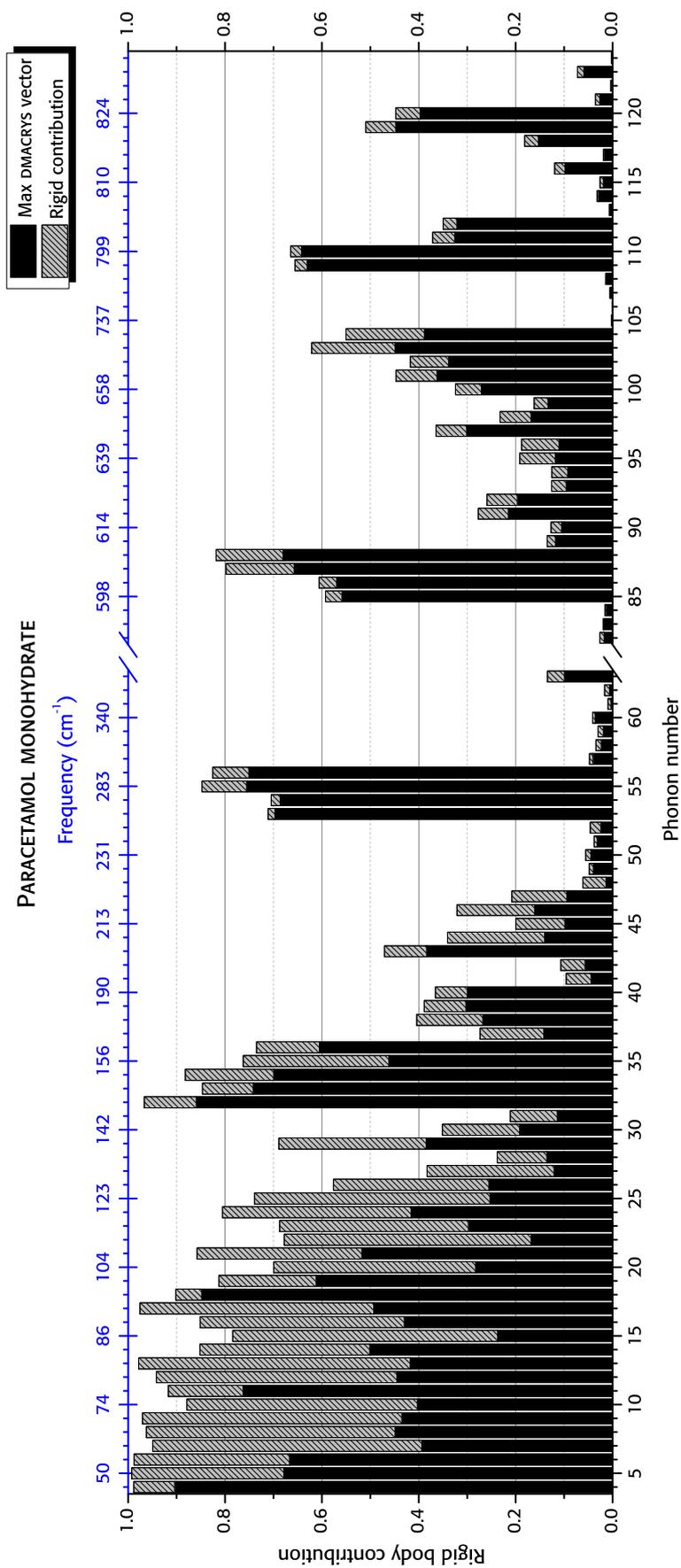


Figure 9.30. Breakdown of the rigid body contributions to the eigenvectors of paracetamol monohydrate calculated with CASTEP 5, PBE functional, and the most important contribution calculated with DMACRYS, w99 force field

### 9.3.2 Nitrofurantoin hydrate form I

As we mentioned in the introduction, the nitrofurantoin hydrate calculation is characterised by a incompletely translational acoustic phonons (about 22 % internal vibration), which might affect the reliability of the rigid body analysis.

The fraction of the rigid molecule displacement in the lower part of the spectrum (Figure 9.31) is characterised by some low energy internal vibrations (the first with a frequency of  $37.97\text{ cm}^{-1}$ ), a lower value than in the CASTEP calculations for either of the neat forms of nitrofurantoin.

Once again, the agreement between the two methods in the lower frequency part of the spectrum is not very good except for some of the modes, but the agreement improves for modes 54–64 (in the frequency range  $200\text{ cm}^{-1}$  to  $240\text{ cm}^{-1}$ ) and there is excellent agreement for the water rotational modes at higher frequencies.

As for paracetamol monohydrate, the first modes to have a perfect 1:1 match are modes 73–80 in region 2, which perfectly match (DMACRYS modes 29–36).

## 9.4 CONCLUSIONS

In this chapter, we have analysed the effect of water on the lattice modes and frequencies of the hydrate systems we have studied.

The main observation we obtain from our analysis is the clear separation into frequency regions according to the importance of water contributions to the vibrational mode. Region 3 is particularly well defined, especially regarding the rotational modes in water due to its very low moment of inertia. Another important parameter is the type of interaction, and the frequency increases with the number of hydrogen bonds probed by the vibration. The frequency is therefore very low when no hydrogen bond is involved (as in the low energy lattice modes in carbamazepine dihydrate), and increases along with the complexity of the vibration, where the mutual movement of the molecules affect an increasing number of close intermolecular contacts.

The number of modes in region 2 depends on the number of stretching water-API hydrogen bonds; due to the type of the vibration considered,

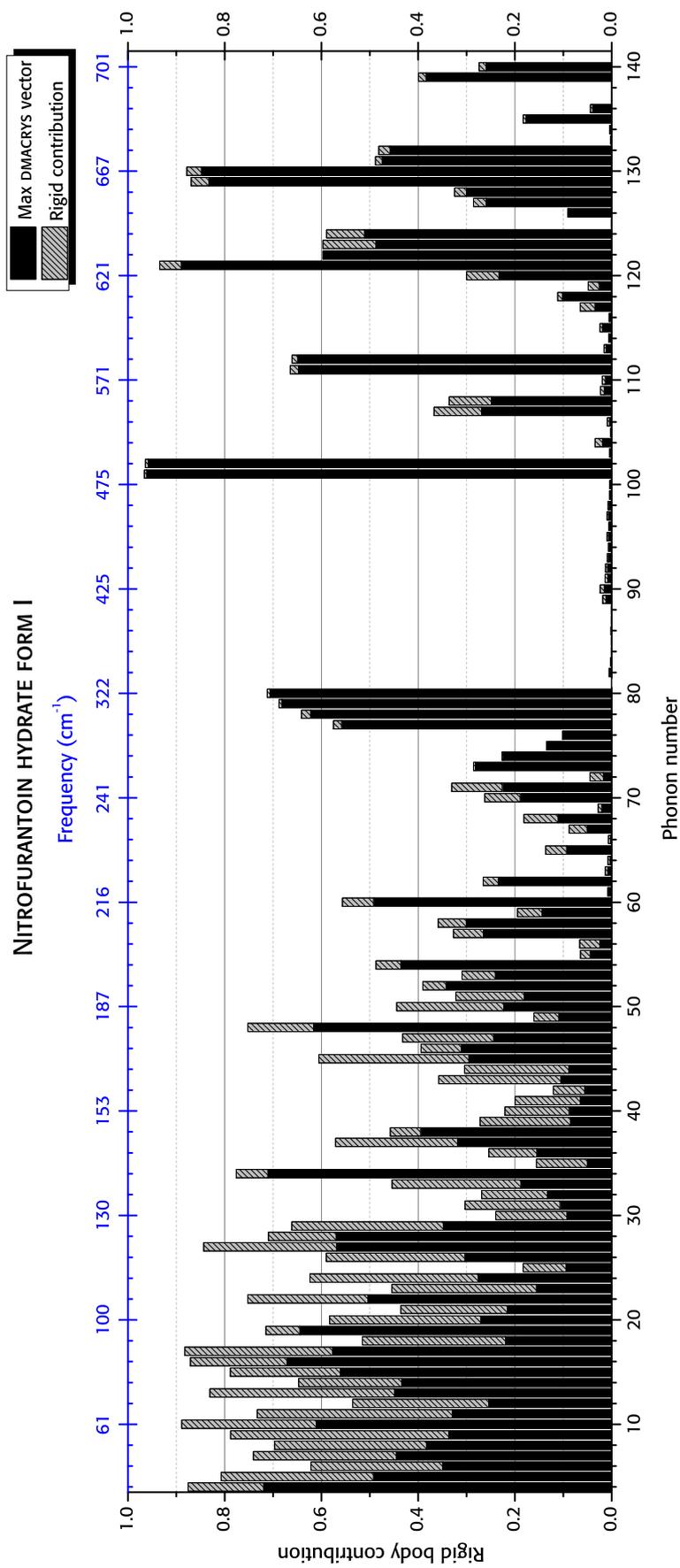


Figure 9.31. Breakdown of the rigid body contributions to the eigenvectors of nitrofurantoin hydrate form I (HAXBUD01) calculated with CASTEP 5.5, dispersion corrected PBE, and the most important contribution calculated with DMACRYS, w99 force field

this is related only to the translation of the water molecules. It is easy to predict the number of modes in this intermediate region just by looking at the structure, and consequently considering all the possible symmetry adapted combinations. This observation explains why in some of the systems (paracetamol monohydrates and both polymorphs of nitrofurantoin hydrates) all the water translations stretch a water-API interaction, while in the case of theophylline monohydrate only water translations in one direction are important.

From the eigenvector comparison between DMACRYS and CASTEP we see that the agreement is not exceedingly good for the modes at lower frequency, but the agreement gradually improves for the modes in region 2 and region 3. The agreement at higher frequency seems to suggest that the description of the vibrations is better at higher frequencies: this is probably due to the higher energy involved in these vibrations, and in the lower relative error in the assessment of the intermolecular forces.

## THE EFFECT OF POLARISATION

---

**I**N THIS CHAPTER we repeated the calculations reported in chapter 7, but including effect of polarisation induced by the surrounding molecules within the crystal. We used two different approaches:

**FULL MINIMISATION (METHOD 1):** evaluation of the electrostatic distribution of each molecule surrounded by the charges, and its use for the DMACRYS minimisation and the “standard” calculation of the phonon frequencies and intensities, according to the procedure of diagram 5.1;

**INTENSITIES ONLY (METHOD 2):** minimisation performed using a normal DMACRYS calculation of the phonon frequencies and eigenvectors. The absorption intensities are not calculated with RUDOLPH standard tool (estimate of the cell dipole moment derivative using the rotational fraction of each phonon vector), but with the recalculation of the molecular electrostatic distribution according to each phonon displacement (diagram 5.4), and estimation of the dipole derivative by interpolation on all the conformations.

None of the above methods is completely free of drawbacks. The parameters in the Williams forcefield (as in the case of any other forcefield) are parameterised together with an electrostatic method of choice: each individual atom-atom energy is modelled as a sum of terms

$$E_{jk} = -A_{jk}r_{jk}^{-6} + B_{jk}e^{-C_{jk}r_{jk}} + q_jq_kr_{jk}^{-1}, \quad (10.1)$$

and the values for the parameters  $A$ ,  $B$  and  $C$  depends on the charges used for their fit. We can therefore consider that the effect of polarisation is somewhat included in the optimisation of the parameters, and by adding the polarisation we would be counting these effects twice (using method 1).

With method 2 we rely on the “self consistent” frequencies and eigenvectors and recalculate the intensities by considering the dipoles as they would be inside the crystal, and their change with the displacement associated with each eigenvector perturbation from the equilibrium. With this method, anyway, for each perturbation of the molecular position the changes in electrostatic were not taken into account to begin with, and the change in dipole (related to the intensity) might reflect artificial effects.

## 10.1 DIPOLE MOMENT VARIATIONS IN THE CRYSTAL

The changes in molecular dipole moments in the crystal are reported in Table 10.1. Since the electronic distribution of the molecules is not affected by the type of forcefield used, all of the reported structures are from the w99 determinations, including a 30 Å cloud of surrounding molecules.

The first thing we notice is that we have a molecular electric dipole enhancement for every molecule we considered. The dipole changes display a considerable variation between different structure, from a minimum of 0.75 % for the API molecules in paracetamol hydrate to 50.4 % in carbamazepine form III.

The dipole enhancement in water is much less variable than in the API molecules varying from 16 % to 28 % (around 20 % for most of the structures). The enhancement is somewhat reduced to what is observed in liquid water [174], ice [173] and water clusters of various sizes [222]: these calculations find an increase of 40 % in the value of the molecular dipole. The dipole magnitude in the water molecules depends on their environment within the crystal: if water molecules are arranged in pocket-like structures (nitrofurantoin hydrates, paracetamol monohydrate) the dipole moment change are smaller compared to systems where water–water interactions are more significant, as in carbamazepine dihydrate (close to 30 % enhancement).

The findings for the molecular dipole enhancement in molecular crystals do not always agree with measurements of charge density analysed from X-ray determinations: in two comprehensive studies [172, 223] Spackman summarises the measured dipole moments in molecular

Table 10.1. Dipole moment enhancement in the molecular crystals we studied, calculated with GAUSSIAN03, B3LYP 6-31G\*\*, reported for all the molecules in the asymmetric unit cell, with and without the polarisation due to the cloud charges. Molecule orientation is in DMACRYS Cartesian coordinates ( $z \parallel c$ ,  $y \in bc$  plane,  $x \perp yz$  plane and forming a right-handed axis system). Cloud radius = 30 Å

Crystal form	Molecule	Unpolarised dipole moment / Debye			Polarised dipole moment / Debye			Magnitude change / %		
		x	y	z	Magnitude	x	y		z	Magnitude
Benzoic acid	API	-1.78	0.22	1.96	2.66	-1.95	0.21	2.12	2.89	8.64
Anhydrous theophylline	API	-3.28	1.01	-0.67	3.53	-3.50	1.70	0.22	3.90	10.48
Paracetamol form I	API	-2.47	0.21	4.14	4.83	-3.46	0.25	5.93	6.87	42.23
Paracetamol form II	API	-2.02	-4.41	-0.54	4.89	-3.33	-5.99	-0.97	6.92	41.51
Paracetamol form III	API 1	-4.54	2.32	0.77	5.16	-6.36	3.53	0.67	7.30	41.47
	API 2	4.58	-2.23	-0.08	5.09	6.47	-5.99	-3.51	7.36	44.59
Methyl Phenyl-acetamide I	API	-0.29	-3.35	-0.55	3.41	0.06	-4.64	-0.81	4.71	38.12
Methyl Phenyl-acetamide II	API	-3.29	0.71	0.29	3.38	-4.91	0.91	0.36	5.01	48.22
Fluoro Phenyl-acetamide	API	0.35	2.03	3.51	4.07	0.11	1.97	4.09	4.54	11.54
Chloro Phenyl-acetamide	API	3.73	1.16	-1.85	4.32	5.44	1.15	-1.77	5.84	7.35

— continues on next page

Table 10.1. — continued from previous page

Crystal form	Molecule	Unpolarised dipole moment / Debye			Polarised dipole moment / Debye			Magnitude change / %		
		x	y	z	Magnitude	x	y		z	Magnitude
Bromo Phenyl-acetamide	API	-3.75	-0.43	1.61	4.10	-5.36	-0.21	1.55	5.59	36.34
Iodo Phenyl-acetamide	API	-0.05	0.41	4.38	4.4	0.11	0.22	6.36	6.36	44.54
Nitrofurantoin form $\alpha$	API	-1.80	6.33	-2.72	7.13	-1.92	8.66	-3.20	9.43	32.25
Nitrofurantoin form $\beta$	API	6.33	2.70	-4.19	8.06	9.07	2.90	-4.98	10.75	18.52
Carbamazepine form I	API 1	-1.87	-3.02	2.05	4.10	-2.84	-2.99	3.90	5.68	38.53
	API 2	1.42	-3.34	-1.99	4.14	2.28	-3.59	-1.66	4.56	10.14
	API 3	1.36	3.63	-1.49	4.15	-2.07	4.43	-2.13	5.33	28.43
	API 4	1.40	-0.10	3.84	4.08	2.37	0.33	4.03	4.68	14.70
Carbamazepine form II	API	-3.52	-2.05	-0.66	4.13	-3.34	-2.50	-1.43	4.41	6.77
Carbamazepine form III	API	-2.96	1.78	2.26	4.13	-4.64	2.19	3.48	6.21	50.36
Carbamazepine form IV	API	-1.38	-2.85	2.64	4.12	-1.53	-3.94	3.96	5.80	40.77

— continues on next page

Table 10.1. — continued from previous page

Crystal form	Molecule	Unpolarised dipole moment / Debye			Polarised dipole moment / Debye			Magnitude change / %		
		x	y	z	Magnitude	x	y		z	Magnitude
Theophylline hydrate	API 1	-2.12	-0.46	-2.81	3.56	-2.76	-0.44	-3.62	4.58	28.65
	API 2	2.12	0.44	2.80	3.55	2.73	0.42	3.66	4.59	29.29
	WATER 1	1.63	0.72	1.18	2.14	2.00	0.92	1.46	2.65	23.83
	WATER 2	1.47	-0.81	-1.33	2.14	1.79	-1.02	-1.63	2.63	22.89
Paracetamol monohydrate	API	-0.14	0.45	3.92	3.95	-0.14	0.45	3.95	3.98	0.75
	WATER	-0.04	0.09	2.07	2.07	-0.05	0.11	2.48	2.49	20.28
Paracetamol trihydrate	API	-1.24	4.47	-1.08	4.76	-1.70	6.70	-1.21	7.02	47.47
	WATER 1	-0.42	1.52	1.36	2.08	0.50	1.89	1.68	2.58	24.03
	WATER 2	-0.63	1.84	0.53	2.01	-0.78	2.25	0.64	2.47	22.88
	WATER 3	0.04	1.62	1.08	1.95	0.04	1.95	1.37	2.38	22.05
Nitrofurantoin form I	API	-6.61	2.43	2.54	7.49	-9.20	2.98	3.26	10.21	36.31
	WATER	-1.59	1.01	-0.77	2.04	-0.65	-1.79	1.54	2.45	16.66

— continues on next page

Table 10.1. — continued from previous page

Crystal form	Molecule	Unpolarised dipole moment / Debye			Polarised dipole moment / Debye			Magnitude change / %		
		x	y	z	Magnitude	x	y		z	
Nitrofurantoin form II	API	-0.97	6.12	3.41	7.08	-1.12	8.20	3.52	9.00	27.11
	WATER	-0.54	-1.49	1.27	2.03	-0.65	-1.79	1.54	2.45	20.07
Carbamazepine dihydrate	API	-3.85	1.72	0.27	4.22	-6.25	0.91	0.42	6.33	50.00
	WATER 1	-1.20	0.70	1.48	2.03	-1.58	0.91	1.87	2.61	28.57
	WATER 2	-0.99	-1.47	1.07	2.07	-1.22	-1.82	1.32	2.56	23.67

Table 10.1. (Continued from previous page) Dipole moment enhancement in the molecular crystals we studied, calculated with GAUSSIAN03, B3LYP 6-31G\*\*, reported for all the molecules in the asymmetric unit cell, with and without the polarisation due to the cloud charges. Molecule orientation is in DMACRYs Cartesian coordinates ( $z \parallel c$ ,  $y \in bc$  plane,  $x \perp yz$  plane and forming a right-handed axis system). Cloud radius = 30 Å

crystals, mostly changed by a factor in the range of  $-50\%$  to  $100\%$  compared to isolated molecule calculations. Furthermore, a series of periodic DFT calculations in the same article showed that dipole enhancements larger than  $75\%$  do not seem to be backed up by theoretical calculation, and might be due to errors in the refinement procedure of the experimental charges.

In another article by the same author [224] it is suggested that the use of an iterative self consistent method similar to the one we implemented can give a close estimation of the dipole moment of a molecule in the crystal. The electronic distribution around a molecule is described as dipoles, centred on the molecule itself, and summed up to infinity using the Ewald method [225]; the electronic distribution is calculated in the obtained electric field, and it is repeated until self consistency. This method produces an enhancement of the molecular dipole for a set of molecules (unfortunately, none of them present in our set) varying from  $10\%$  to  $40\%$ , very similar to what was found in periodic DFT calculations and to what we calculate for our systems.

The electronic distribution of a molecule is dependent on the crystal packing: for carbamazepine the dipole enhancement varies from  $7\%$  to  $50\%$  depending on the polymorph, and it can of course vary for symmetry-independent molecules in the same unit cell (see carbamazepine form I, with  $Z' = 4$ ).

The percentile variation in dipole is evident is the nitrofurantoin polymorphs as well, with the enhancement in form  $\alpha$  almost double of what we have for form  $\beta$ .

The most peculiar behaviour is the one related to paracetamol: the molecular dipole of the isolated molecule is roughly constant ( $4.9\text{ D}$ ) for all the conformations in the neat crystals, but it is reduced by almost  $20\%$  in the molecular conformation in the hydrate, ( $4.0\text{ D}$ ) and it remains almost constant when we include the condensed state.

Our treatment is a first approximation of the electronic properties and it cannot take into account charge transfers or quantum-mechanical effects. The most important contribution we are not considering is due to hydrogen bonds: by its definition [226, 227] it cannot be fully accounted for by an electrostatic approximation [228].

## 10.2 EFFECTS OF THE FULL MINIMISATION ON THE PHONON MODES

In this section we briefly discuss the differences between the experimental and unpolarised spectra, and we present a comparison between the eigenvectors in the polarisation corrected and uncorrected calculations.

For all our calculations we analysed the eigenvectors of the polarised calculation with those from the unpolarised results, by calculating the square of the scalar product between the two complete sets of vectors; once again, this approach is meaningful if the final geometries are comparable to each other, which is the case for all of our calculations. The results are visually presented in the form of a matrix, with eigenvectors from each calculation ordered by energy on one of the axes. If the two sets are the same, we see a diagonal trend (as for nitrofurantoin  $\beta$  in Figure 10.3a); a non-diagonal 1:1 match highlights a different energy order in the vibrational modes.

### 10.2.1 Structural changes

The electrostatic part is a very important contribution to the total lattice energy in DMACRYS, but it does not seem to affect the final geometry of the crystal. For most of the systems the introduction of the polarisation correction seems to stabilise the structure to the initial geometry, as the  $F$  values are reduced in almost all the cases (see Table 10.2) and we do not notice significant variation between minimised structures with the two methods (if we exclude the hydrates, which were the systems with the biggest changes among our set).

### 10.2.2 Benzoic acid

The overall agreement with the experimental spectrum (Figures 10.1 and 10.2 for w99 and FIT forcefields respectively) is visibly reduced: the frequency shift to higher frequencies with the inclusion of polarisation worsens the agreement with the experimental spectrum compared to the unpolarised frequency, and the shift is particularly strong for all the

System	FIT forcefield		w99 forcefield	
	Unpol.	Pol.	Unpol.	Pol.
Benzoic acid	28.17	18.87	23.21	20.60
Anhydrous theophylline	8.45	24.74	14.09	35.25
Paracetamol form I	19.24	24.02	10.98	12.10
Paracetamol form II	17.47	23.19	13.39	17.63
Paracetamol form III	38.53	16.58	42.05	24.66
Nitrofurantoin $\alpha$	33.65	27.20	47.07	26.78
Nitrofurantoin $\beta$	33.55	16.02	30.14	8.56
Carbamazepine form I	4.57	6.70	4.34	4.14
Carbamazepine form II	1.58	3.41	1.40	0.97
Carbamazepine form III	14.48	14.28	10.75	13.35
Carbamazepine form IV	9.25	5.79	12.57	4.74
Paracetamol monohydrate	26.83	52.89	47.64	71.72
Paracetamol trihydrate	15.97	37.24	6.05	20.06
Nitrofurantoin form I	286.24	241.53	118.24	124.39
Nitrofurantoin form II	85.90	53.96	15.42	11.31
Theophylline hydrate	151.66	186.41	218.52	222.69
Carbamazepine dihydrate	22.32	18.27	28.95	25.74

Table 10.2. Variation of the structure upon minimisation with DMACRYS, FIT and w99 forcefields, as described by the  $F$  value with (Pol.) and without (Unpol.) considering the effect of polarisation

frequencies over  $100 \text{ cm}^{-1}$ . The eigenvectors (Figure 10.3) are almost unvaried except for two modes with the FIT forcefield; there still mainly 1:1 correspondence of the eigenvectors, even if some of the energies are reshuffled as a consequence of polarisation for some of the modes.

The highest frequency lattice modes are shifted up to  $40 \text{ cm}^{-1}$  for the highest energy IR active modes, and even more ( $87 \text{ cm}^{-1}$ ) for the IR-forbidden (Raman active) modes ( $288 \text{ cm}^{-1}$  vs  $201 \text{ cm}^{-1}$ ).

### 10.2.3 Anhydrous theophylline

The frequency shift in the case of theophylline improves the position of the strongest absorption features, but the overall agreement over the entire spectrum is not improved: we can also notice that the polarised spectra from FIT and w99 calculations are much more similar to each other than the corresponding unpolarised spectra (Figures 10.4 and 10.5). The highest energy modes, where there is a stretch of the hydrogen bonds, are shifted by almost  $40 \text{ cm}^{-1}$  (from  $150 \text{ cm}^{-1}$  to  $190 \text{ cm}^{-1}$ ).

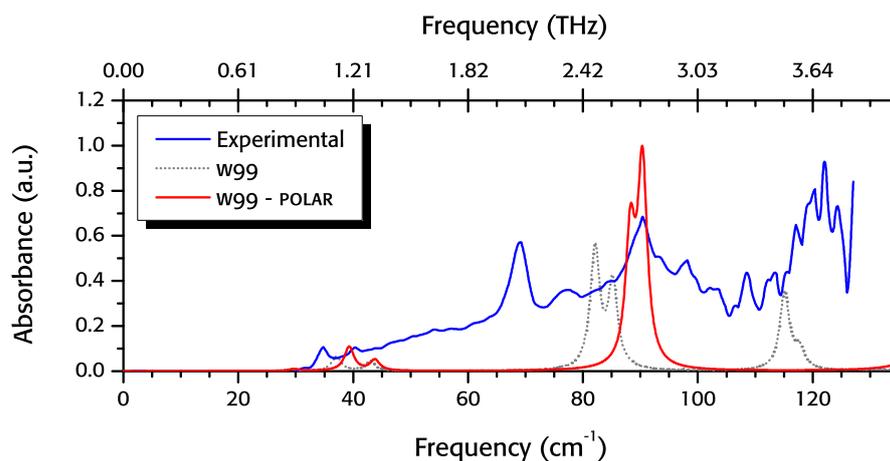


Figure 10.1. Experimental (100 K) and DMACRYS, w99 forcefield spectrum of benzoic acid. The unpolarised and polarised spectra are reported

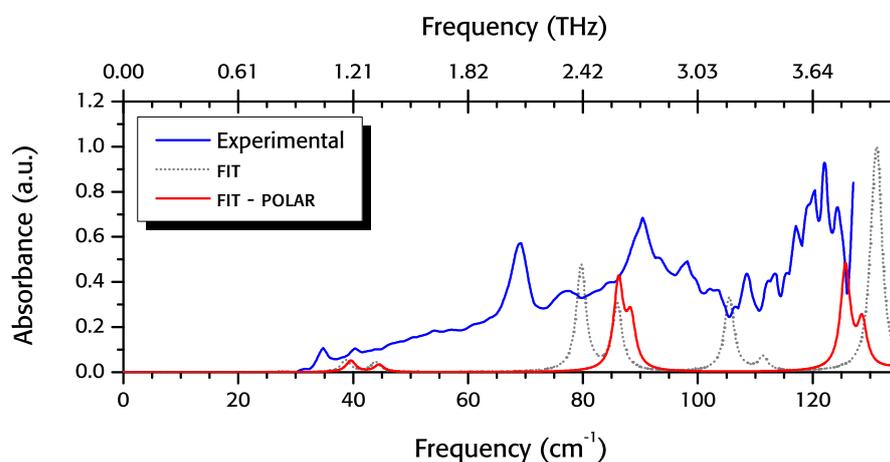


Figure 10.2. Experimental (100 K) and DMACRYS, FIT forcefield spectrum of benzoic acid. The unpolarised and polarised spectra are reported

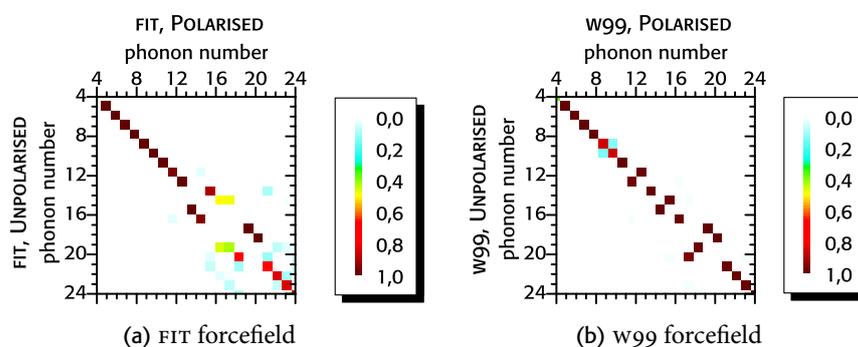


Figure 10.3. Comparison of the DMACRYS eigenvectors of benzoic acid, polarised and unpolarised methods. The square of the magnitude of the projections of each vector upon each other is plotted

The eigenvectors are almost unaffected by the inclusion of polarisation, except for modes 13–15 at a frequency of  $\approx 70 \text{ cm}^{-1}$ , where the eigenvectors of the polarised calculation are a mixture of those from the unpolarised eigenvectors (Figure 10.6).

#### 10.2.4 Theophylline monohydrate

We report only one of the polarisation calculations from the four symmetry-corrected structures discussed in chapter 7. There is an overall improvement in the agreement with the experimental spectrum, due to a better match of the strong absorption features at  $60 \text{ cm}^{-1}$  and  $85 \text{ cm}^{-1}$ . However, there are still no calculated peaks in the region  $45 \text{ cm}^{-1}$  to  $70 \text{ cm}^{-1}$ , where there are probably absorption features hidden under the wide peak in the measured spectrum. We can notice how the peaks obtained with the calculations with FIT and w99 forcefields are much more similar than the corresponding unpolarised calculation. For the lattice modes in the experimental range (modes 4–28), up to  $120 \text{ cm}^{-1}$  the frequency shift is less than  $12 \text{ cm}^{-1}$ .

The change in the eigenvectors is particularly evident in the modes 12–28 in FIT (Figure 10.9), where there is strong non-diagonal component in the polarised-unpolarised eigenvectors (Figure 10.9). The eigenvectors with the w99 forcefield are almost untouched by the polarisation correction.

#### 10.2.5 Paracetamol form I

Without taking into account the frequency shift it would appear that the agreement in the region  $90 \text{ cm}^{-1}$  to  $120 \text{ cm}^{-1}$  for the w99 determination is improved (Figure 10.10). However, the peaks in the region  $40 \text{ cm}^{-1}$  to  $60 \text{ cm}^{-1}$  are completely unaccounted for, due to the shift to higher frequency in the polarised calculation. We therefore have to conclude that the agreement at higher frequency is probably fortuitous match (unless we take into account the possibility of intramolecular vibrations, which is ruled out by the CASTEP analysis of the eigenvectors, see chapter 8). The FIT calculation (Figure 10.11) suffers from the same problems, without showing the same level of agreement, except for one mode at  $106 \text{ cm}^{-1}$ .

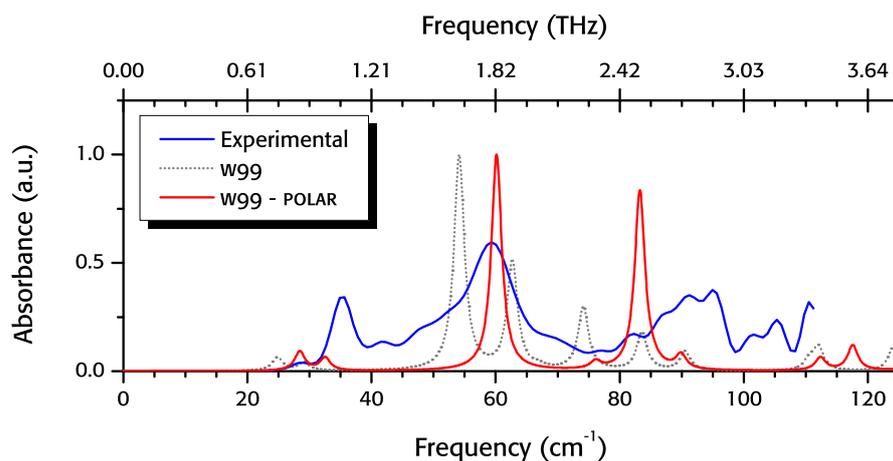


Figure 10.4. Experimental (80 K) and DMACRYS, w99 forcefield spectrum of anhydrous theophylline

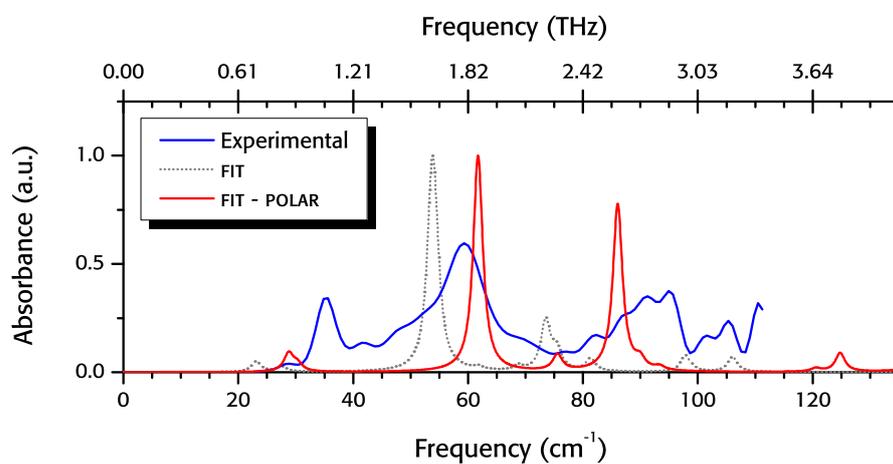


Figure 10.5. Experimental (80 K) and DMACRYS, FIT forcefield spectrum of anhydrous theophylline

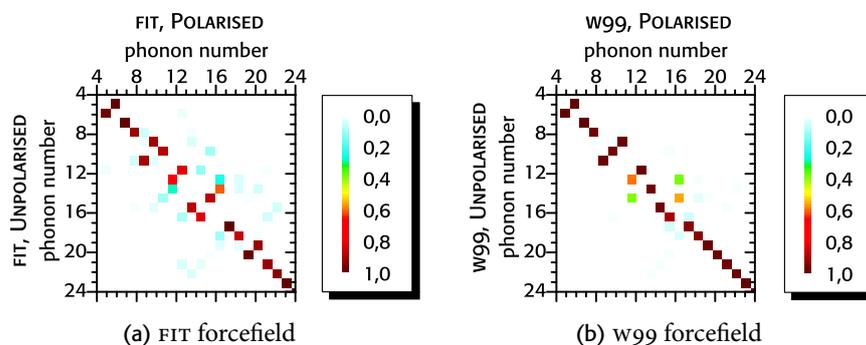


Figure 10.6. Comparison of the DMACRYS eigenvectors of anhydrous theophylline, polarised and unpolarised methods. The square of the magnitude of the projections of each vector upon each other is plotted

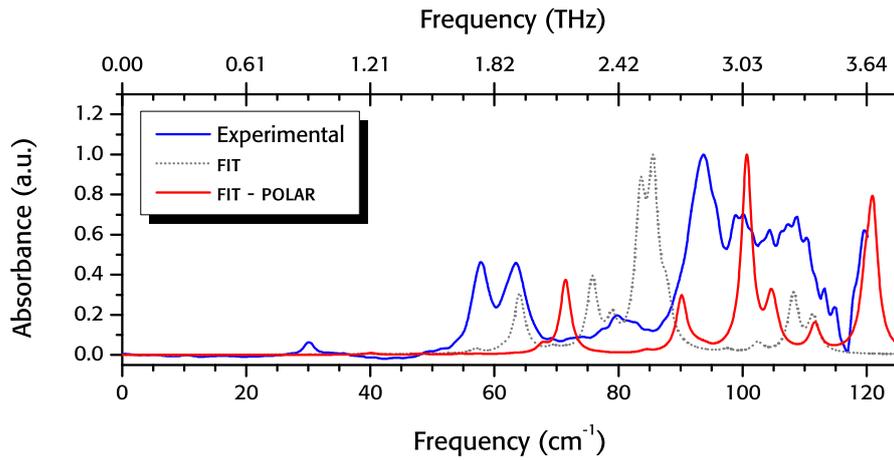


Figure 10.7. Experimental (100 K) and DMACRYS, w99 forcefield spectrum of theophylline monohydrate

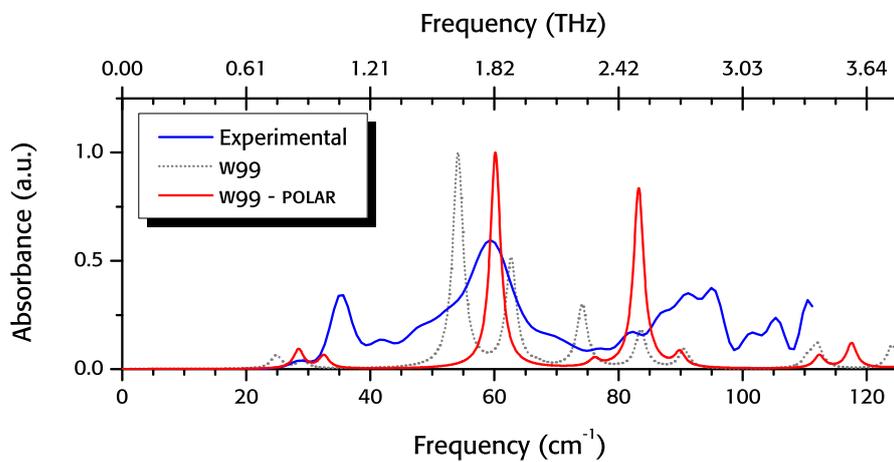


Figure 10.8. Experimental (100 K) and DMACRYS, FIT forcefield spectrum of theophylline monohydrate

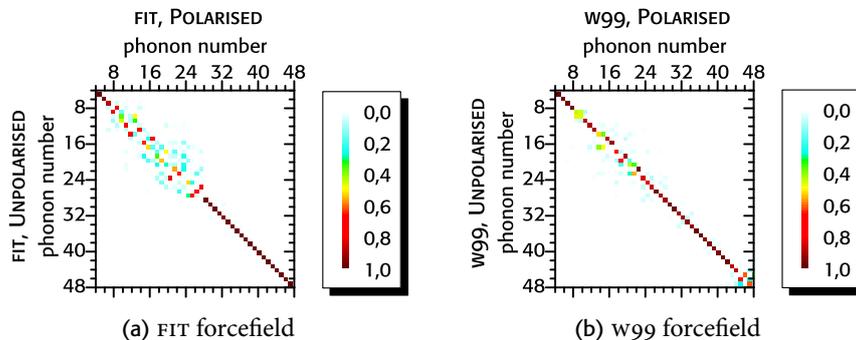


Figure 10.9. Comparison of the DMACRYS eigenvectors of theophylline hydrate, polarised and unpolarised methods. The square of the magnitude of the projections of each vector upon each other is plotted

The eigenvectors for both the forcefields (Figure 10.12) are almost completely unchanged by the polarisation, again except for the energy order of a few modes.

#### 10.2.6 Nitrofurantoin form $\beta$

The spectra derived from the use of the polarisation correction with FIT and w99 forcefields are reported in Figures 10.13 and 10.14. There are only minor differences between corresponding polarised and unpolarised spectra. The small changes are even more evident in the eigenvector comparison, with hardly any visible change (Figure 10.18).

#### 10.2.7 Nitrofurantoin hydrate, form II

The polarisation corrected spectrum calculated with the FIT forcefield (Figure 10.17) is very flat up to  $113\text{ cm}^{-1}$ , with all of the features in the region  $40\text{ cm}^{-1}$  to  $100\text{ cm}^{-1}$  much weaker than their experimental counterpart, and a general loss of agreement compared to the non-polarisation correction calculation.

The spectra from the w99 calculation (Figure 10.16) has a much better agreement with experiment than the FIT calculation, but it is worse if compared with the corresponding unpolarised w99 calculation. In particular, except for the highest frequency absorption peak (a doublet at  $107\text{ cm}^{-1}$ ), which is almost superimposable with an experimental feature, we notice a loss of agreement in the details between  $70\text{ cm}^{-1}$  and  $100\text{ cm}^{-1}$ .

Nitrofurantoin hydrate form II has 16 molecules in the unit cell, which leads to into a higher number of eigenvectors (Figure 10.18). There is a strong 1:1 match for most of the modes except between modes 40–60. Within the structure of this crystal form there is a complicated four-atom hydrogen bond interaction between the API and the water molecule (see Figure 6.11, page 84). The eigenvectors perturbing this interaction are mainly between modes 28–50 (see for example Figure 9.21, page 193); the reason behind the strong non-diagonality in Figure 10.18 is probably the small differences between the forces in different modes, which are affected by small changes in the electrostatic distribution.

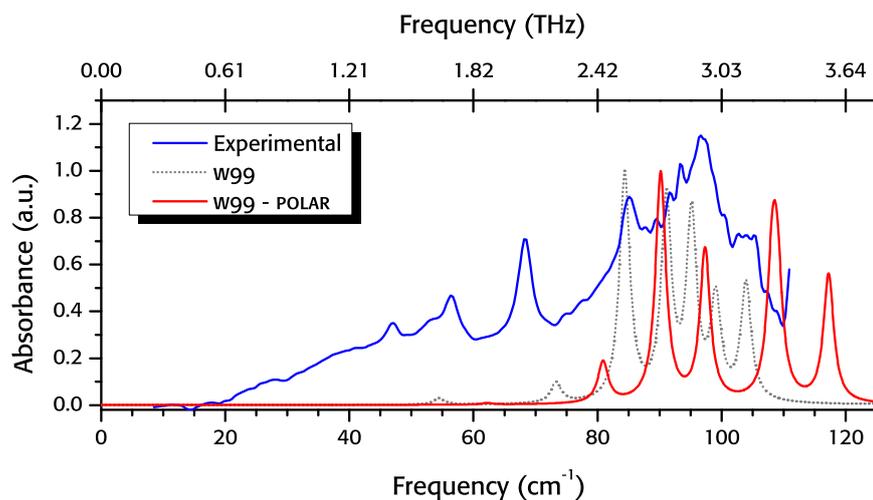


Figure 10.10. Experimental (29 K) and DMACRYS, w99 forcefield spectrum of paracetamol form I

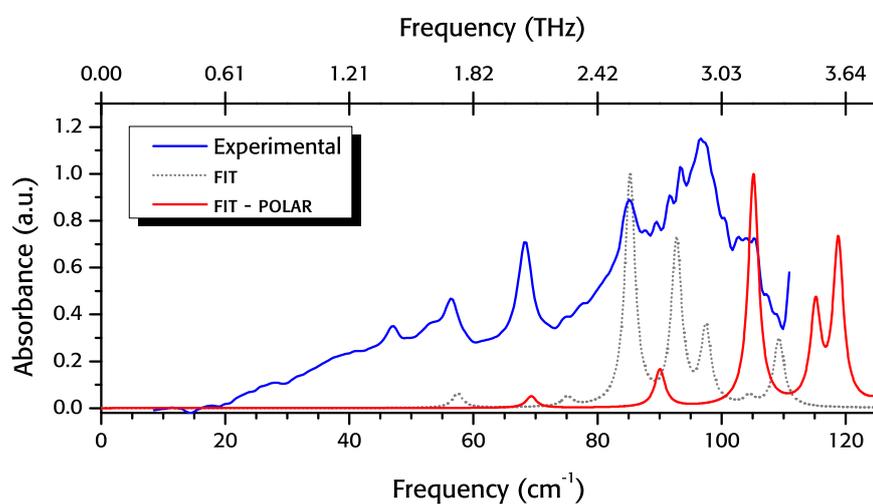


Figure 10.11. Experimental (29 K) and DMACRYS, FIT forcefield spectrum of paracetamol form I

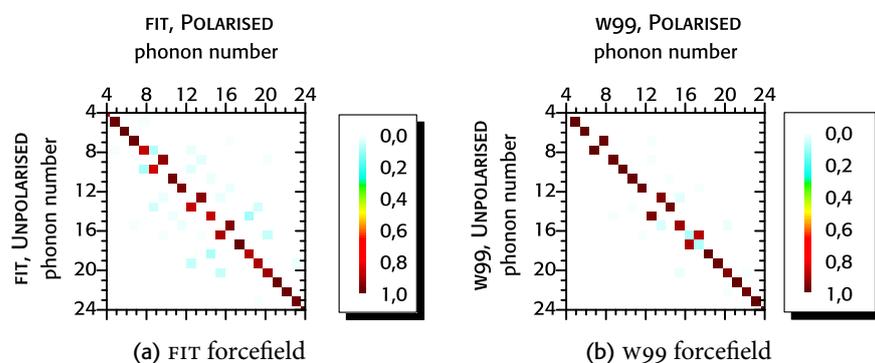


Figure 10.12. Comparison of the DMACRYS eigenvectors of paracetamol form I, polarised and unpolarised methods. The square of the magnitude of the projections of each vector upon each other is plotted

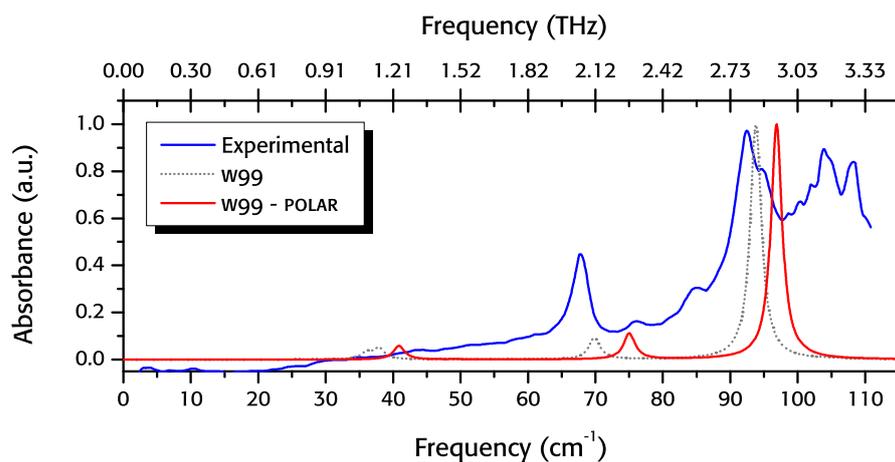


Figure 10.13. Experimental (29 K) and DMACRYS, w99 forcefield spectrum of nitrofurantoin form  $\beta$

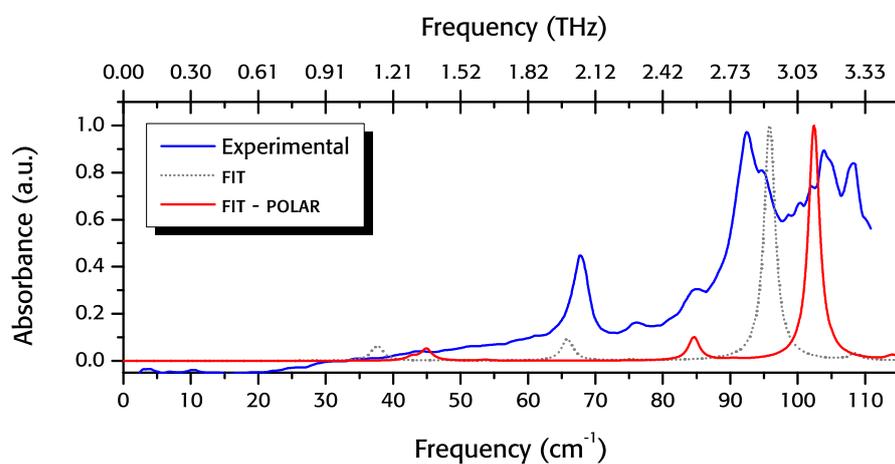


Figure 10.14. Experimental (29 K) and DMACRYS, FIT forcefield spectrum of nitrofurantoin form  $\beta$

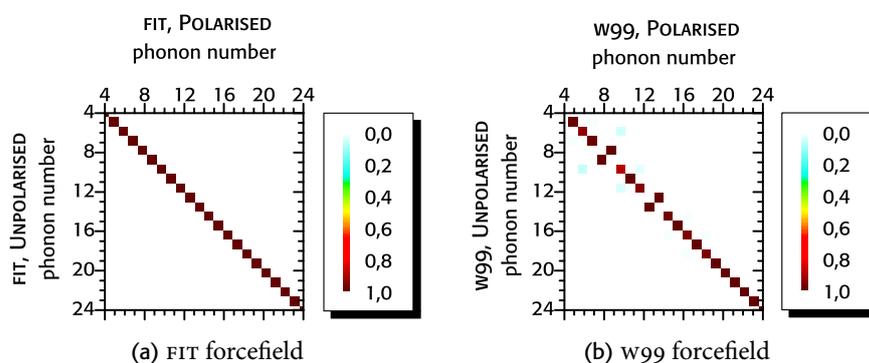


Figure 10.15. Comparison of the DMACRYS eigenvectors of nitrofurantoin  $\beta$ , polarised and unpolarised methods. The square of the magnitude of the projections of each vector upon each other is plotted

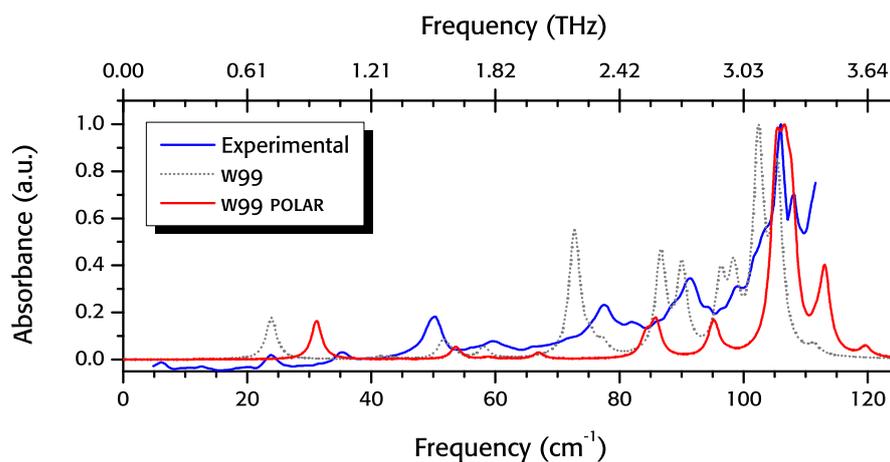


Figure 10.16. Experimental (80 K) and DMACRYS, w99 forcefield spectrum of nitrofurantoin hydrate form II

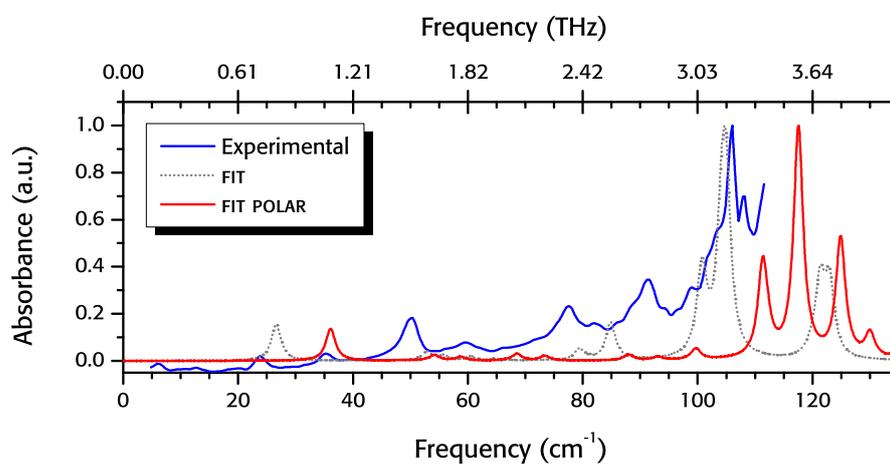


Figure 10.17. Experimental (80 K) and DMACRYS, FIT forcefield spectrum of nitrofurantoin hydrate form II

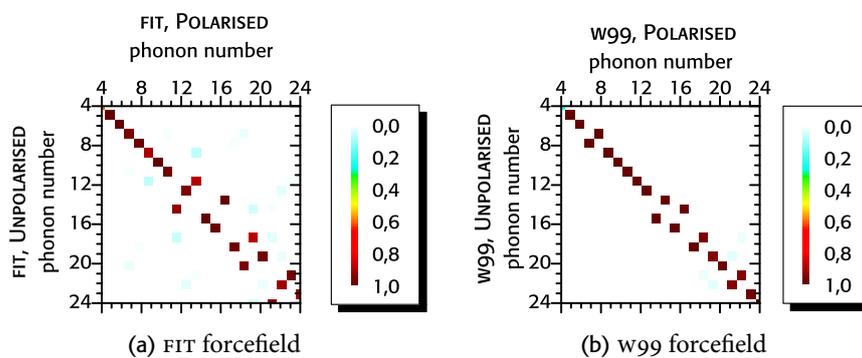


Figure 10.18. Comparison of the DMACRYS eigenvectors of nitrofurantoin II, polarised and unpolarised methods. The square of the magnitude of the projections of each vector upon each other is plotted

### 10.2.8 Carbamazepine form I

The calculated spectra with polarisation correction are reported in Figures 10.19 and 10.20. We can notice how the polarisation in the FIT calculation improves the position of some of the peaks, closer to experimental determination than the corresponding unpolarised calculation, especially for  $40\text{ cm}^{-1}$  to  $90\text{ cm}^{-1}$  (the highest peak is  $7\text{ cm}^{-1}$  higher than the experimental feature, but with a more accurate shape).

The agreement with the experimental peaks is even more evident with the w99 forcefield: the agreement with shape and position is excellent for most of the peaks. The use of a bigger FWHM for the Lorentzian shape (Figure 10.22) can show that the convolution of some of the calculated peaks form the bigger experimental feature.

The eigenvectors in the two w99 calculations (Figure 10.25) do not differ much, but in the FIT calculation there are some off diagonal contribution (even if it mainly involves two unpolarised modes for each polarised).

### 10.2.9 Carbamazepine form III

The changes in the spectra for carbamazepine form III are not as evident as in form I, as we only see small shifts of the absorption features to higher frequencies. Since there was not a particularly good agreement without polarisation, we cannot judge if there is any improvement by the use of polarisation.

The non-diagonality in the eigenvector comparison is very limited (10.25), with almost a perfect 1:1 match between the modes, once again with some energy swaps between a few modes.

### 10.2.10 Conclusions

The polarisation effect is clearly visible in all of the spectra: as we mentioned in the previous chapter, the lattice mode frequencies are very sensitive to small variations in the setup of the calculation, and the electrostatic distribution on the molecule is no exception. As a general observation, the frequencies are always increased in the polarised calculation, as a result of the increase in the magnitude of the charges, which

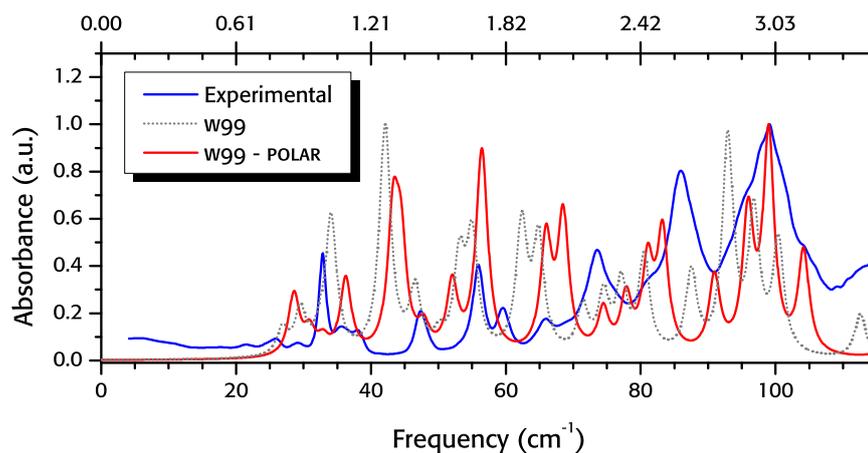


Figure 10.19. Experimental (10 K) and DMACRYS, w99 forcefield spectrum of carbamazepine form I

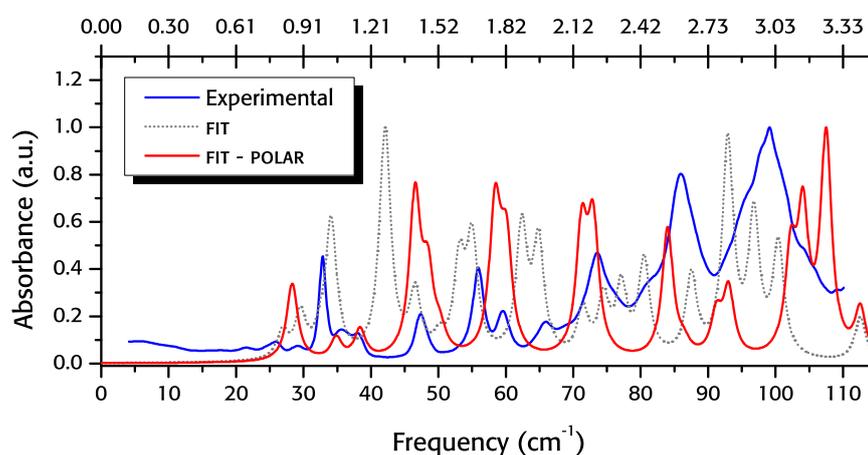


Figure 10.20. Experimental (10 K) and DMACRYS, FIT forcefield spectrum of carbamazepine form I

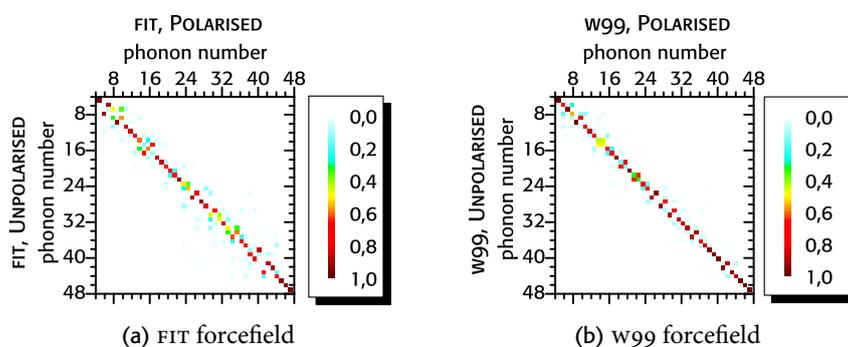


Figure 10.21. Comparison of the DMACRYS eigenvectors of carbamazepine form I, polarised and unpolarised methods. The square of the magnitude of the projections of each vector upon each other is plotted

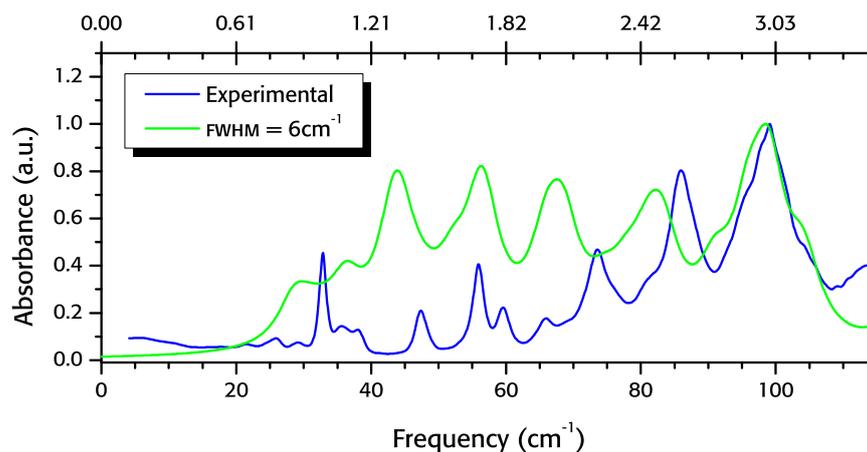


Figure 10.22. Experimental (10 K) and DMACRYS, w99 forcefield spectrum of carbamazepine form I, Lorentzian width =  $6 \text{ cm}^{-1}$

translates into an increase in the magnitude of the forces. The magnitude of the frequency shift is not constant among all the structures, and it does not appear to be solely related to the variation of the molecular dipole magnitude: for example the maximum frequency shift in benzoic acid is  $40 \text{ cm}^{-1}$  (where the molecular dipole increases by 8.64 %), but it is only  $9 \text{ cm}^{-1}$  in nitrofurantoin form  $\alpha$  (where the molecular dipole increases by 36 %).

By far the greatest shifts in lattice mode frequencies happen at higher frequency, in particular to the water-dominated modes outside the experimental terahertz range in the hydrate structures (not shown here, frequency higher than  $150 \text{ cm}^{-1}$  for the regions 2 and 3, as defined in chapter 9.2, page 178), where the highest frequency is enhanced from  $100 \text{ cm}^{-1}$  to  $250 \text{ cm}^{-1}$  depending on the structure. This is probably due to the water molecule being much more polarisable than typical organic systems, and more sensitive even to small changes of the forces in the system, since it is lighter.

The vibrational eigenvectors are much more resilient to change with respect to the polarisation effect in the anhydrate systems as the frequencies, and the majority of the modes are almost unvaried. In the hydrate systems the intermediate frequency modes and some of the highest frequency modes display more mixing: again, this might be due to the greater effect polarisation of water has on the forces acting on light molecules.

The use of the polarisation seems to have a “levelling” effect on the

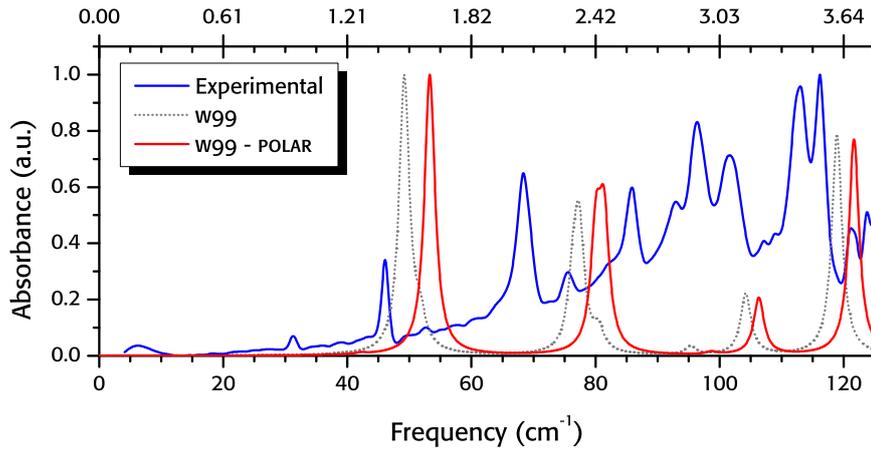


Figure 10.23. Experimental (10 K) and DMACRYS, w99 forcefield spectrum of carbamazepine form III

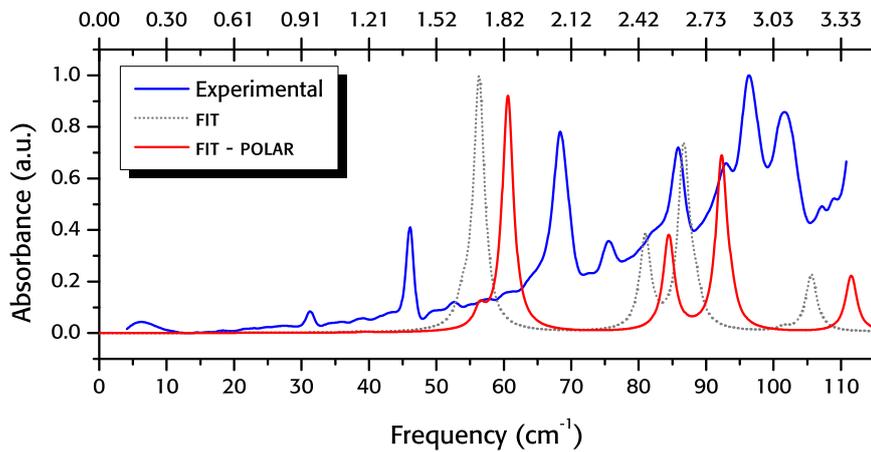


Figure 10.24. Experimental (10 K) and DMACRYS, FIT forcefield spectrum of carbamazepine form III

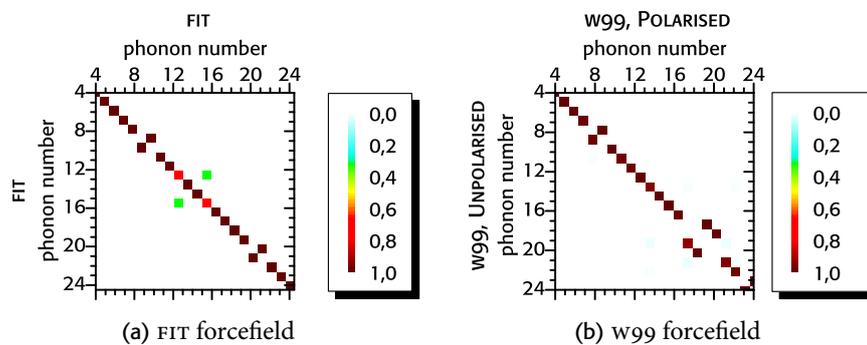


Figure 10.25. Comparison of the DMACRYS eigenvectors of carbamazepine form III, polarised and unpolarised methods. The square of the magnitude of the projections of each vector upon each other is plotted

calculations: the comparison between FIT and w99 spectra shows that they are much more similar than the corresponding unpolarised calculations.

We can notice that the frequency shift sometimes improve the agreement of the higher frequency lattice modes (paracetamol form I, nitrofurantoin hydrate); however the overall effect of the polarisation is to worsen the agreement with experimental spectra, particularly for the low frequency modes. A notable exception is the w99 simulation of carbamazepine form I, where the polarisation seems to improve the positions and intensities of almost all of the modes. This is probably due to the big differences in the polarisation of the four API molecules in the crystal, which were not taken into account in any way in the unpolarised calculation.

### 10.3 INTENSITIES RECALCULATION

In this section we take into account the effect of the polarisation on the intensities of the modes calculated from a DMACRYS calculation without polarisation using the SANTHA script. We briefly compare the calculated with the experimental spectrum for each of our systems: we only analyse the w99 calculation, since we are not recalculating the frequencies. However, we analyse the FIT calculation for benzoic acid, since this is the forcefield that reproduces the spectrum best.

#### 10.3.1 Benzoic acid

The calculated spectrum for benzoic acid is reported in Figure 10.26. We notice that the strongest experimental absorptions (at  $69\text{ cm}^{-1}$  and  $91\text{ cm}^{-1}$ ) are better reproduced in the SANTHA calculation; also, the very weak unpolarised absorption at  $112\text{ cm}^{-1}$  is now stronger, and can be more easily assigned to one of the peaks in the range  $92\text{ cm}^{-1}$  to  $104\text{ cm}^{-1}$ .

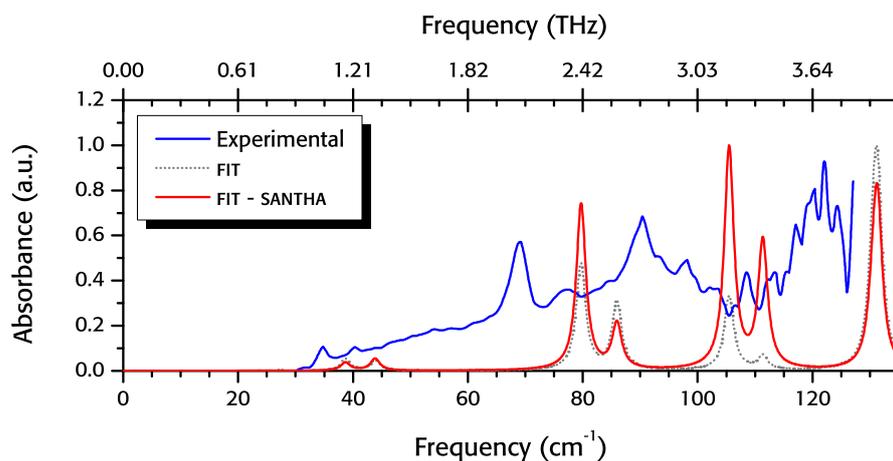


Figure 10.26. Experimental and DMACRYS, FIT forcefield spectrum of benzoic acid calculated with SANTHA

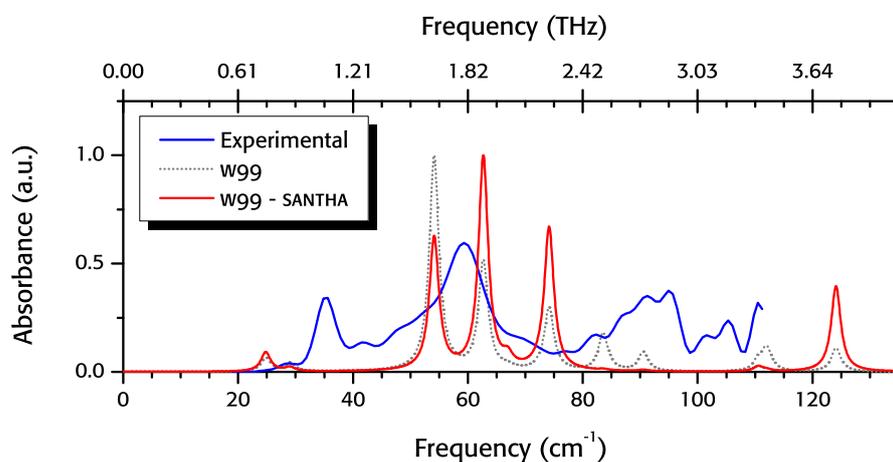


Figure 10.27. Experimental and DMACRYS, w99 forcefield spectrum of anhydrous theophylline calculated with SANTHA

### 10.3.2 Anhydrous theophylline

The calculated spectrum for anhydrous theophylline is reported in Figure 10.27. The new intensities of the peaks in the range from  $50\text{ cm}^{-1}$  to  $80\text{ cm}^{-1}$  agree better with the broad feature centred at  $60\text{ cm}^{-1}$ . The peaks from  $80\text{ cm}^{-1}$  to  $110\text{ cm}^{-1}$  have a much reduced intensity, this time in disagreement with the experimental determination.

### 10.3.3 Theophylline monohydrate

We report only one of the four SANTHA calculation from the symmetry-corrected structures of theophylline monohydrate. For this calculation (Figure 10.28) the SANTHA intensities are very similar to the unpolarised calculation. The intensity attenuation for the peak at  $106\text{ cm}^{-1}$  agrees with the experimental observation, but worsens the agreement for the peak at  $62\text{ cm}^{-1}$ .

#### 10.3.4 Paracetamol form I

The spectrum for paracetamol form I is reported in Figure 10.29. The polarisation has a small effect on the absorption intensities: there is an enhancement of the low frequency absorption, and a minor reshaping of the group in the range  $85\text{ cm}^{-1}$  to  $105\text{ cm}^{-1}$ , whose convolution agrees better with the feature around  $105\text{ cm}^{-1}$ .

#### 10.3.5 Nitrofurantoin $\beta$

The polarised spectrum of nitrofurantoin  $\beta$  is almost indistinguishable from the unpolarised spectrum (Figure 7.17, page 111) and it is therefore not reported.

#### 10.3.6 Nitrofurantoin hydrate form II

The polarised spectrum of nitrofurantoin hydrate form II is reported in Figure 10.30. The peak reshaping highlights a better description of the intensities of most of the modes, improving the agreement in the region from  $70\text{ cm}^{-1}$  to  $110\text{ cm}^{-1}$ .

#### 10.3.7 Carbamazepine form I

The polarised spectrum of carbamazepine form I is reported in Figure 10.31. In the SANTHA calculation the most intense absorption is for the peak at  $96\text{ cm}^{-1}$ , agreeing with the experimental determination. The new peaks are more easily recognisable as convolution of the bigger peaks: for this reason we can more confidently assign the vibrational modes as suggested by the arrow in the graph. As in the recalculation case,

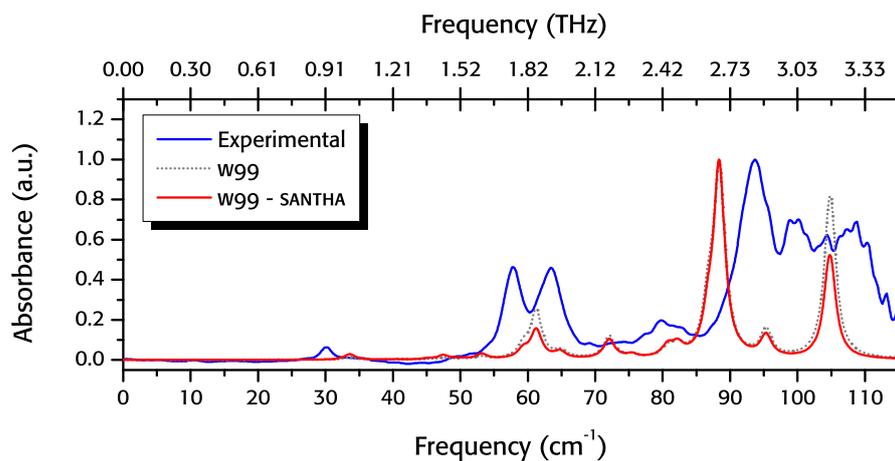


Figure 10.28. Experimental and DMACRYS, w99 forcefield spectrum of theophylline monohydrate calculated with SANTHA

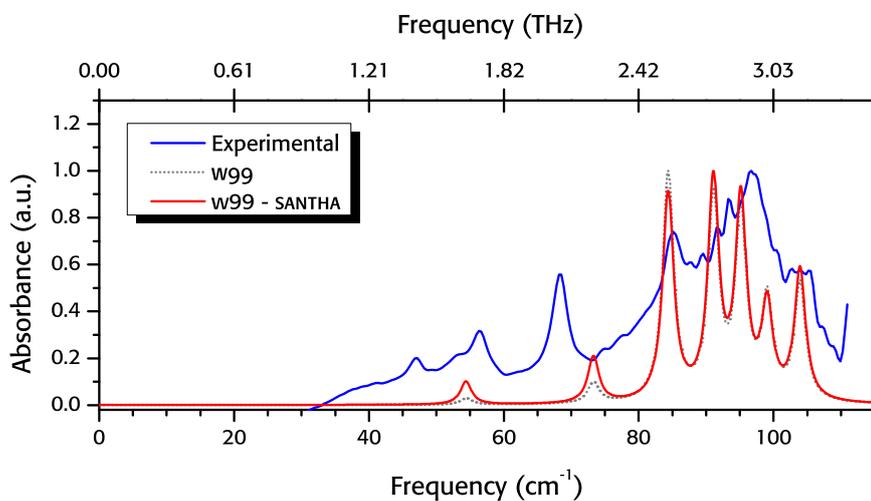


Figure 10.29. Experimental and DMACRYS, w99 forcefield spectrum of paracetamol form I calculated with SANTHA

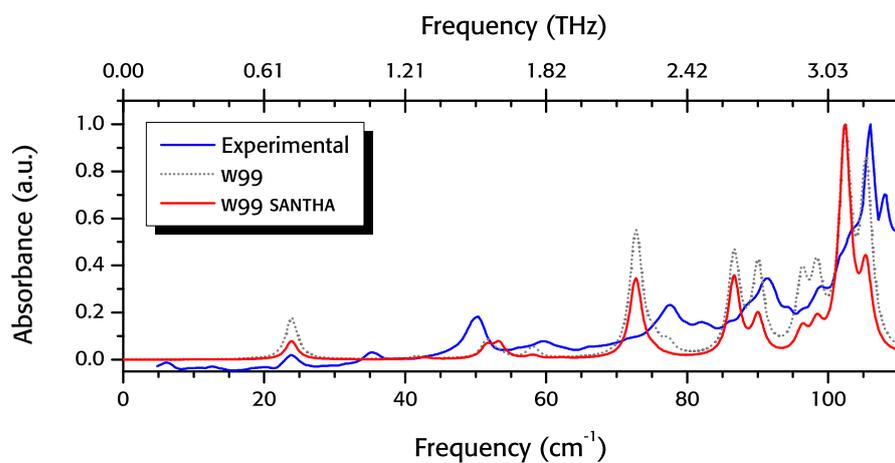


Figure 10.30. Experimental and DMACRYS, w99 forcefield spectrum of nitrofurantoin hydrate form II calculated with SANTHA

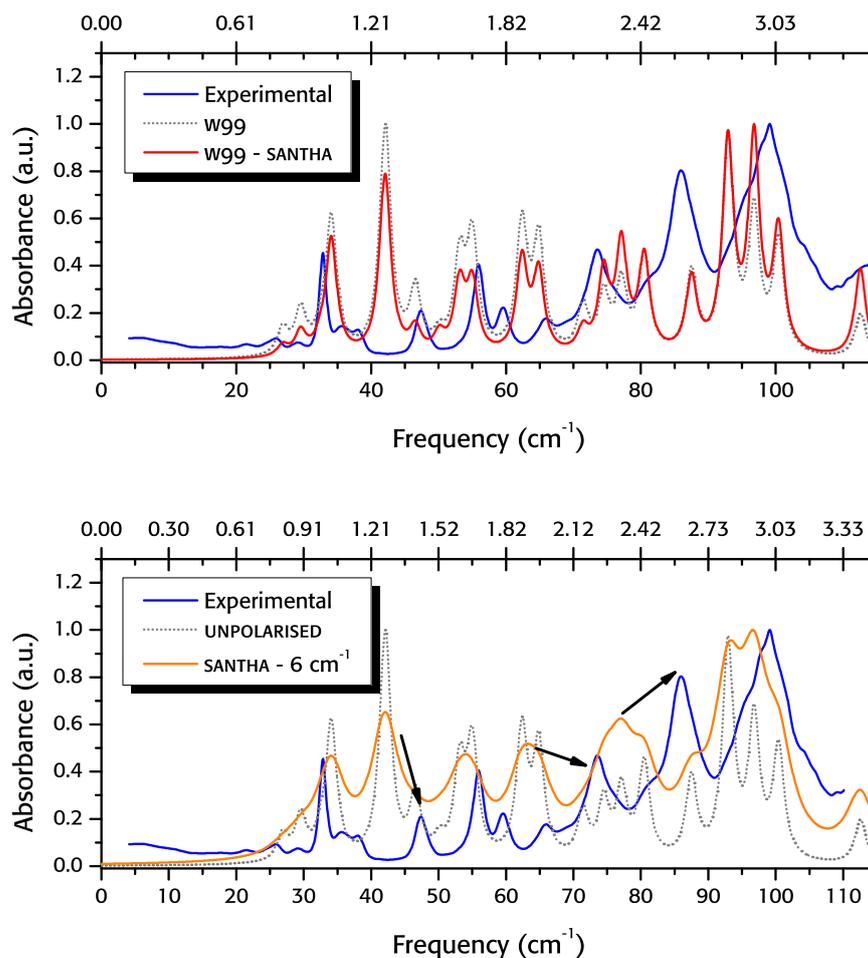


Figure 10.31. Experimental and DMACRYS, w99 forcefield spectrum of carbamazepine form I calculated with SANTHA. Possible peak assignments are suggested by the arrows. In the bottom Figure, the Lorentzian FWHM =  $6 \text{ cm}^{-1}$

the agreement is visually much better than it was, another sign that the polarisation and the differences between the molecules play a big role.

### 10.3.8 Carbamazepine form III

As in the case of the unpolarised calculation, we see a gross overestimation of the absorption of the  $49 \text{ cm}^{-1}$  peak (Figure 10.32). The intensities of the peaks in the region  $70 \text{ cm}^{-1}$  to  $110 \text{ cm}^{-1}$  are reduced by the polarisation effect.

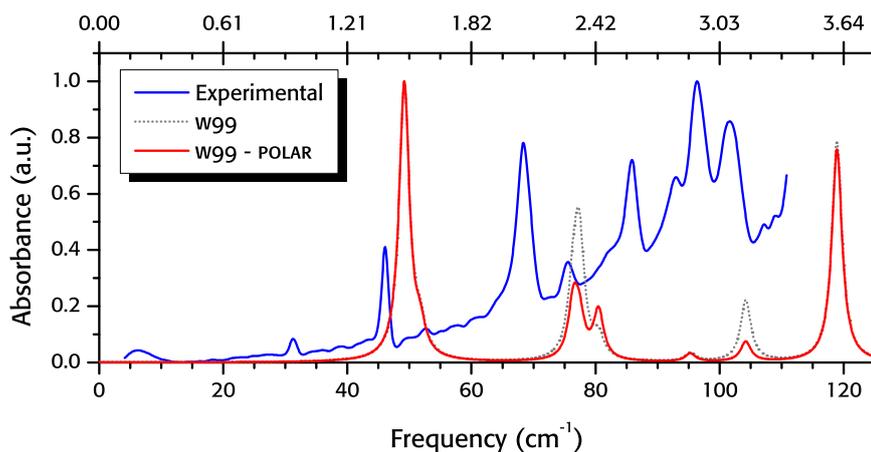


Figure 10.32. Experimental and DMACRYS, w99 forcefield spectrum of carbamazepine form III calculated with SANTHA

### 10.3.9 Conclusions

The recalculation of the absorption intensities by taking into account the local effects of the crystal on the electronic distribution in the molecules improves the agreement with the experimental spectra in most of the cases analysed. This confirms that the local field is an important factor for the determination of the dipole moment variation in the crystal under the vibrations, which determines the absorption intensity. Therefore, despite the theoretical issues (use of an electronic distribution different than the one used for the calculation of the eigenvectors) the use of a self-consistent method would be a more accurate method for the comparison of the calculation with experimental data. Unfortunately, the SANTHA calculation is several orders of magnitude slower than the RUDOLPH methods (hours instead of seconds), but still much cheaper than a periodic DFT calculation.

## 10.4 CONCLUSIONS

In this chapter we analysed the effect of the use of polarisation in the analysis of the absorption spectra. With the exception of carbamazepine form I, the use of a self-consistent electronic distribution seem to worsen the agreement with experimental determinations.

The use of eigenvectors calculated with the unpolarised method and

eigenvalues obtained considering a more accurate intensity determination seems to improve the results, despite the use of an electrostatic distribution not consistent with the parameterisation of the forcefield. The success of this analysis is probably helped by the resilience of the eigenvector towards the change in the electrostatic distribution: therefore we are using what would probably be a good approximation of the “real” eigenvectors, obtainable by a self-consistent combination of a forcefield with polarised electrostatics, and a correct description of the intensities. An interesting test of this theory would be the use of a new forcefield optimised for the use with polarised electrostatics, and a comparison of both the eigenvectors and the absorption intensities with this method.

## ISOSTRUCTURAL CRYSTALS

---

**I**N THIS CHAPTER we study the terahertz spectra of a set of N-phenyl-acetamide derivatives in the temperature range from 80 K to 240 K. The aim of this study was to consider the effect of small molecular changes in isostructural or structurally similar systems on the terahertz spectra. The experimental samples, which are a part of a larger subset, were provided from Suzanne Huth and Michael Hursthouse (University of Southampton); the crystal structures of the fluoro- [229] and methyl- [230] derivatives are available from the eCrystal repository of the University of Southampton [231]. All of the other crystal structures, except for the iodo-derivative (privately provided by Suzanne Huth) are also available from the CSD.

All of the experimental samples have been validated by Powder X-Ray Diffraction (PXRD) with the help of Bhavnita Patel. Powder patterns of the bulk materials were compared to simulated patterns from the known crystal structures to confirm the sample did not undergo any structural modification.

The samples of all of the halogen derivatives (fluoro-, chloro-, bromo- and iodo-derivatives) were confirmed to be the isostructural forms; the methyl-derivative was found to be the non-isostructural polymorph (CSD code ACTOLD01, see Table 6.3).

### 11.1 EXPERIMENTAL AND CALCULATED SPECTRA

In this section we present the experimental and calculated spectra, as we did in chapter 7.

For each of the samples we took five measurements of the terahertz spectrum (see appendix A), at temperatures ranging from 80 K (nitrogen cooled) to 295 K, with the clearest determinations of the absorption peaks for all the samples determined in the lowest temperature measurement.

We used the dispersion corrected PBE functional for all of the reported CASTEP calculations, and we allowed for the relaxation of the unit cell parameters in all calculations.

### 11.1.1 The halogens forcefield

For DMACRYS calculations, the repulsion-dispersion parameters for the halogen atoms are not available in the FIT forcefield: all the DMACRYS phonon calculations of the halogen-derivatives (all of the samples except for the methyl-phenyl acetamide) were consequently performed with the w99 potential only, for which parameters have been fitted.

The parameterisation of the exchange-repulsion interaction for a halogen atom  $X$  includes an anisotropic term, depending on the relative orientation of the atoms (i.e, dependence on the angle  $\widehat{YXZ}$  in a system  $Y-X \cdots Z$ ): for halogen atoms (as demonstrated by Day [232] for the family of chloro-benzenes) the anisotropy of the interaction is not a negligible effect.

With this approach there is a change in the usual Buckingham potential between atoms  $i$  and  $k$  of type  $\iota$  and  $\kappa$  (see Equation 3.18), with the introduction of additional parameters:

$$U_{ik}^{\text{Anis}} = A^{\iota\kappa} \exp(-B^{\iota\kappa}(R_{ik} - \rho^{\iota\kappa}(\Omega_{ik}))) - C^{\iota\kappa} R_{ik}^{-6} \quad (11.1)$$

The function  $\rho$  depends on the relative orientation  $\Omega_{ik}$  of the two atoms with respect to the intramolecular bond to the halogen.  $\rho$  is defined as

$$\rho^{\iota\kappa}(\Omega_{\iota\kappa}) = +\rho_2^{\iota} \frac{3(\widehat{\mathbf{z}}_i \cdot \widehat{\mathbf{R}}_{ik})^2 - 1}{2} + \rho_2^{\kappa} \frac{3(\widehat{\mathbf{z}}_k \cdot \widehat{\mathbf{R}}_{ik})^2 - 1}{2} \quad (11.2)$$

where  $\widehat{\mathbf{R}}_{ik}$  is the interatomic unitary distance vector and  $\widehat{\mathbf{z}}_k$  is the direction vector of the intramolecular bond with the halogen (drawn in blue in Figure 11.1). The values of the parameterisation of halogens are reported in Table 11.1

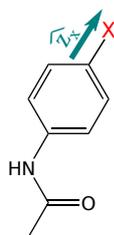


Figure 11.1.

Atom pair	$A''$ kJ mol <sup>-1</sup>	$B''$ Å <sup>-1</sup>	$C''$ kJÅ <sup>6</sup> /mol	$\rho_2$ Å
F . . . F	362 881	3.34	688.91	-0.0352
Cl . . . Cl	569 746	3.34	8366.24	-0.0939
Br . . . Br	1 184 070	3.30	16 254.88	-0.1064
I . . . I	1 261 260	3.14	16 632.14	-0.1215

Table 11.1. Parameters  $A''$ ,  $B''$ ,  $C''$  and anisotropic parameter  $\rho_2$  of the exp-6 w99 potentials. Only the parameters for homonuclear pairs are reported

### 11.1.2 The experimental spectrum of the methyl-derivative

The measured spectrum of the methyl derivative at 80 K is reported in Figure 11.2.

There are four main recognisable peaks: two minor signals at 53 cm<sup>-1</sup> and 79 cm<sup>-1</sup>, and two main signals at 99 cm<sup>-1</sup> and 105 cm<sup>-1</sup>. No feature at frequency higher than 110 cm<sup>-1</sup> can be unambiguously identified due to the measurement uncertainties.

### 11.1.3 DMACRYS calculations of the methyl-derivative

The variation of the unit cell using both the potentials is very small, especially for the w99 calculation. The most visible change between the two structures (see Table 11.2) is the  $c$  axis dimension and (consequently) the unit cell volume, which is almost unvaried with FIT, and expands with w99.

	Exp. (293 K)	Crystal parameters variation			
		w99	Change/%	FIT	Change/%
$a / \text{Å}$	11.74	11.65	-0.08	11.68	-0.45
$b / \text{Å}$	9.59	9.57	-0.20	9.45	-1.41
$c / \text{Å}$	7.56	7.54	-0.23	7.73	2.36
$\beta / ^\circ$	106.0	103.51	-2.34	103.62	-2.24
Vol. / Å <sup>3</sup>	818.18	817.72	-0.05	831.01	1.56
$F$ value		12.13		28.46	

Table 11.2. Variation of the lattice parameters for the DMACRYS geometry optimisation of methyl derivative, with FIT and w99 forcefield. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$ . The initial structure parameters are taken from ACTOLD01

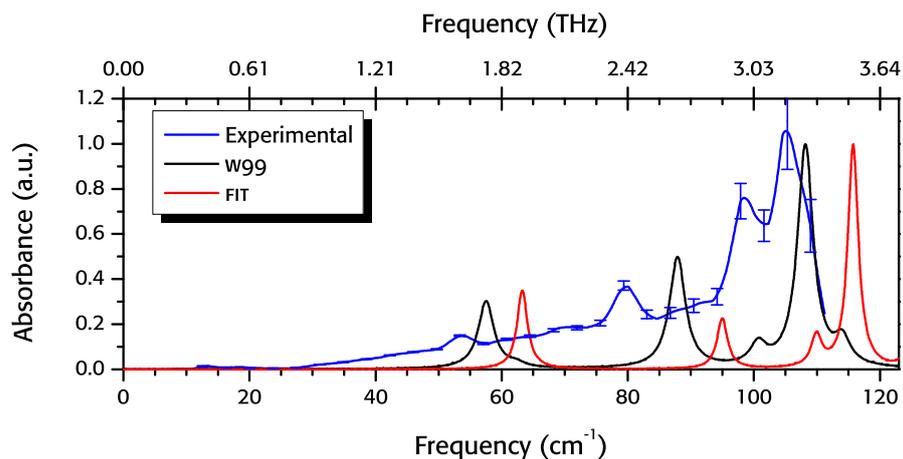


Figure 11.2. Experimental spectrum of the methyl-derivative collected at 80 K, compared with DMACRYS calculations with the w99 and FIT force-fields

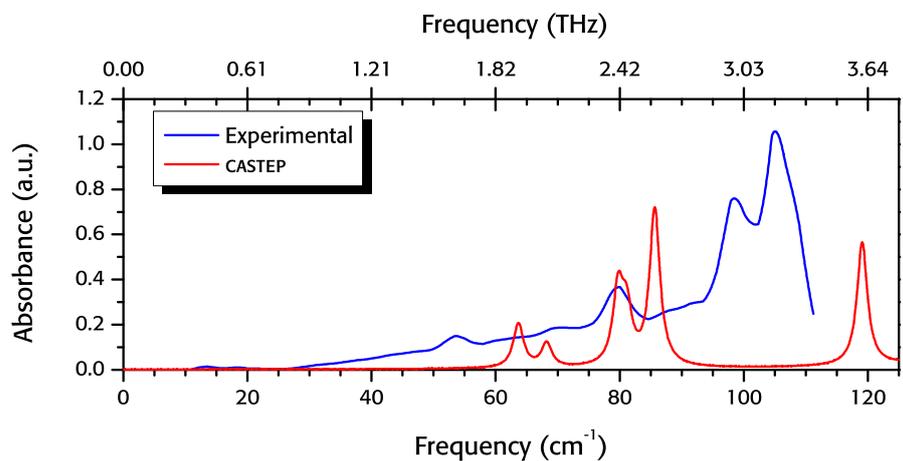


Figure 11.3. Experimental spectrum of the methyl-derivative collected at 80 K, compared with CASTEP calculation, dispersion corrected PBE

	Experimental (293 K)	Crystal parameters variation	
		PBE/DFT-D	Change /%
$a / \text{\AA}$	11.74	11.42	-2.72
$b / \text{\AA}$	9.59	9.15	-4.58
$c / \text{\AA}$	7.56	7.08	-6.34
$\beta / ^\circ$	106.0	106.85	0.80
Volume / $\text{\AA}^3$	754.51	709.35	-5.98

Table 11.3. Change in the lattice parameters for the CASTEP geometry optimisation of the methyl-derivative, dispersion corrected PBE. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$

The volume difference (which is itself a reflection of the different evaluation of the forces) is the reason behind the absorption frequencies shift to higher frequencies that we observe in the FIT calculation.

The agreement in the number and position of the calculated peaks with the experimental features is particularly good using the w99 potential; we can see in Figure 11.2 that the agreement is excellent for all the peaks except for the absorption at  $87\text{ cm}^{-1}$ , which is calculated to be  $7\text{ cm}^{-1}$  higher in frequency than the observed absorption peaks. Also, the intensity of the absorption at  $\approx 100\text{ cm}^{-1}$  is predicted to be too weak. The agreement is slightly less satisfying for the FIT calculation: if we look at the spectrum shape, all of the absorption features are present and easily assignable to their experimental counterparts, but their positions are shifted from  $10\text{ cm}^{-1}$  to  $20\text{ cm}^{-1}$  higher in frequency compared to the experimental data.

#### 11.1.4 CASTEP calculation of the methyl-derivative

All of the cell axes are decreased in the CASTEP minimisation, with a volume contraction of almost 6%. The agreement with the experimental spectrum is poor: if we assign the calculated doublet centred around  $85\text{ cm}^{-1}$  to the most prominent experimental peaks at  $98\text{ cm}^{-1}$  and  $106\text{ cm}^{-1}$  based on their shape, (Figure 11.3) the frequencies do not match, as they are shifted by almost  $20\text{ cm}^{-1}$  to lower frequencies; the lowest experimental peak has a very good frequency match with a very weak calculated feature at  $55\text{ cm}^{-1}$ .

#### 11.1.5 The experimental spectrum of the fluoro-derivative

The experimental spectrum of the fluoro-derivative at 80 K is reported in Figure 11.4. The spectrum is essentially flat up to  $55\text{ cm}^{-1}$  (with the possible exception of a very weak feature at  $26\text{ cm}^{-1}$ , which might be due an artifact of the baseline removal). The strongest absorptions are observed at  $81\text{ cm}^{-1}$  and  $103\text{ cm}^{-1}$ ; the shoulder at  $74\text{ cm}^{-1}$  is probably a hidden peak, and some of the other strong background noise features between  $85\text{ cm}^{-1}$  to  $110\text{ cm}^{-1}$  could possibly be absorption peaks as well.

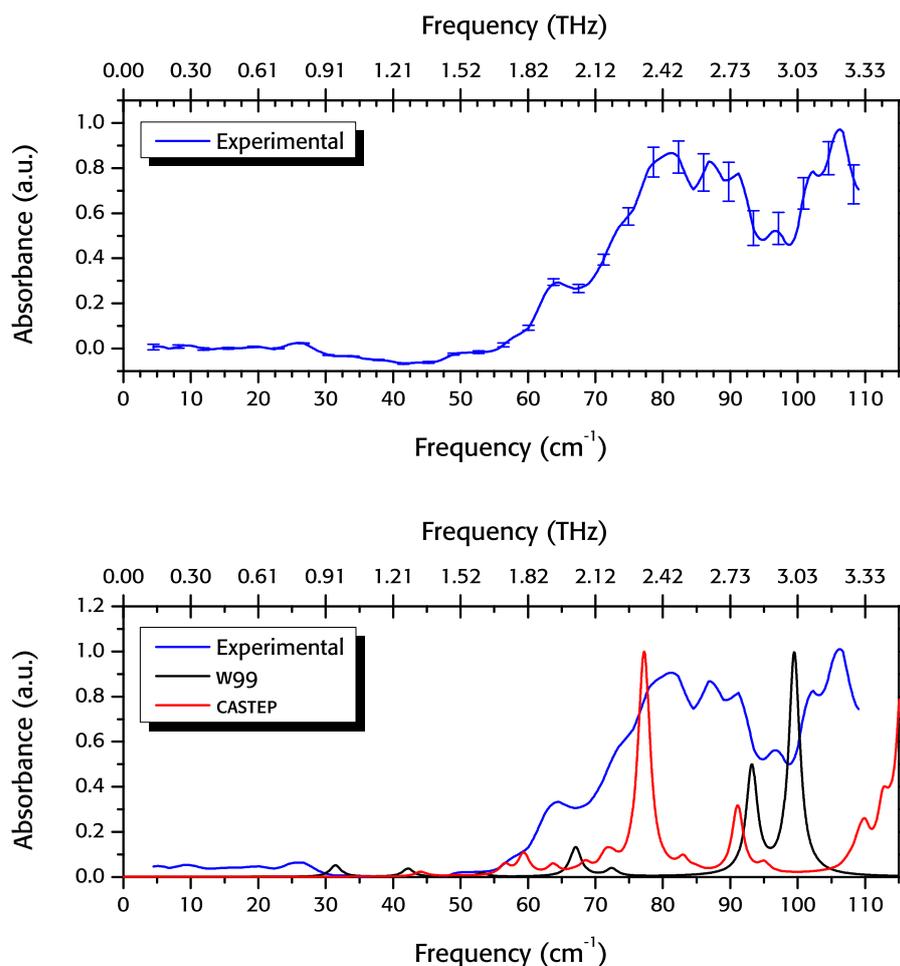


Figure 11.4. Experimental spectrum of the fluoro-derivative collected at 80 K (top) and its comparison with calculations using DMACRYS with the w99 forcefield and CASTEP, PBE/DFT-D (bottom)

	Exp. (293 K)	Crystal parameters variation			
		w99	Change /%	DFT-D	Change /%
$a / \text{\AA}$	4.73	4.48	-5.29	4.36	-7.82
$b / \text{\AA}$	17.07	17.81	4.38	17.16	0.52
$c / \text{\AA}$	9.63	9.68	0.52	9.33	-3.11
$\beta / ^\circ$	92.87	96.78	4.21	92.44	-0.46
Vol. / $\text{\AA}^3$	776.86	767.53	-1.20	699.30	-9.98
$F$ value		103.91		-	

Table 11.4. Change in the lattice parameters in the geometry optimisation of the fluoro-derivative with DMACRYS, w99 force field, and CASTEP, PBE/DFT-D. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$ . The initial structure parameters are taken from SAZLEL

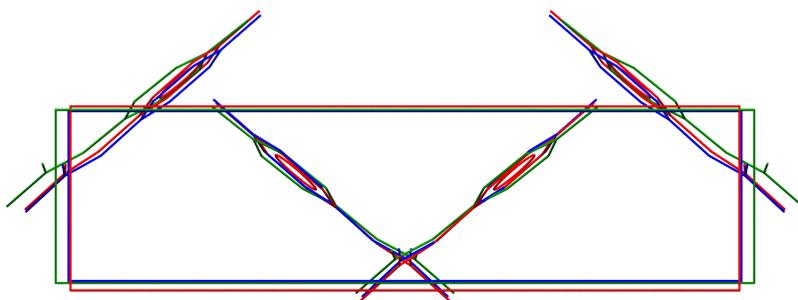


Figure 11.5. Superimposition of the experimental unit cell of the fluoro-derivative (in red) with the CASTEP (in blue) and the DMACRYS (in green) optimised geometry, as seen down the  $c$  axis. The unit cell axes are in the same colour of the structure. Hydrogen atoms are not shown

#### 11.1.6 DMACRYS calculation of the fluoro-derivative

Despite large changes in the magnitude of the  $a$  and  $b$  axes during lattice minimisation, the volume of the structure optimised with DMACRYS is almost equivalent to the experimental value, because the positive and negative changes cancel out (see Table 11.4). As a result, the  $F$  value is quite big, a sign that either the parameters of the atom-atom potential for fluorine are not good enough, or that the structure is not well described by the existing force field.

The calculated spectrum (Figure 11.4, in red) has the same flatness as the experimental one, but the first strong experimental feature is found shifted of more than  $20\text{ cm}^{-1}$  to higher frequencies in the simulation; additionally, there is no calculated higher energy feature, except for a weak one at  $130\text{ cm}^{-1}$  (not shown).

#### 11.1.7 CASTEP calculation of the fluoro-derivative

The biggest variation of the final CASTEP geometry is the  $a$  axis, with almost an 8% contraction. There is an increase of the  $b$  axis, as it was observed in the DMACRYS calculation.

Figure 11.5 shows the superimposition of the unit cell of the calculated structures (blue for CASTEP, green from DMACRYS) with the experimental determination at 295 K (in red) as seen through the  $c$  axis. The most noticeable difference between the structures is in the unit cell parameters, with the large increase of the  $b$  axis in DMACRYS which, despite not being the biggest in percentage, affects the longest axis: the CASTEP

structure has therefore a better visual agreement. Furthermore, we can see that the CASTEP molecules overlay better with the experimental determination and the V shape formed by the central molecules is more similar: the angle is almost unvaried for the CASTEP calculation ( $99.36^\circ$  vs  $99.22^\circ$  in the experimentally determined structure) but it widens to  $103.88^\circ$  in DMACRYS.

The similarities of the CASTEP geometry with the experimentally determined structure suggests that we should see a good agreement for the CASTEP phonon spectrum, as we observe in Figure 11.3. The calculation predicts a flat spectrum up to  $77.4\text{ cm}^{-1}$ , in agreement with the strongest measured absorption; another strong absorption is centred at  $115\text{ cm}^{-1}$ .

#### 11.1.8 The experimental spectrum of the chloro-derivative

The experimental spectrum of the chloro-derivative at 80 K is reported in Figure 11.6, top graph. There are several clear absorption peaks: the strongest recorded features are found at  $59\text{ cm}^{-1}$ ,  $76\text{ cm}^{-1}$  and  $101\text{ cm}^{-1}$ . There are other weaker absorptions at  $40\text{ cm}^{-1}$ , a very weak feature at  $48\text{ cm}^{-1}$  and a shoulder at  $91\text{ cm}^{-1}$ .

#### 11.1.9 DMACRYS calculations of the chloro-derivative

The optimisation of the structure with DMACRYS displays a very good agreement with the experimental structure, obtained from the most recent of the two room temperature crystal structure determinations in the CSD (CLACTNO2 [233]). The minimised structure has minimal variation in both the unit cell parameters (see Table 11.5) and in the molecular positions ( $F < 6$ ).

The agreement with the experimental spectrum is excellent for the position of the peaks up to  $90\text{ cm}^{-1}$ , with the strong calculated absorption at  $79.6\text{ cm}^{-1}$  shifted by less than  $5\text{ cm}^{-1}$  to higher frequency. The experimental features at  $101\text{ cm}^{-1}$  and  $121\text{ cm}^{-1}$  are not clearly seen in the calculated spectrum, with the highest active vibrational phonons are found at  $110\text{ cm}^{-1}$  and  $120\text{ cm}^{-1}$  (very weak, not shown).

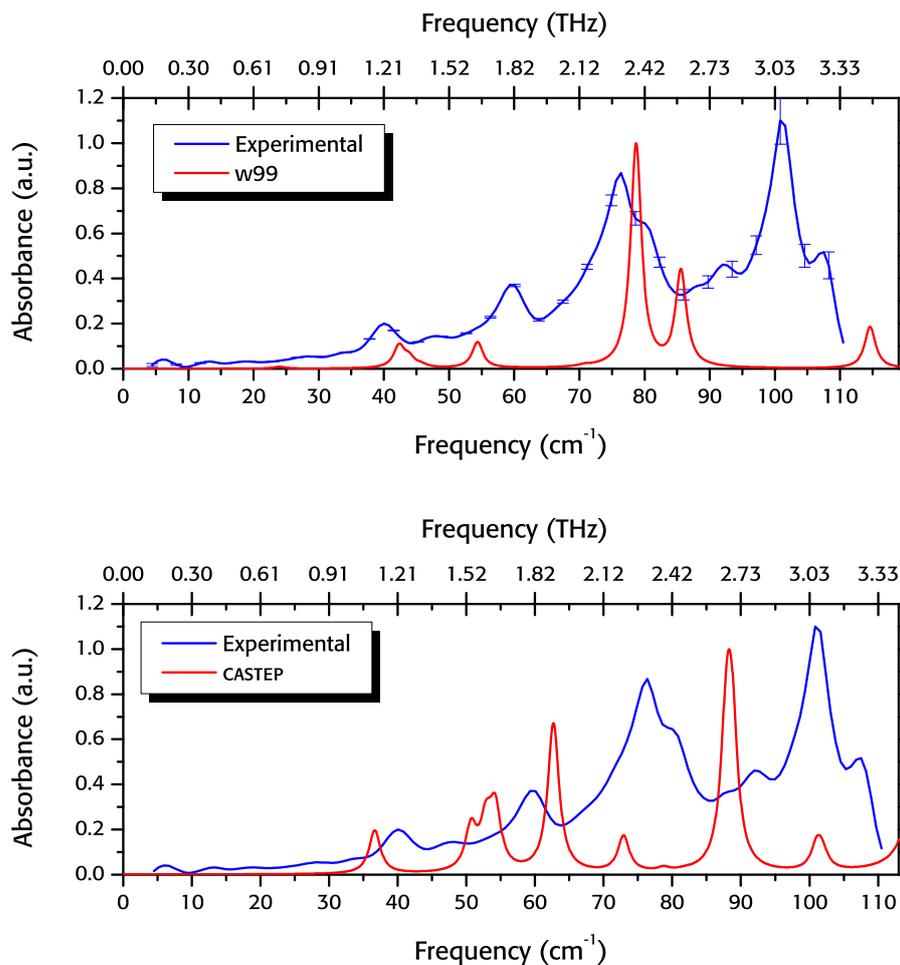


Figure 11.6. Experimental spectrum of the chloro-derivative collected at 80 K compared with the spectrum calculated with DMACRYS, w99 force-field (top) and with CASTEP, PBE/DFT-D

	Crystal parameters variation				
	Exp. (293 K)	w99	Change/%	DFT-D	Change/%
$a / \text{\AA}$	9.71	9.84	1.34	9.54	-1.75
$b / \text{\AA}$	12.75	12.88	1.08	11.72	-8.07
$c / \text{\AA}$	6.52	6.57	0.72	6.60	1.22
Vol. / $\text{\AA}^3$	807.47	833.21	3.18	738.63	-8.55
<i>F</i> value		5.50		-	

Table 11.5. Change in the lattice parameters in the geometry optimisation of the chloro derivative with DMACRYS, w99 force field, and CASTEP, PBE/DFT-D. Symmetry constraints: unit cell angles  $\alpha = \beta = \gamma = 90^\circ$ . The initial structure parameters are taken from CLACTN02

#### 11.1.10 CASTEP calculation of the chloro-derivative

The comparison of the optimised CASTEP unit cell with the experimental values shows an anisotropic change in the unit cell dimensions, as in the case of the fluoro-derivative; the  $a$  axis is much more affected by the minimisation ( $-8\%$ ) than the others, and the  $c$  axis is slightly increased. The overall volume compression ( $8.5\%$ ) is comparable to what we observed in the other systems.

There is a good qualitative agreement between the experimental and calculated spectra, even if the frequencies of individual absorptions do not agree perfectly. All of the calculated absorption features are shifted to lower frequencies, but not uniformly: the frequencies of the four lower experimental peaks cover the range from  $40\text{ cm}^{-1}$  to  $78\text{ cm}^{-1}$ , while their calculated counterparts are compressed in the region from  $36\text{ cm}^{-1}$  to  $64\text{ cm}^{-1}$ . The stronger experimental absorption at  $101\text{ cm}^{-1}$  has a calculated homologue at  $88\text{ cm}^{-1}$ .

#### 11.1.11 The experimental spectrum of the bromo-derivative

The experimental spectrum of the bromo-derivative at  $80\text{ K}$  is reported in Figure 11.7, top graph. There are many similarities with the spectrum of the chloro-derivative for the lowest frequency absorptions, at  $38\text{ cm}^{-1}$ ,  $60\text{ cm}^{-1}$  and  $72\text{ cm}^{-1}$ . The strongest absorption feature is at  $91\text{ cm}^{-1}$ , with two shoulders on its side. At higher frequencies, the measurement uncertainties allow only the clear identification of one peak ( $101\text{ cm}^{-1}$ ).

#### 11.1.12 DMACRYS calculations of the bromo-derivative

The  $b$  axis is the only unit cell parameter to show any significant variation during the geometry optimisation of the bromo-derivative. The  $F$  value is not high ( $38.11$ ), and the biggest deviations from the experimental structure (apart from the aforementioned  $b$  axis) is the rotational reorientation of the molecules (see Figure 11.7b).

The calculated spectrum has two strong peaks at  $81\text{ cm}^{-1}$  and  $102\text{ cm}^{-1}$ , with other weak absorptions in the region from  $20\text{ cm}^{-1}$  to  $65\text{ cm}^{-1}$ . As in the case of the chloro derivative, the experimental peaks at higher frequency are not clearly seen in the DMACRYS calculation. The

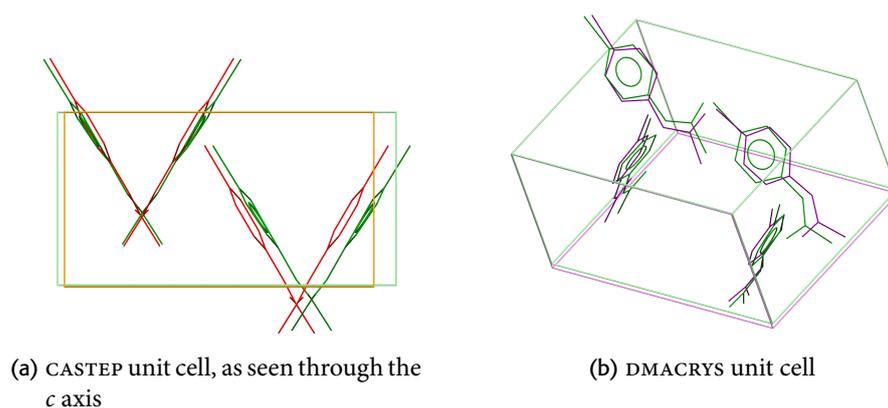


Figure 11.7. Superimposition of the experimental unit cell of the bromo-derivative (in green) with the CASTEP (in red, left picture) and the DMACRYS (in violet, right picture) optimised geometry. The unit cell axes are in the same colour as the structure. Hydrogen atoms are not shown for clarity

two main experimental peaks at  $91\text{ cm}^{-1}$  and  $72\text{ cm}^{-1}$  can be assigned to calculated features, that are shifted  $10\text{ cm}^{-1}$  to higher frequency in the calculated spectrum; the experimental signal at  $60\text{ cm}^{-1}$  is not accounted for in the calculated spectrum.

### 11.1.13 CASTEP calculation of the bromo-derivative

As in the case of the other halogen-derivatives, the structural changes after the energy minimisation with CASTEP are not isotropic, with the *c* axis showing the biggest variation (an 8 % contraction); this is very similar to the change we observed in the value of the *c* axis in the chloro-derivative.

The effect is clearly visible in Figure 11.7a: the V shape of the molecule is not affected, but the separation of the molecules along the *c* axis is clearly reduced.

The calculated absorption at  $40\text{ cm}^{-1}$  is stronger than any of the experimental peaks in that region from  $55\text{ cm}^{-1}$  to  $85\text{ cm}^{-1}$ , while the peaks in the region  $90\text{ cm}^{-1}$  to  $105\text{ cm}^{-1}$  are weaker than their experimental counterparts. The peaks in the region from  $55\text{ cm}^{-1}$  to  $85\text{ cm}^{-1}$ , instead, agree very well with experimental features if we shift them by  $10\text{ cm}^{-1}$  to higher frequency.

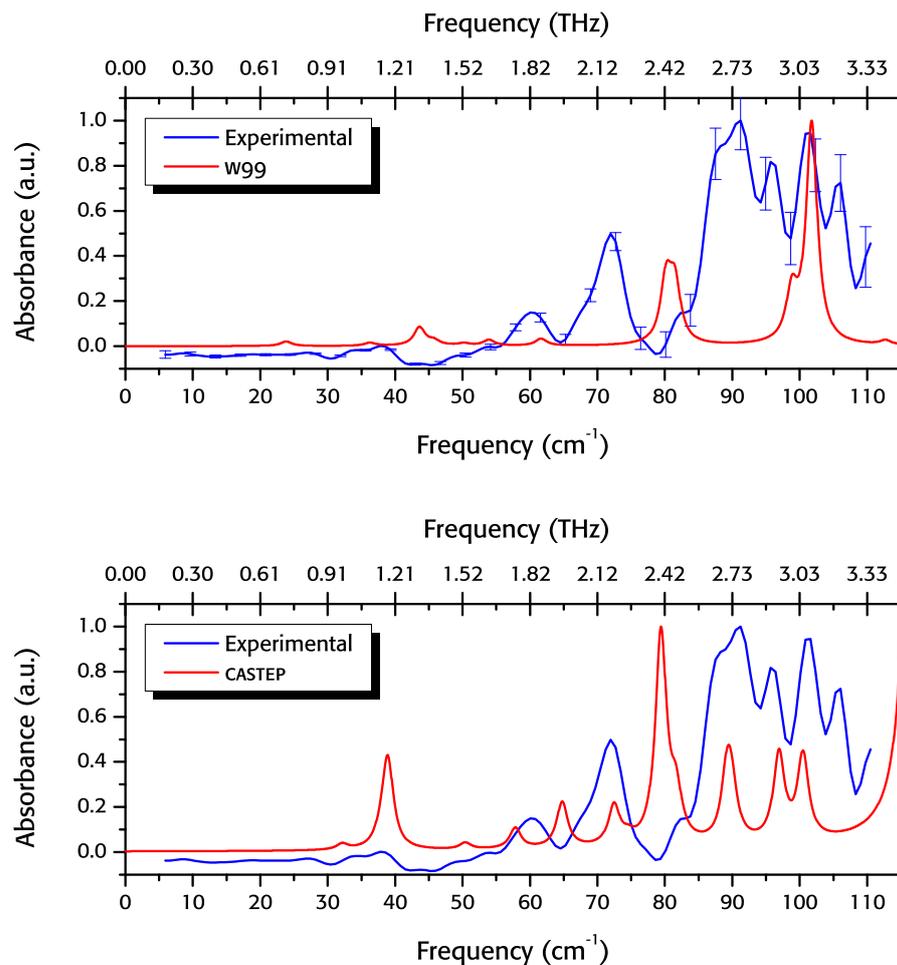


Figure 11.8. Experimental spectrum of the bromo-derivative collected at 80 K compared with the spectrum calculated with DMACRYS, w99 force-field (top) and with CASTEP, PBE/DFT-D (bottom figure)

	Crystal parameters variation				
	Exp. (293 K)	w99	Change/%	DFT-D	Change/%
$a / \text{\AA}$	9.70	9.74	0.50	9.58	-1.23
$b / \text{\AA}$	13.00	13.01	0.09	11.87	-8.69
$c / \text{\AA}$	6.67	6.89	3.34	6.73	0.89
Vol. / $\text{\AA}^3$	841.08	874.40	3.96	766.31	-8.91
$F$ value		38.11		-	

Table 11.6. Change in the lattice parameters in the geometry optimisation of the bromo-derivative with DMACRYS, w99 force field, and CASTEP, PBE/DFT-D. Symmetry constraints: unit cell angles  $\alpha = \beta = \gamma = 90^\circ$ . The initial structure parameters are taken from BRAC TN

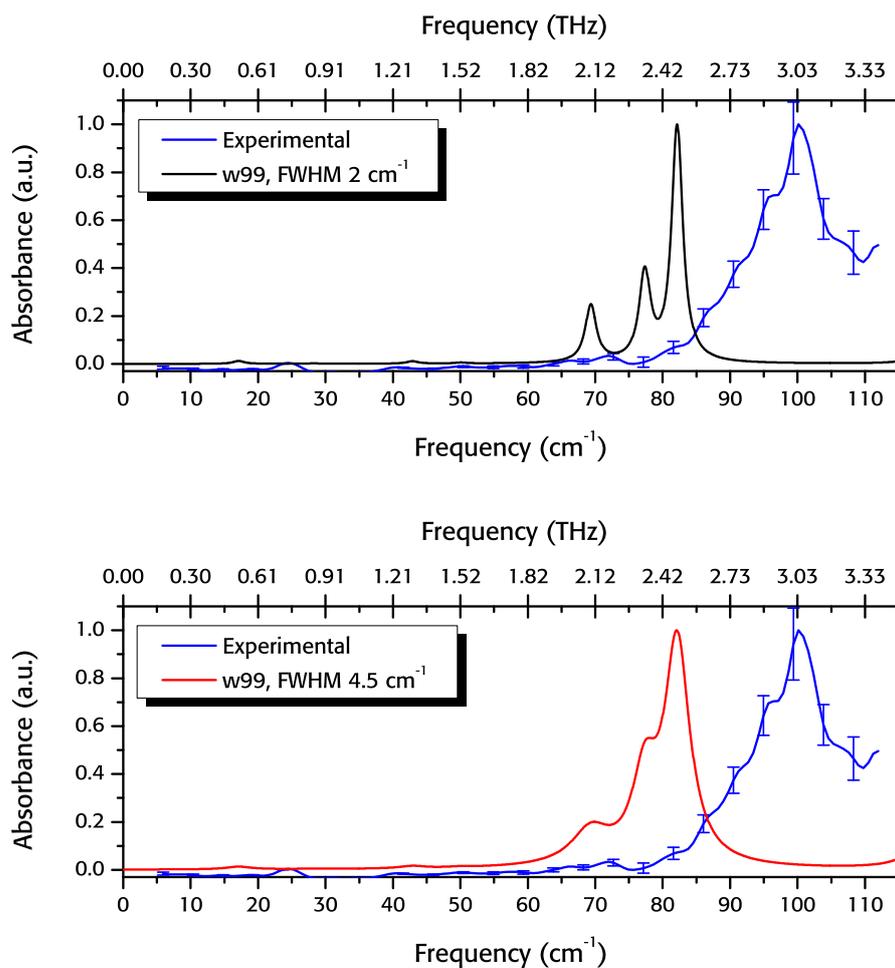


Figure 11.9. Experimental spectrum of the iodo-derivative collected at 80 K, compared with the spectra calculated with DMACRYS, w99. In the top graph, the Lorentzian line width is set to  $2\text{ cm}^{-1}$ , in the bottom graph is set to  $4.5\text{ cm}^{-1}$

	Experimental (293 K)	Crystal parameters variation	
		w99	Change/%
$a / \text{\AA}$	9.57	9.41	-1.75
$b / \text{\AA}$	10.30	10.60	2.85
$c / \text{\AA}$	9.42	9.66	2.46
$\beta / ^\circ$	109.28	6.89	3.34
Volume / $\text{\AA}^3$	841.08	906.78	3.26
$F$ value		25.79	

Table 11.7. Change in the lattice parameters in the geometry optimisation of the iodo-derivative with DMACRYS, w99 force field. Symmetry constraints: unit cell angles  $\alpha = \gamma = 90^\circ$

#### 11.1.14 The experimental spectrum of the iodo-derivative

For this molecular system we have only the DMACRYS calculation, because the CASTEP geometry optimisation failed to converge: this is the only calculation where this happened, and it is probably due to deficiencies of the iodine pseudopotential.

The spectrum of the iodo-derivative (see Figure 11.9) has a lot of similarities with the spectrum of the fluoro-derivative: it is flat up to  $80\text{ cm}^{-1}$  and has a wide peak centred at  $100\text{ cm}^{-1}$ . Since the experimental uncertainties are comparable to the peaks heights we cannot say anything about the other peaks at higher frequency. There are two very weak absorptions at  $23\text{ cm}^{-1}$  and  $73\text{ cm}^{-1}$ , although the lower frequency one might be an artifact of the baseline removal.

#### 11.1.15 DMACRYS calculations of the iodo-derivative

The unit cell axes, which display a change of less than 3 % during lattice minimisation, are not significantly affected by the minimisation process (Table 11.7). As in the case of the bromo-derivative, there is almost no translation of the molecules during lattice minimisation, but there is some rotational reorientation, although the contribution to the total  $F$  value is anyway small (7.2 out of 25.8).

The calculated phonon spectrum shows a very good agreement in shape with the experimental determination, with the three peaks from  $65\text{ cm}^{-1}$  to  $85\text{ cm}^{-1}$  assignable to the experimental feature centred at  $100\text{ cm}^{-1}$ . The very wide experimental peak is probably a sum of overlapping peaks: to highlight this we show in Figure 11.9 (bottom) the w99 calculated spectrum with a wider Lorentzian peak shape ( $\text{FWHM} = 4.5\text{ cm}^{-1}$ ), which better represents the experimental features.

There are other weak peaks calculated at  $17\text{ cm}^{-1}$ ,  $42\text{ cm}^{-1}$  and  $49\text{ cm}^{-1}$ , confirming that the experimental low frequency peaks might be real absorptions.

## 11.2 SPECTRA SIMILARITIES BETWEEN THE SYSTEMS

In Figure 11.10 we report the measured low temperature terahertz spectra of the four halogen derivatives. The absorption features are quite distinctive for each of the different crystals. The most similar spectra belong to the chloro- and bromo derivative, systems that share an almost superimposable crystal structure (note that the measured methyl-derivative is not the isostructural form).

The numbered peaks in Figure 11.10 identify corresponding absorptions in the spectrum of the chloro- and bromo-derivatives. While peaks 1 and 2 are almost in the same position, peak 3 to 6 are at higher frequency for the bromo-derivative. The shift is expected, as bromine has a greater mass than chlorine, and this difference directly affects the frequencies (see Equation 3.5).

According to DMACRYS, the most relevant absorption modes of the chloro-derivative are a doublet around  $80\text{ cm}^{-1}$ , and an overlapping doublet at  $114.63\text{ cm}^{-1} - 114.67\text{ cm}^{-1}$ . Similarly, the bromo-derivative has almost superimposing features at  $80.2\text{ cm}^{-1}$  and  $81.5\text{ cm}^{-1}$  and a slightly separated doublet at  $98.8\text{ cm}^{-1}$  and  $101.8\text{ cm}^{-1}$ .

From the analysis of the eigenvectors of the two calculations, the vibrational modes in the high energy part of the spectra (frequency higher than  $70\text{ cm}^{-1}$ ) are rotational in nature, with the molecular motion around different axis: the types of motion are shown in Figure 11.11.

The strongest absorption in both systems is of the type in Figure 11.11a, which has the largest dipole change (it reorients the dipole of the N–C–O group) is at  $78.7\text{ cm}^{-1}$  for the chloro derivative, but at  $101.8\text{ cm}^{-1}$  for the bromo derivative. Modes at  $114\text{ cm}^{-1}$  in the chloro-derivative (Figure 11.11b) corresponds to the peak at  $98\text{ cm}^{-1}$  in the bromo derivative. Although the result based on the calculation seems to suggest that the assignment based on frequency alone is not entirely correct, it must be stressed that the method does not have the accuracy needed to be conclusive.

The CASTEP spectra are quite different from each other, and we did not attempt a 1:1 assignment based on frequency alone. The eigenvector decomposition is reported and discussed later, in section 11.4.1.

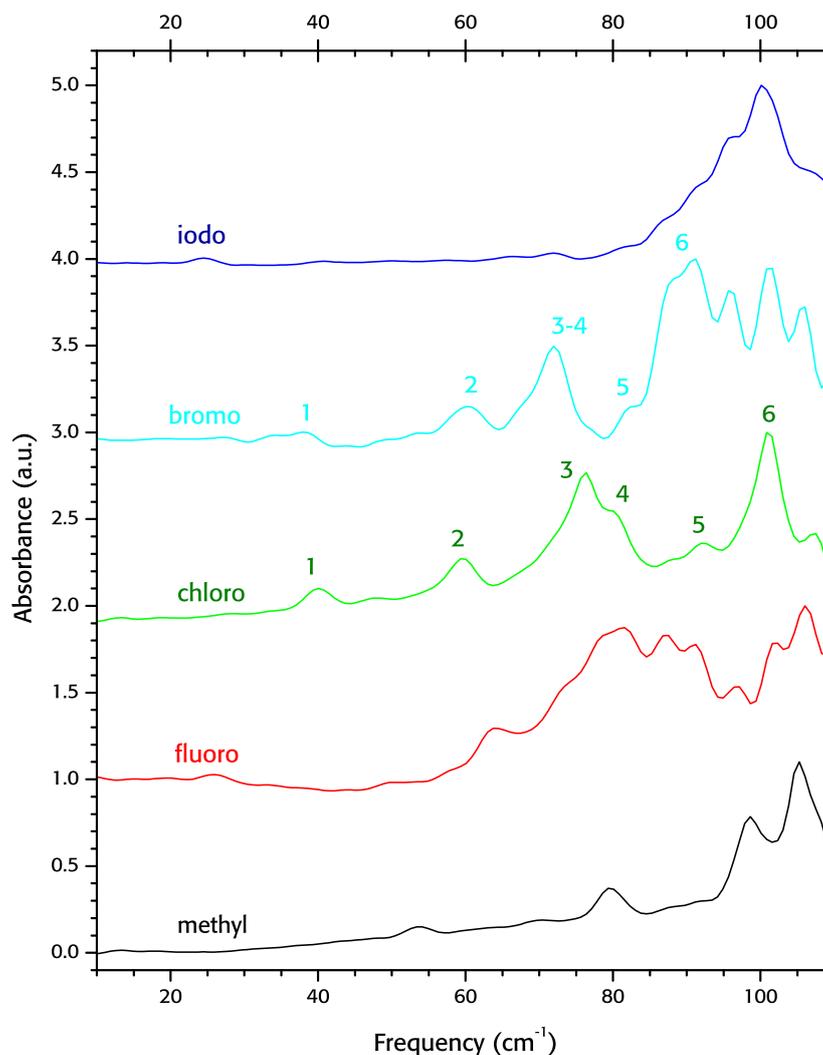


Figure 11.10. Low temperature determinations of the measured phenylacetamide-derivatives. Numbers on the chloro- and bromo-derivatives identify corresponding features in the spectra

### 11.3 POLARISATION EFFECT

We already discussed the effect of polarisation on the spectra with DMACRYS in a dedicated chapter for most of the systems, both with complete recalculation of the eigenvectors and the SANTHA method for the intensities only (chapter 10).

The results for the substituted phenyl acetamides with the complete recalculation of the eigenvalues confirm our previous findings: the *F*-value upon lattice minimisation are of the same magnitude as the unpolarised calculations, with the exception of the iodo-derivative (Table 11.8), highlighting small changes in the structures despite the different elec-

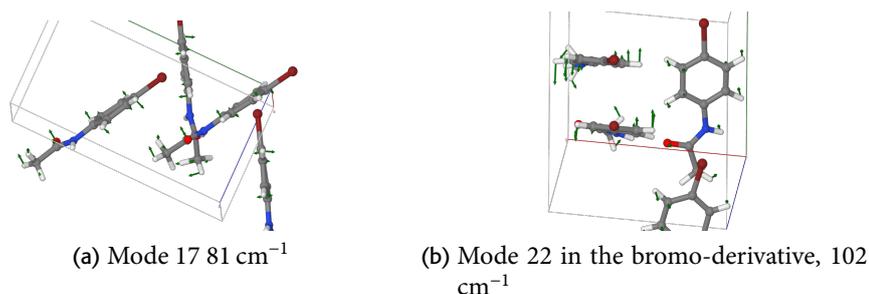


Figure 11.11. Vibrational eigenvectors of the bromo-derivative calculated with DMACRYS, w99 forcefield

System	FIT forcefield		w99 forcefield	
	Unpol.	Pol.	Unpol.	Pol.
Methyl Phenyl-acetamide I	15.09	4.31	23.77	7.55
Methyl Phenyl-acetamide II	9.67	9.90	18.29	14.36
Fluoro Phenyl-acetamide	–	–	103.91	61.39
Chloro Phenyl-acetamide	–	–	5.50	4.42
Bromo Phenyl-acetamide	–	–	38.11	24.71
Iodo Phenyl-acetamide	–	–	25.80	55.72

Table 11.8. Variation of the structure upon minimisation with DMACRYS, FIT and w99 forcefields of the acetamide-derivatives, as described by the  $F$  value with and without the effect of polarisation. The FIT forcefield does not include parameters for the halogen atoms, and the calculations were not performed

trostatics. We do not report the spectra obtained with the full minimisation with polarisation for the acetamide-derivatives, since for all the systems we found a reduced agreement with the experimentally determined spectra (as an example, see the spectrum of N-(4-methyl-phenyl) acetamide in Figure 11.12), similarly to what we found in chapter 10.

The changes in intensities using the SANTHA program were only very minor in the case of the halogen derivatives, mainly affecting only the weakest absorptions (see for example Figure 11.13 for the recalculation of the spectrum of the fluoro-derivative).

The spectrum calculated with SANTHA for the methyl derivative is reported in Figure 11.14. There are not big changes, but the attenuation of the absorption intensity at  $60\text{ cm}^{-1}$  and the strengthening of the  $101\text{ cm}^{-1}$  peaks agrees very well with the experimental determinations: the use of the polarisation effect in this system does improve the agreement with the experimentally determined spectrum.

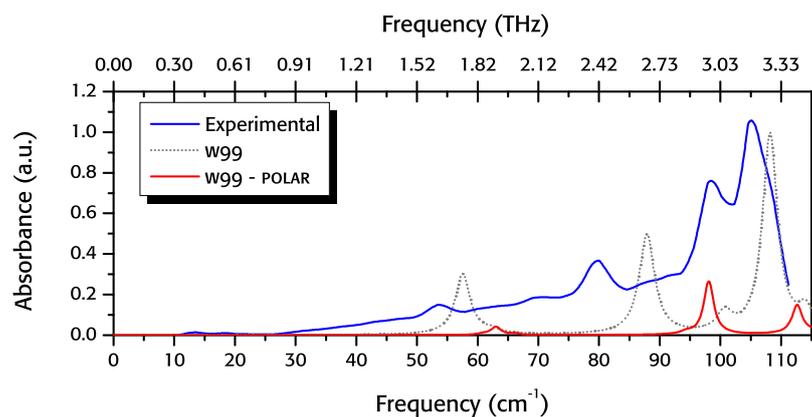


Figure 11.12. Experimental and DMACRYS, w99 forcefield spectrum of N-(4-methyl-phenyl) acetamide with full recalculation using polarisation

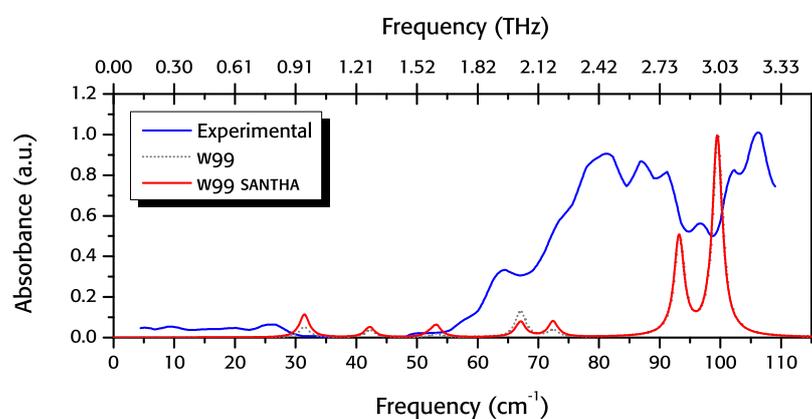


Figure 11.13. Experimental and DMACRYS, w99 forcefield spectrum of N-(4-fluoro-phenyl) acetamide calculated with SANTHA

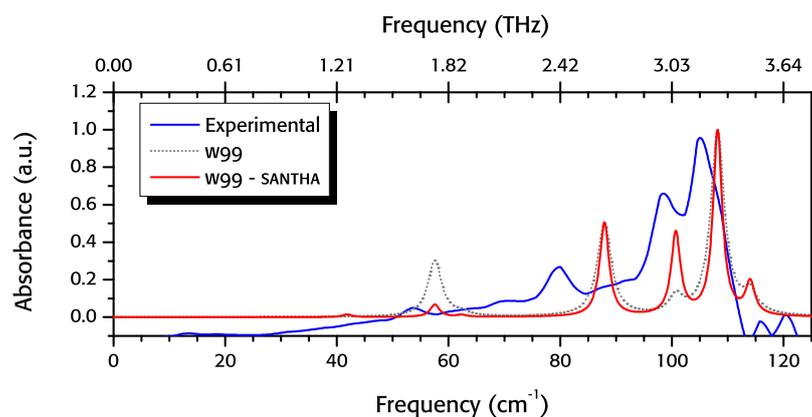


Figure 11.14. Experimental and DMACRYS, w99 forcefield spectrum of N-(4-methyl-phenyl) acetamide calculated with SANTHA

## 11.4 RIGIDITY ANALYSIS

In this section we complement the analysis of the molecular rigidity reported in chapter 8. We consider the calculation relative to the most similar crystals, with the same space group and almost equivalent unit cell parameters: the chloro-, bromo- and methyl derivatives.

We expected these three systems to have similar frequency for the lowest normal mode of vibration of the isolated molecule, due to an almost identical molecular skeleton. The calculated torsions of the acetamide group at the B3LYP/6-31G\*\* level of theory have values that are indeed very close to the paracetamol case (54.11 cm<sup>-1</sup>, 54.4 cm<sup>-1</sup> and 53.7 cm<sup>-1</sup> for the methyl-, fluoro-, chloro-derivatives respectively, compared to the 50.1 cm<sup>-1</sup> of paracetamol). However, for the rotation of the methyl group we calculated a frequency as low as 21.26 cm<sup>-1</sup>, 23.92 cm<sup>-1</sup> and 35.4 cm<sup>-1</sup> in the methyl-, chloro- and bromo-derivative respectively, compared to the higher 42.8 cm<sup>-1</sup> value of paracetamol.

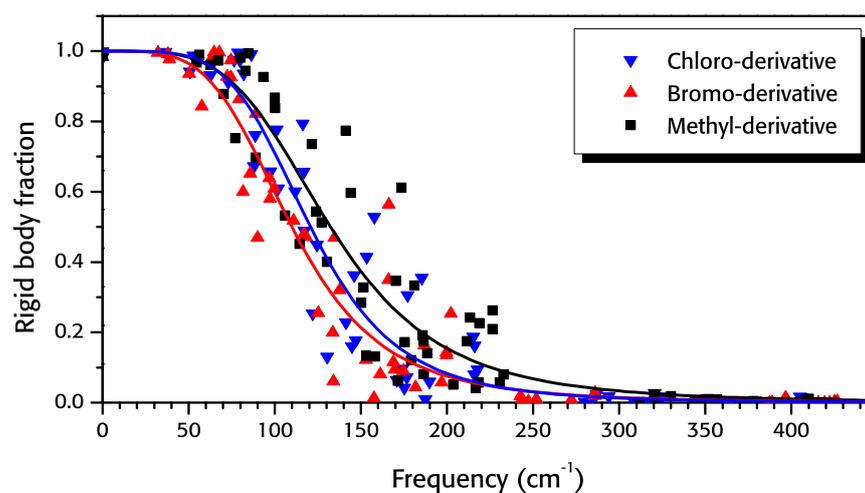
The eigenvectors for the chloro-, bromo- and methyl-derivatives calculated lattice modes with CASTEP were also expected to be very similar, taking into account that the three molecules share almost superimposable crystal structures and due to the molecule similarity.

The rigid body fractions from the calculations of the three systems are reported in Figure 11.15. There is a sharp separation between the different regions of the spectra, as in the case of theophylline (page 146). The rigid-body region is very well defined up to 100 cm<sup>-1</sup>, with the first internal vibration falling between 81 cm<sup>-1</sup> (for the bromo-derivative) and 89 cm<sup>-1</sup> (methyl-derivative).

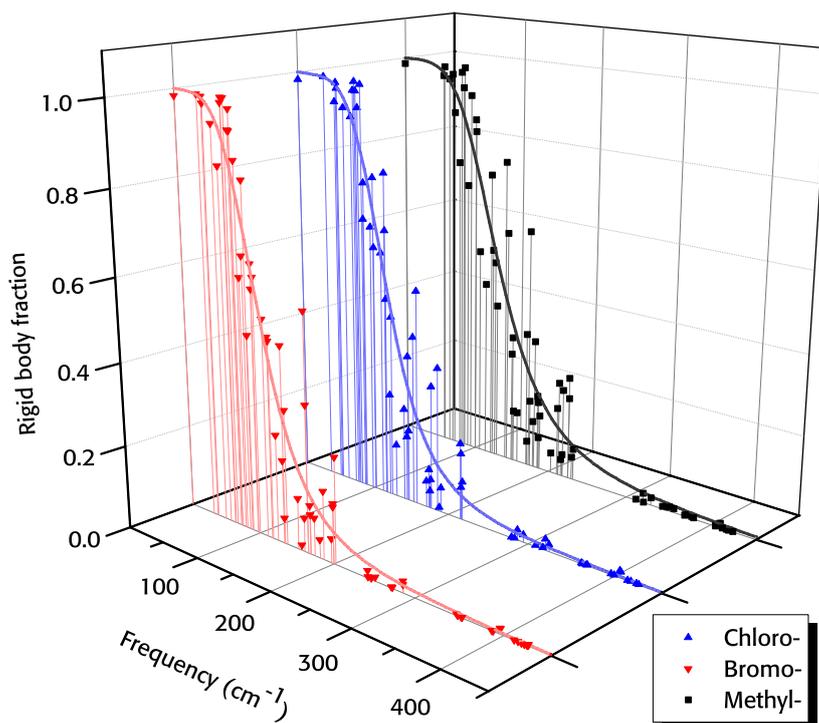
The strongest intermolecular interaction for all of the structures is the O ··· HN hydrogen bond, and in all of the lowest energy modes of vibration it is not stretched. The internal rotation of the substituted phenyl ring relative to the acetamide group is the most important internal contribution in all the partially internal modes in the range 0 cm<sup>-1</sup> to 100 cm<sup>-1</sup> for the three systems.

The results of the logistic fits highlight the high similarity between the three structures. The frequency range over which rigid molecule modes are seen in Figure 11.15a depends on the molecular mass of the group: bromine (79 u), chlorine (35 u) and methyl (15 u).

The energy-minimised structures of the bromo- and chloro- derivat-



(a) Rigid body contribution of the methyl-, chloro- and bromo-derivative with CASTEP calculations, dispersion corrected PBE



(b) Exploded view of the above graph

Figure 11.15. Rigid body contribution to the phonon modes of N-(4-chloro-phenyl) acetamide (blue), N-(4-bromo-phenyl) acetamide (red) and N-(4-methyl-phenyl) acetamide (black) in the region from  $0\text{ cm}^{-1}$  to  $450\text{ cm}^{-1}$ , as calculated with CASTEP, dispersion corrected PBE. The logistic curves are of the same colour as the data

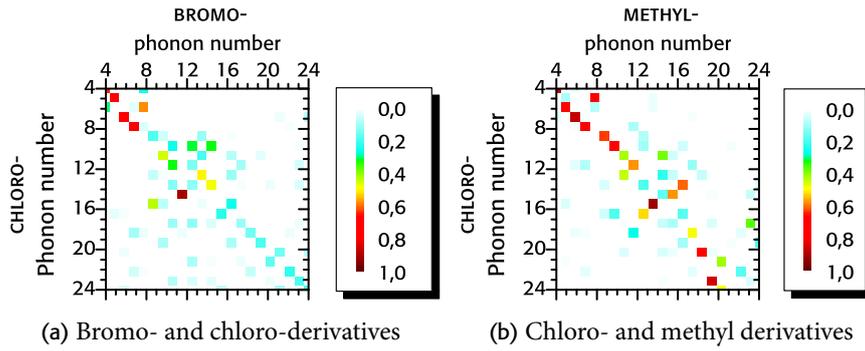


Figure 11.16. Comparison of the CASTEP eigenvectors, PBE/DFT-D functional, chloro-, bromo and methyl-N-phenyl-acetamides. The colour represents the square of the magnitude of the projections of each vector

Type	Form	# mode	Rigid body fraction	Frequency (cm <sup>-1</sup> )
$\omega_{in}$	Chloro-	12 of 225	0.67	87.93
	Bromo-	13 of 225	0.60	81.66
	Methyl-	10 of 261	0.69	88.77
$\omega_{nrig}$	Chloro-	29 of 225	0.53	157.90
	Bromo-	32 of 225	0.56	166.19
	Methyl-	28 of 261	0.61	173.44
$\omega_{high}$	Chloro-	40 of 225	0.16	216.06
	Bromo-	41 of 225	0.25	202.31
	Methyl-	43 of 261	0.26	226.44

Table 11.9. Characterising frequencies of the optical phonons of N-(4-chloro-phenyl) acetamide, N-(4-bromo-phenyl) acetamide and N-(4-methyl-phenyl) acetamide from the CASTEP calculation, dispersion corrected PBE

Form	$x_0$ (cm <sup>-1</sup> )	$p$	$\chi^2$	$\varsigma$
Chloro-	120.71	4.75	0.012	0.640
Bromo-	109.91	4.19	0.012	0.609
Methyl-	132.38	4.12	0.015	0.774

Table 11.10. Characterising parameters of the logistic fit of the rigid body fraction in the phonon eigenvectors of N-(4-chloro-phenyl) acetamide, N-(4-bromo-phenyl) acetamide and N-(4-methyl-phenyl) acetamide from a CASTEP calculation, PBE functional

ive are very similar, and they share the same number and type of atoms (except for the halogen type). In Figure 11.16a we compare the eigenvectors from the two calculations by projecting the eigenvectors upon each other (similarly to the comparison between eigenvectors of PBE vs PBE/DFT-D, page 157), and taking the scalar product between them. We can see a fair diagonal trend in the graph, which confirm the similarity not only in the structures, but in the eigenvectors as well.

The same very good agreement between the different sets of eigenvectors is seen in the case of the chloro- and methyl-derivative (Figure 11.16b). The agreement is more pronounced than it appears from the graph: the scalar product is calculated excluding the hydrogen atoms in the methyl group in order to have the same number of atoms in the two structures. As a consequence, the highest possible superimposition is always less than 1 (depending on the contribution of the hydrogen atoms in the methyl group, which is different for each phonon).

#### 11.4.1 Eigenvector agreement between CASTEP and DMACRYS

In this section we analyse the eigenvector agreement between CASTEP and DMACRYS, as described in section 8.4, page 158.

The bromo-derivative is the system displaying the highest level of agreement, even though the results are quite close for all three systems, around 75 % (see Table 11.11) for the sum of the most important DMACRYS vectors in the deconstruction of each of the CASTEP eigenvectors. The most representative DMACRYS eigenvector contributions for the first 35 optical phonons of the three systems is shown in Figure 11.17.

The CASTEP eigenvectors of the bromo-derivative calculation exhibit a good 1:1 agreement with single DMACRYS eigenvector, except for two pairs of coupled modes (8, 10 and 12, 15).

For the chloro-derivative, the agreement between the DMACRYS and CASTEP eigenvectors is not very good, but is better than what we observe in the various paracetamol calculations. In this case some of the CASTEP modes have a very good agreement with single DMACRYS modes (4, 5, 6, 8 in particular), while the majority of the others are a mixture of two DMACRYS modes (modes 7 and 11, 9 and 13, 10 and 12, 16 and 21 are the lowest frequency ones).

In the methyl-derivative, none of the CASTEP eigenvectors exhibit are

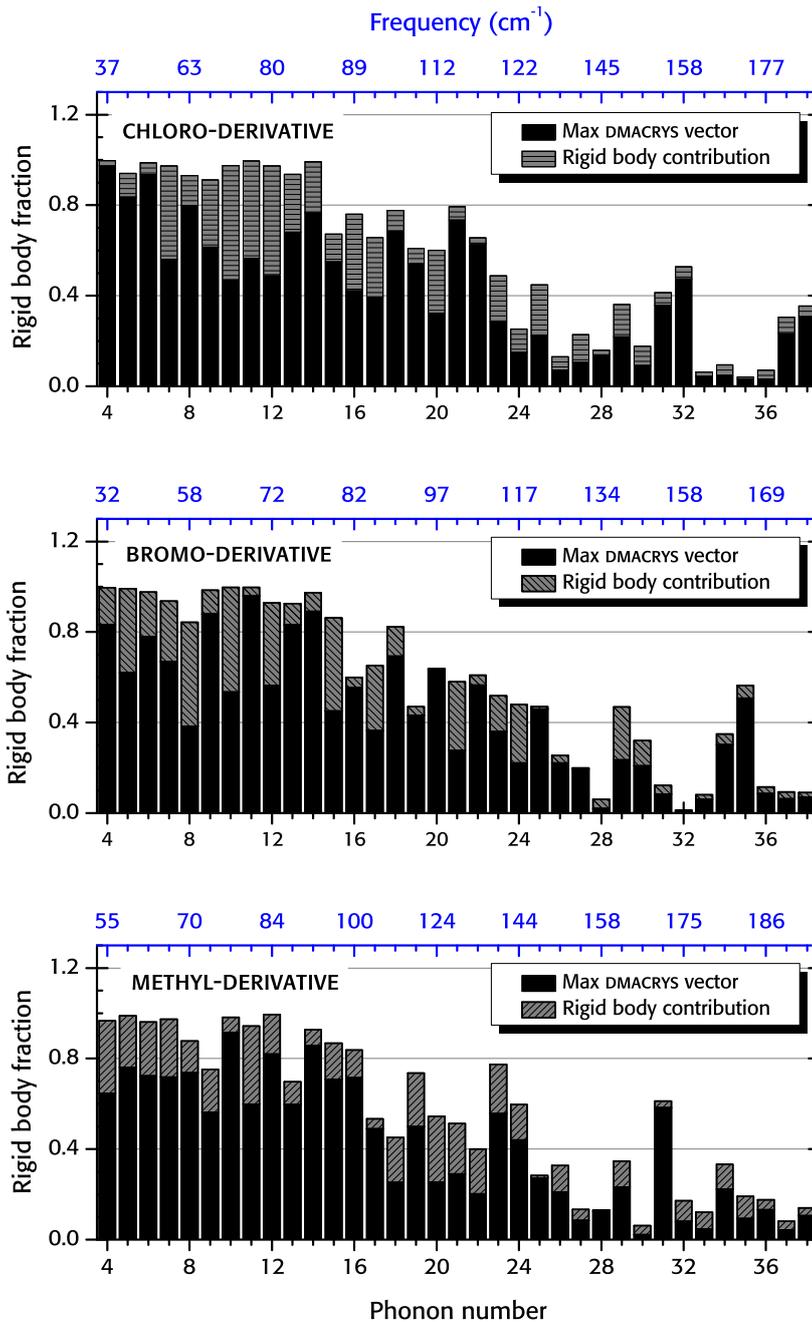


Figure 11.17. Sum of the most representative DMACRYs vector for the measured substituted acetamides

a mixture of two DMACRYs modes, even though the individual contribution of the DMACRYs vectors are not as good as for the other systems.

Crystal system	Overall DMACRYS vectors sum / %
N-(4-methyl-phenyl) –	73.93
N-(4-chloro-phenyl) –	76.57
N-(4-bromo-phenyl) –	77.40

Table 11.11. Sum of the most representative DMACRYS vector in the dispersion corrected CASTEP eigenvectors

## 11.5 CONCLUSIONS

From the analysis of the N-phenyl-acetamide derivatives we have another example of the high effectiveness of terahertz spectroscopy in the discrimination between similar crystals. We found that, despite close similarity in the structure, only two out of the five molecular crystals resulted in very similar terahertz spectra: the bromo- and chloro derivative, the only two sharing an almost superimposable structure and belonging to the same space group. Even these two systems have differentiating features, especially in the higher frequency part of the terahertz spectrum.

The simulation of the terahertz spectra with DMACRYS once again highlighted that the method has to be improved: for most of the features in the simulations it is hard to make precise assignments with the experimental spectrum. The assignments of absorption peaks with CASTEP were easier, but the frequencies do not agree with the experimental values, probably because of the contraction of the unit cell.

The analysis of the eigenvectors shows a rigid molecule behaviour similar to what we observe for paracetamol, with a clear separation between rigid molecule and internal vibrational modes. The eigenvectors of the isostructural crystals (chloro-, bromo- and methyl-) were found to be similar, which is another confirmation of the resilience of the vibrational modes to small changes in molecular structure. The position of the absorption frequencies were found to be the most sensitive parameter to changes, and therefore the most characterising feature of an absorption peak.

## CONCLUSIONS AND OUTLOOK

---

**I**N THIS THESIS we discussed the use and development of methods aimed at the interpretation of terahertz spectroscopy, applied to the study of small organic molecules of pharmaceutical interest. Our approach was based on the calculation of the normal modes of vibration of the molecular crystals in the harmonic approximation. We used two different methods: DMACRYS, a rigid body, atom-atom method and CASTEP, a DFT-based simulation method. We selected a few polymorphic systems to study, and we analysed the similarity of calculated spectra with experimentally determined absorptions.

The experimental acquisition of the spectra was performed with a state-of-the-art spectrometer at low temperature, because the absorption features are enhanced when the temperature is decreased.

As discussed in chapters 7 and 11, the DMACRYS calculations displayed varying agreement with the experimental determinations, depending on the systems studied. For most of the molecular crystals we attempted assignments of the calculated peaks to the experimental features, based on the agreement of the peak position and the calculated intensity of each absorption. While in a few cases the match is good or satisfactory, the agreement between experiment and calculation can be better defined as qualitative in most cases, because the frequencies of the calculated peaks are often far off their experimental counterparts, and sometimes are even missing (as observed for the peak at  $65\text{ cm}^{-1}$  in paracetamol form I in figure 7.14, page 108). Furthermore, different crystal forms of a molecule display a different level of agreement: for example, in the calculated spectra of nitrofurantoin form  $\beta$  several experimental peaks are missing, while for nitrofurantoin hydrate form II there is a nice match between calculated and experimental peaks up to  $100\text{ cm}^{-1}$ .

The CASTEP calculations, although providing in principle a more accurate description of the crystal, failed to deliver a quantitative description of the experimental features as well. The only simulation displaying

a very good agreement with its experimental counterpart for both positions of the peaks and their absorption intensities is the non-dispersion corrected calculation of paracetamol form I. The spectrum of some of the other dispersion-corrected calculations (carbamazepine form III, 4-chloro-phenyl-acetamide) are in good agreement with the shape of the experimental spectra, but the position of the absorption peaks is far off.

The use of the dispersion correction in the DFT calculations failed to improve the match with the experimental determinations. A possible reason for this result is in the inadequate parametrisation of the Grimme  $c_6$  parameters for organic molecular systems, as suggests by a study [212] that we reported at page 133.

In chapter 8 we analysed the eigenvectors of the CASTEP calculations for our molecular systems and we compared them to the rigid-molecule DMACRYS calculations. The comparison aimed to check for which systems and up to which frequency our rigid molecule atom-atom model was justified in the simulations of terahertz spectra. For each of the systems we studied, the rigidity of the eigenvectors was well represented by a logistic curve, highlighting the presence of three separated frequency regions: one region (at low frequencies) characterised by rigid-molecule vibrations; the highest frequency region, where modes are comprised of only intramolecular vibration; and an intermediate region, where there is mixing of intermolecular and intramolecular vibrations. The frequency range of our spectrometer ( $4\text{ cm}^{-1}$  to  $110\text{ cm}^{-1}$ ) allows for the probing of mainly rigid-body motions: according to the results of the eigenvector analysis, the intramolecular contribution to these vibrational modes for all of our systems can be understood in terms of 2–3 of the lowest energy normal modes of vibration of the isolated molecule.

The rigidity of some of the vibrational modes of the hydrate crystal forms we studied (chapter 9) were found to have quite a different behaviour than their neat counterpart. The marked characteristics of the water molecule compared to the organic molecules (its small mass and very small moment of inertia) allowed for an easy identification of the rigid-body dominated eigenvectors where water movement are predominant, which were found at high frequency ( $> 350\text{ cm}^{-1}$ ). Rigid-body vibrational modes where the vibrational modes involve the stretching of the hydrogen bonds between water and the API (of frequencies between  $\approx 150\text{ cm}^{-1}$  to  $250\text{ cm}^{-1}$ ) were also easy to identify in the calculated spectra, although they are found outside the range of our spectrometer. We

hope that future improvements of the experimental apparatus, or the use of a different technique (e.g. FTIR or RAMAN) will allow us to explore this region of the spectrum to verify the presence of these modes in the experimental spectrum as well.

In chapter 10 we explored a modification of the DMACRYS approach, with the use of two PERL scripts (RUDOLPH and SANTHA) that we developed to calculate the effects of the crystal environment on the electrostatics of the molecules. The vibrational eigenvectors and eigenfrequencies we obtained with this method worsened the agreement with the experimental determinations: this is due to the fact that polarisation effects are already averaged into the empirically determined repulsion-dispersion parameters of the w99 forcefield, obtained with the use of electrostatics from unpolarised, isolated molecule calculations. However, the use of SANTHA to evaluate the absorption intensities improved the agreement of the calculations with the experimental determinations; the improved molecular electrostatic gave a better description of the dipole variations with the vibrational movement, taking advantage of the fact that the eigenvectors and eigenvalues are calculated with a good starting model.

In chapter 11 we analysed a series of structurally similar molecular crystals. The experimental spectra were found to be comparable only for crystals with the same space groups, and sharing an almost superimposable crystal structure (the chloro- and bromo-acetanilides): once again this shows the strength of terahertz spectroscopy in differentiating between very similar systems.

## FUTURE WORK

The use of computational methods to identify the molecular vibrations behind the experimental absorption peaks in molecular crystals needs a better modelling of the systems to be effective. For DFT methods, an unavoidable improvement will be in the development and use of new  $c_6$  dispersion parameters.

The analysis of DMACRYS results points towards a few directions for the improvement of our computational method. We believe that the use of a forcefield optimised for the use of polarised charges will improve the

agreement of the calculated vibrational frequencies with their experimental counterparts: therefore, we are planning a reparameterisation of w99, to have a forcefield consistent with the use of polarised electrostatics.

Other improvements will be based on the modification of the DMACRYS code. In chapter 8 we have verified that only a few isolated molecule intramolecular modes of vibration are important to describe correctly the lattice modes in the range from  $4\text{ cm}^{-1}$  to  $120\text{ cm}^{-1}$  for the systems we studied.

The introduction of flexibility inside the dynamical matrix (see Equation 3.4) requires the evaluation of the numerical derivatives:

$$H_{ij} = \frac{\partial^2 \varphi}{\partial u_i \partial u_j} \quad (12.1)$$

where the displacements  $u$  can either be molecular or intramolecular. At the same time the energy  $\varphi$  depends both on the lattice energy and on the conformational changes.

For the calculation of these derivatives we need to calculate the energy of several molecular geometries of each molecule in the unit cell with a sufficiently accurate electronic structure method, according to the intramolecular displacement of the normal modes we wish to consider; furthermore, we need to calculate the lattice energy for each of the resulting conformations of the molecules. This procedure will require a significant modification of the DMACRYS subroutine dedicated to the calculation of the vibrational properties of the crystal, but it will allow us to use the good multipole model for the description of the electrostatics, and introduce flexibility without the need to define parameters for the description of the intramolecular degrees of freedom.

The temperature dependence of the positions and absorption intensities of the absorption peaks is interesting to study, as it gives information on the anharmonicity of the vibrational modes, and of the population of the excited vibrational states at finite temperature. These effects are automatically taken into account by molecular dynamics methods [234], which have the drawback of being slow compared to our DMACRYS calculations (if the forces are modelled by forcefield, while *ab initio* molecular dynamics methods [235] are prohibitively expensive except for very small crystals, like ammonia [236]).

We plan to take into account these effects by sampling the potential

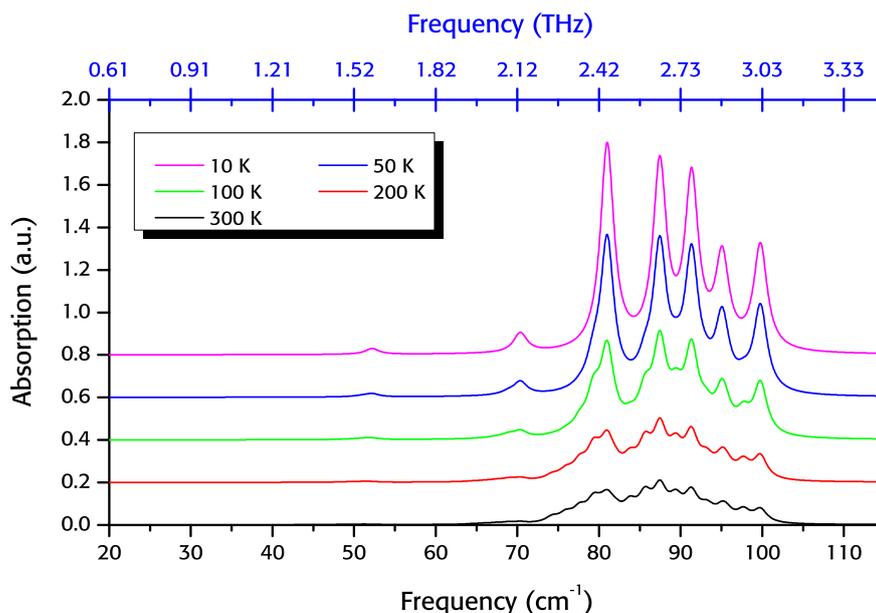


Figure 12.1. Simulated DMACRYS spectra of paracetamol form I, at different temperatures. The absorption frequencies with the formula in Equation 12.4, with  $\chi_e = 0.01$ . Five energy levels for each absorption mode are plotted. The absorption intensities are normalised to the strongest peak at 10 K

energy surface of the crystal, by considering the energy variation in the lattice energy associated with displacements along each of the eigenvectors. Once we obtain the shape of the potential energy landscape, we can approximate it to an anharmonic potential energy function with known vibrational states and energies. For example, the energy levels  $E_n$  of the Morse potential

$$V(R) = D_e(1 - e^{-a(R-R_0)})^2 \quad (12.2)$$

for the displacement of a molecule from its equilibrium position  $R_0$  can be expressed as

$$E_n = h\nu_0(n + 1/2) - \frac{(h\nu_0(n + 1/2))^2}{4D_e}, \quad (12.3)$$

with  $\nu_0$  depending on the parameters  $D_e$  and  $a$ . Absorption intensities can therefore be modelled by considering the Boltzmann population of the vibrational levels and the frequency difference between two energy levels  $E_{n+1}$  and  $E_n$ .

To exemplify this approach, we have simulated the absorption spec-

tra at varying temperatures for paracetamol form I, by considering the lowest energy levels with a model system

$$E_n = \omega_e(n + 1/2) - \omega_e\chi_e(n + 1/2)^2, \quad (12.4)$$

where  $\omega_e$  are the calculated harmonic frequencies and  $\chi_e$  is an anharmonicity parameter. As shown in Figure 12.1, there is significant change in the spectrum as temperature is increased, with the hot bands of each absorption becoming more important at higher temperatures.

In this example we used an arbitrary value for the anharmonicity,  $\chi_e = 0.01$ ; with the sampling of the potential energy surface it will be possible to get the anharmonicity constants (which are in principle different for each of the absorption modes) directly from our calculations, and to compare the data with the temperature dependent spectra to validate our model.

## VARIABLE TEMPERATURE SPECTRA

In this appendix we report all the measured variable temperature spectra, not normalised. For each measurement we used a tablet containing 360 mg polyethylene and 15 mg API. All the measurements in each graph are relative to one tablet at different temperatures.

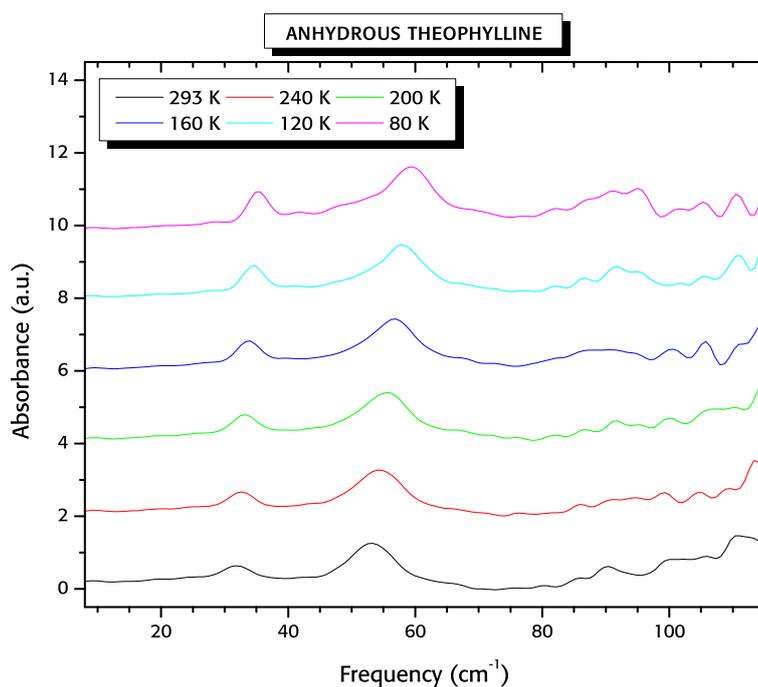


Figure A.1. Spectra of anhydrous theophylline in the range 80 K to 293 K

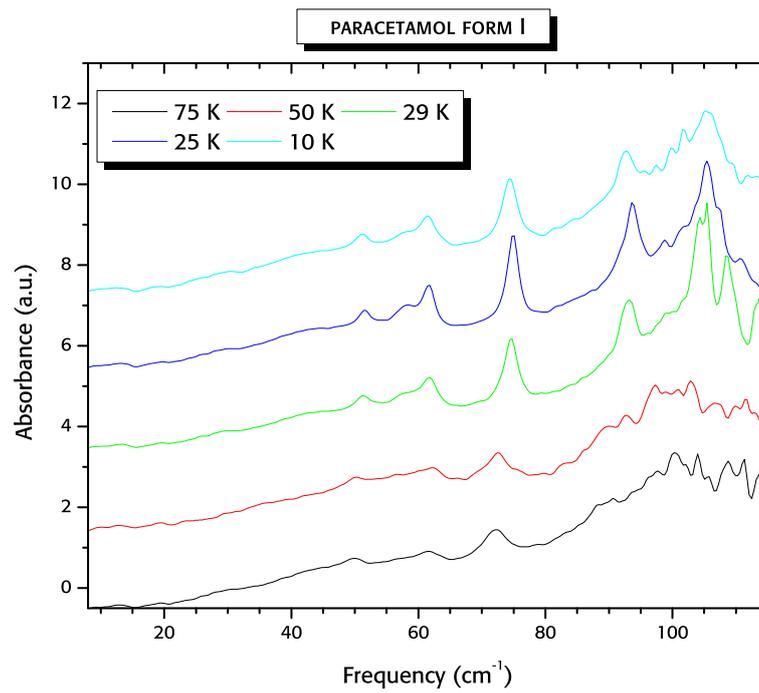
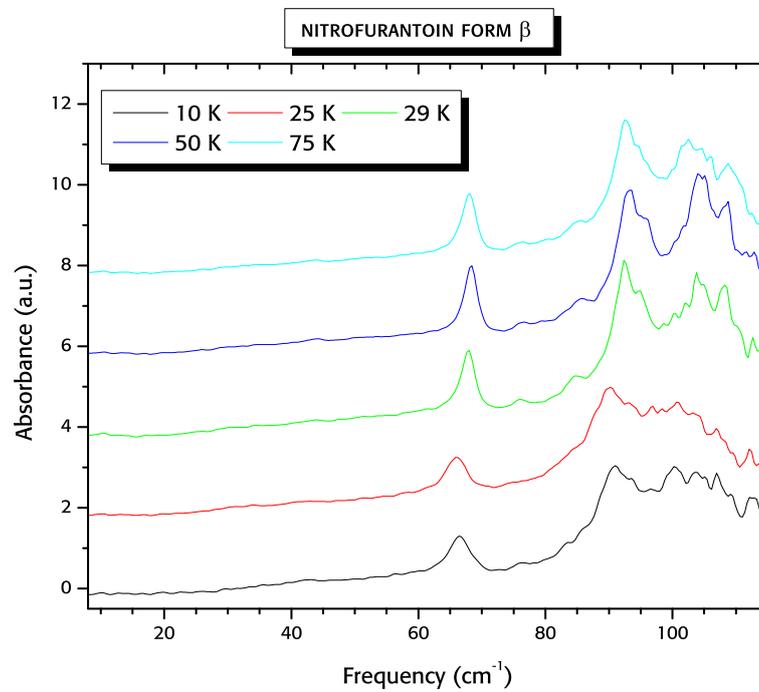


Figure A.2. Spectra of paracetamol form 1 in the range 80 K to 293 K

Figure A.3. Spectra of nitrofurantoin  $\beta$  in the range 80 K to 293 K

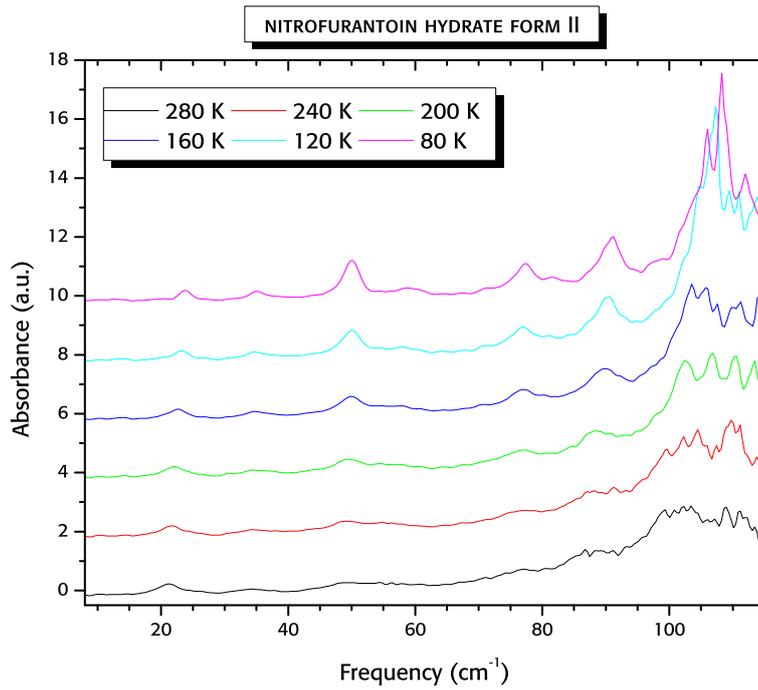


Figure A.4. Spectra of nitrofurantoin hydrate form II in the range 80 K to 293 K

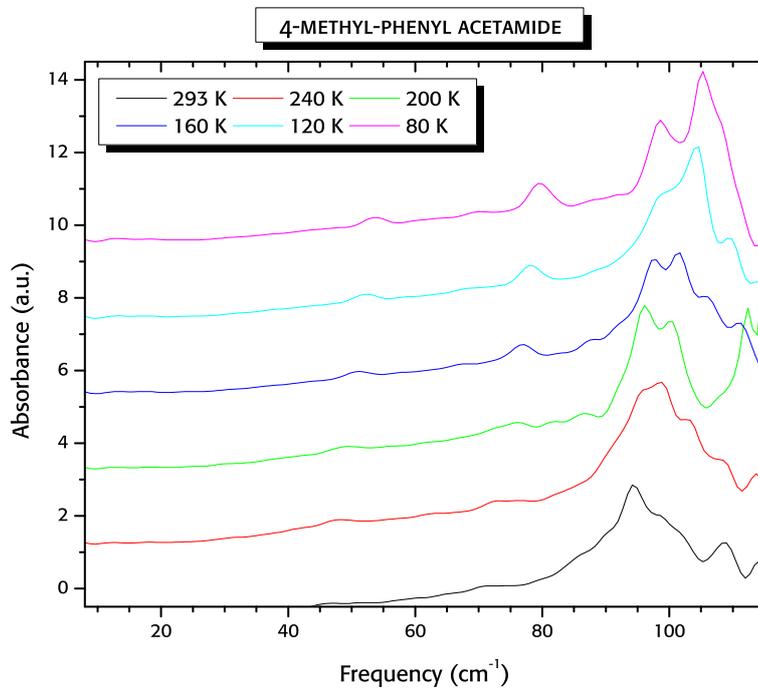


Figure A.5. Spectra of 4-methyl-phenyl acetamide in the range 80 K to 293 K

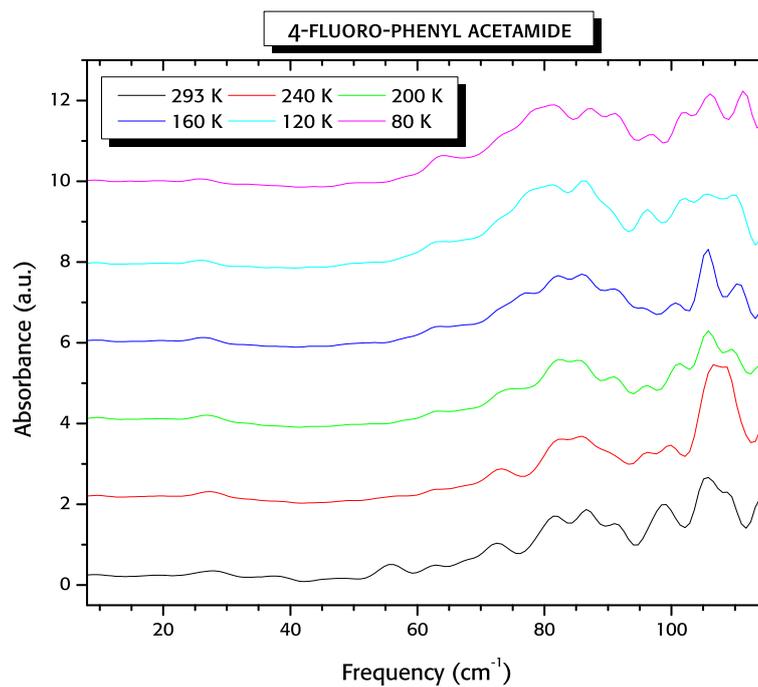


Figure A.6. Spectra of 4-fluoro-phenyl acetamide in the range 80 K to 293 K

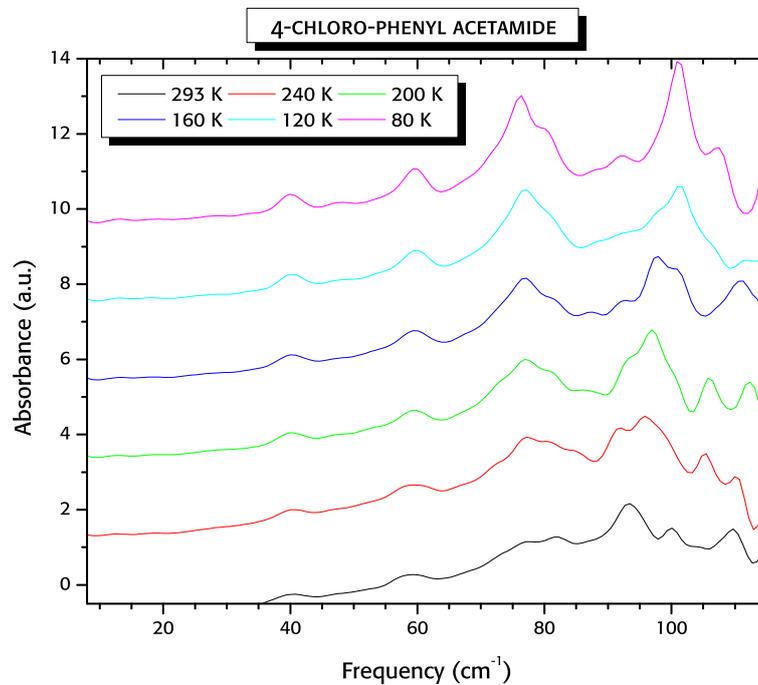


Figure A.7. Spectra of 4-chloro-phenyl acetamide in the range 80 K to 293 K

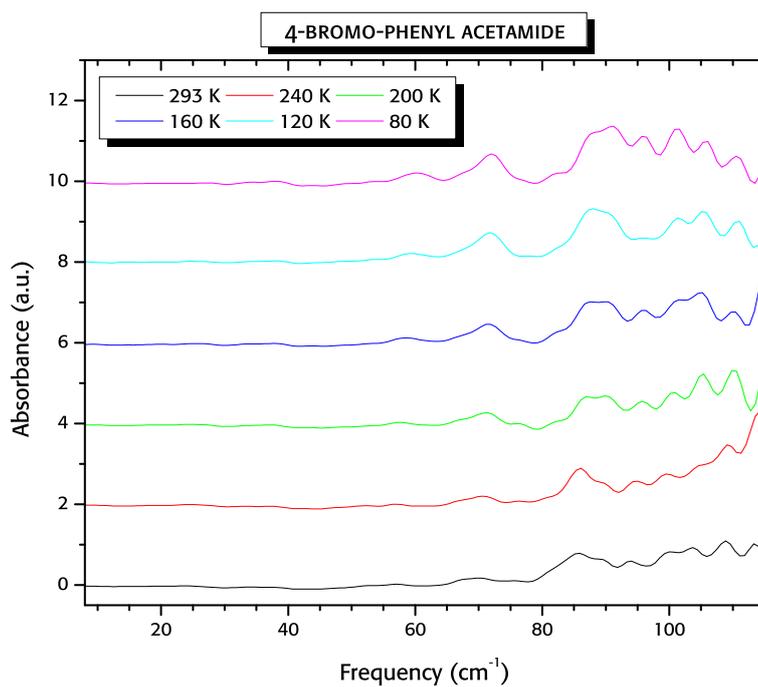


Figure A.8. Spectra of 4-bromo-phenyl acetamide in the range 80 K to 293 K

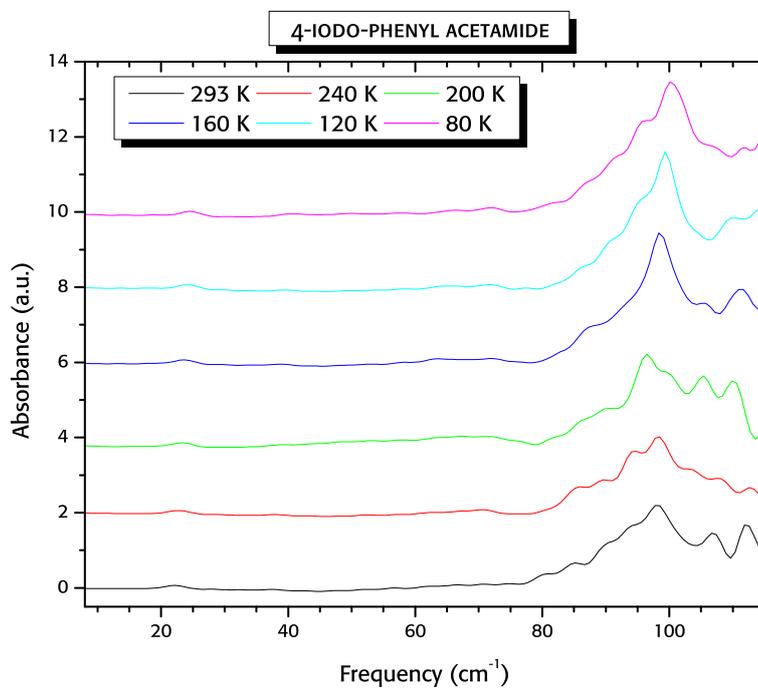


Figure A.9. Spectra of 4-iodo-phenyl acetamide in the range 80 K to 293 K



## BIBLIOGRAPHY

---

- [1] N. Sirota, Certain Problems of Polymorphism (I), *Crystal Research and Technology*, 1982, **17**, 661–691.
- [2] C. G. Salzmann *et al.*, The polymorphism of ice: five unresolved questions, *Physical Chemistry Chemical Physics*, 2011, **13**, 18468–18480.
- [3] R. Rogers, *Polymorphism in crystals: fundamentals, prediction and industrial practice : special issue*, ACS Publications, 2003.
- [4] T. L. Threlfall, Analysis of organic polymorphs. A Review, *Analyst*, 1995, **120**, 2435–2460.
- [5] S. Datta and D. J. W. Grant, Crystal structures of drugs: advances in determination, prediction and engineering, *Nature Reviews Drug Discovery*, 2004, **3**, 42–57.
- [6] A. Grunenberg *et al.*, Theoretical derivation and practical application of energy/temperature diagrams as an instrument in preformulation studies of polymorphic drug substances, *International Journal of Pharmaceutics*, 1996, **129**, 147–158.
- [7] R. Storey *et al.*, Automation of solid form screening procedures in the pharmaceutical industry – How to avoid the bottlenecks, *Crystallography Reviews*, 2004, **10**, 45–56.
- [8] R. Hilfiker, *Polymorphism in the pharmaceutical industry*, Wiley-VCH, 2006.
- [9] R. Storey and I. Ymén, *Solid State Characterization of Pharmaceuticals*, John Wiley & Sons, 2011.
- [10] J. Haleblan and W. McCrone, Pharmaceutical applications of polymorphism, *Journal of Pharmaceutical Sciences*, 1969, **58**, 911–929.
- [11] H. G. Brittain, Effects of mechanical processing on phase composition, *Journal of Pharmaceutical Sciences*, 2002, **91**, 1573–1580.
- [12] J. K. Haleblan, Characterization of habits and crystalline modification of solids and their pharmaceutical applications, *Journal of Pharmaceutical Sciences*, 1975, **64**, 1269–1288.
- [13] J. Swarbrick and J. Boylan, *Encyclopedia of Pharmaceutical Technology*, M. Dekker, 2001.

- [14] S. Byrn *et al.*, Pharmaceutical Solids: A Strategic Approach to Regulatory Considerations, *Pharmaceutical Research*, 1995, **12**, 945–954.
- [15] D. A. Snider *et al.*, Polymorphism in generic drug product development, *Advanced Drug Delivery Reviews*, 2004, **56**, 391 – 395.
- [16] D. Singhal and W. Curatolo, Drug polymorphism and dosage form design: a practical perspective, *Advanced Drug Delivery Reviews*, 2004, **56**, 335–347.
- [17] P. Di Martino *et al.*, A new pure paracetamol for direct compression: The orthorhombic form, *International Journal of Pharmaceutics*, 1996, **128**, 1–8.
- [18] E. Joiris *et al.*, Compression Behavior of Orthorhombic Paracetamol, *Pharmaceutical Research*, 1998, **15**, 1122–1130.
- [19] W. McCrone, in *Physics and chemistry of the organic solid state*, Interscience Publishers, 1963, ch. 8.
- [20] A. Gavezzotti, *Molecular aggregation: structure analysis and molecular simulation of crystals and liquids*, Oxford University Press, 2007, vol. 19.
- [21] D. J. Kempf *et al.*, 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*, 1995, **92**, 2484–2488.
- [22] D. D. Ho *et al.*, Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection, *Nature*, 1995, **373**, 123–126.
- [23] J. Bauer *et al.*, Ritonavir: An Extraordinary Example of Conformational Polymorphism, *Pharmaceutical Research*, 2001, **18**, 859–866.
- [24] D. J. Kempf *et al.*, Discovery of Ritonavir, a Potent Inhibitor of HIV Protease with High Oral Bioavailability and Clinical Efficacy, *Journal of Medicinal Chemistry*, 1998, **41**, 602–617.
- [25] S. R. Chemburkar *et al.*, Dealing with the Impact of Ritonavir Polymorphs on the Late Stages of Bulk Drug Process Development, *Organic Process Research and Development*, 2000, **4**, 413–417.
- [26] A. Laboratories, *Highlights of prescribing information: NORVIR*, online [http://www.rxabbott.com/pdf/norviretab\\_pi.pdf](http://www.rxabbott.com/pdf/norviretab_pi.pdf), 2011 (last revision).

- [27] D. J. Grant and S. R. Byrn, A timely re-examination of drug polymorphism in pharmaceutical development and regulation, *Advanced Drug Delivery Reviews*, 2004, **56**, 237–239.
- [28] S. Byrn *et al.*, *Solid-state chemistry of drugs*, SSCI, Inc., 1999.
- [29] S. L. Morissette *et al.*, Elucidation of crystal form diversity of the HIV protease inhibitor ritonavir by high-throughput crystallization, *Proceedings of the National Academy of Sciences*, 2003, **100**, 2180–2184.
- [30] J. Glusker and K. Trueblood, *Crystal Structure Analysis: A Primer*, Oxford University Press, 2010.
- [31] B. Cullity and S. Stock, *Elements of x-ray diffraction*, Prentice Hall, 2001.
- [32] A. J. Florence *et al.*, Solving molecular crystal structures from laboratory X-ray powder diffraction data with *DASH*: the state of the art and challenges, *Journal of Applied Crystallography*, 2005, **38**, 249–259.
- [33] W. David *et al.*, *Structure Determination from Powder Diffraction Data*, Oxford University Press, 2006.
- [34] F. H. Allen, The Cambridge Structural Database: a quarter of a million crystal structures and rising, *Acta Crystallographica Section B*, 2002, **58**, 380–388.
- [35] P. T. A. Galek *et al.*, One in half a million: a solid form informatics study of a pharmaceutical crystal structure, *CrystEngComm*, 2012, **14**, 2391–2403.
- [36] I. J. Bruno *et al.*, New software for searching the Cambridge Structural Database and visualizing crystal structures, *Acta Crystallographica Section B*, 2002, **58**, 389–397.
- [37] D. E. Braun *et al.*, Structural and Thermodynamic Features of Crystal Polymorphs of R-Cinacalcet Hydrochloride, *Crystal Growth & Design*, 2008, **8**, 4109–4119.
- [38] P. Gabbott, *Principles and applications of thermal analysis*, Blackwell Pub., 2008.
- [39] J. Ford and P. Timmins, *Pharmaceutical thermal analysis: techniques and applications*, E. Horwood, 1989.
- [40] P. Ureña Torres, Cinacalcet HCl: A Novel Treatment for Secondary Hyperparathyroidism Caused by Chronic Kidney Disease, *Journal of renal nutrition*, 2006, **16**, 253–258.

- [41] R. Khankari *et al.*, Physical characterization of nedocromil sodium hydrates, *Journal of Pharmaceutical Sciences*, 1998, **87**, 1052–1061.
- [42] K. Kogermann *et al.*, Establishing quantitative in-line analysis of multiple solid-state transformations during dehydration, *Journal of Pharmaceutical Sciences*, 2008, **97**, 1520–6017.
- [43] K. A. Alkhamis *et al.*, Comparison between dehydration and desolvation kinetics of fluconazole monohydrate and fluconazole ethylacetate solvate using three different methods, *Journal of Pharmaceutical Sciences*, 2006, **95**, 859–870.
- [44] U. Holzgrabe *et al.*, *NMR Spectroscopy in Drug Development and Analysis*, John Wiley & Sons, 2008.
- [45] K. Masuda *et al.*, Solid-state  $^{13}\text{C}$  NMR study of indomethacin polymorphism, *International Journal of Pharmaceutics*, 2006, **318**, 146–153.
- [46] R. K. Harris, NMR studies of organic polymorphs & solvates, *Analyst*, 2006, **131**, 351–373.
- [47] W. Guillory, *Introduction to molecular structure and spectroscopy*, Allyn and Bacon, 1977.
- [48] A. Smith, *Applied infrared spectroscopy: fundamentals, techniques, and analytical problem-solving*, Wiley, 1979.
- [49] V. Giannellini *et al.*, Solid-state study of mepivacaine hydrochloride, *Journal of Pharmaceutical and Biomedical Analysis*, 2005, **39**, 444–454.
- [50] K. Kipouros *et al.*, Simultaneous quantification of carbamazepine crystal forms in ternary mixtures (I, III, and IV) by diffuse reflectance FTIR spectroscopy (DRIFTS) and multivariate calibration, *Journal of Pharmaceutical Sciences*, 2006, **95**, 2419–2431.
- [51] T. Kojima *et al.*, Evaluation of hydrate formation of a pharmaceutical solid by using diffuse reflectance infrared Fourier-transform spectroscopy, *Journal of Pharmaceutical and Biomedical Analysis*, 2008, **46**, 788–791.
- [52] P. K. Aldridge *et al.*, Near-IR Detection of Polymorphism and Process-Related Substances, *Analytical Chemistry*, 1996, **68**, 997–1002.
- [53] A. D. Patel *et al.*, Low-level determination of polymorph composition in physical mixtures by near-infrared reflectance spectroscopy, *Journal of Pharmaceutical Sciences*, 2001, **90**, 360–370.

- [54] J. F. Kauffman *et al.*, Near infrared spectroscopy of magnesium stearate hydrates and multivariate calibration of pseudopolymorph composition, *Journal of Pharmaceutical Sciences*, 2008, **97**, 2757–2767.
- [55] Y.-C. Shen, Terahertz pulsed spectroscopy and imaging for pharmaceutical applications: A Review, *International Journal of Pharmaceutics*, 2011, **417**, 48–60.
- [56] J. A. Zeitler *et al.*, Temperature dependent terahertz pulsed spectroscopy of carbamazepine, *Thermochimica Acta*, 2005, **436**, 71–77.
- [57] J. Zeitler *et al.*, Drug hydrate systems and dehydration processes studied by terahertz pulsed spectroscopy, *International Journal of Pharmaceutics*, 2007, **334**, 78–84.
- [58] J. A. Zeitler *et al.*, Characterization of temperature-induced phase transitions in five polymorphic forms of sulfathiazole by terahertz pulsed spectroscopy and differential scanning calorimetry, *Journal of Pharmaceutical Sciences*, 2006, **95**, 2486–2498.
- [59] L. Ning *et al.*, Study on the THz spectrum of methamphetamine, *Optics Express*, 2005, **13**, 6750–6755.
- [60] G. M. Day *et al.*, Understanding the influence of polymorphism on phonon spectra: lattice dynamics calculations and terahertz spectroscopy of carbamazepine, *Journal of Physical Chemistry B*, 2006, **110**, 447–456.
- [61] K. Siegrist *et al.*, High-Resolution Terahertz Spectroscopy of Crystalline Trialanine: Extreme Sensitivity to  $\beta$ -Sheet Structure and Cocrystallized Water, *Journal of the American Chemical Society*, 2006, **128**, 5764–5775.
- [62] M. D. King and T. M. Korter, Application of London-Type Dispersion Corrections in Solid-State Density Functional Theory for Predicting the Temperature-Dependence of Crystal Structures and Terahertz Spectra, *Crystal Growth & Design*, 2011, **11**, 2006–2010.
- [63] J. Zeitler *et al.*, in *Terahertz spectroscopy – Principles and applications*, Taylor & Francis Group, 2008, ch. 8, pp. 299–323.
- [64] C. A. Schmuttenmaer, Exploring Dynamics in the Far-Infrared with Terahertz Spectroscopy, *Chemical Reviews*, 2004, **104**, 1759–1780.
- [65] A. Borak, Toward bridging the terahertz gap with silicon-based lasers, *Science*, 2005, **308**, 638.

- [66] D. H. Martin and K. Mizuno, The generation of coherent submillimetre waves, *Advances in Physics*, 1976, **25**, 211–246.
- [67] S. D. Ganichev and W. Prettl, *Intense terahertz excitation of semiconductors*, Oxford University Press, 2006.
- [68] R. Kohler *et al.*, Terahertz semiconductor-heterostructure laser, *Nature*, 2002, **417**, 156–159.
- [69] M. Inguscio *et al.*, A review of frequency measurements of optically pumped lasers from 0.1 to 8 THz, *Journal of Applied Physics*, 1986, **60**, R161–R192.
- [70] A. Rostami *et al.*, *Terahertz Technology: Fundamentals and Applications*, Springer, 2010.
- [71] D. H. Auston, Picosecond optoelectronic switching and gating in silicon, *Applied Physics Letters*, 1975, **26**, 101–103.
- [72] D. H. Auston and K. P. Cheung, Coherent time-domain far-infrared spectroscopy, *Journal of the Optical Society of America B*, 1985, **2**, 606–612.
- [73] P. Smith *et al.*, Subpicosecond photoconducting dipole antennas, *Journal of Quantum Electronics*, 1988, **24**, 255–260.
- [74] M. van Exter *et al.*, Terahertz time-domain spectroscopy of water vapor, *Optical Letters*, 1989, **14**, 1128–1130.
- [75] C. Balanis, *Advanced engineering electromagnetics*, John Wiley & Sons, 1989.
- [76] R. King and G. Smith, *Antennas in matter*, MIT Press, 1981.
- [77] A. Nahata *et al.*, A wideband coherent terahertz spectroscopy system using optical rectification and electro-optic sampling, *Applied Physics Letters*, 1996, **69**, 2321–2323.
- [78] Q. Wu and X.-C. Zhang, Free-space electro-optic sampling of terahertz beams, *Applied Physics Letters*, 1995, **67**, 3523–3525.
- [79] M. Walther, *Ph.D. thesis*, Albert-Ludwigs-Universität – Freiburg im Breisgau, 2003.
- [80] P. F. Bernath, The spectroscopy of water vapour: Experiment, theory and applications, *Physical Chemistry Chemical Physics*, 2002, **4**, 1501–1509.
- [81] R. J. Foltynowicz *et al.*, Terahertz absorption measurement for gas-phase 2,4-dinitrotoluene from 0.05 THz to 2.7 THz, *Chemical Physics Letters*, 2006, **431**, 34–38.

- [82] O. Pirali *et al.*, Far-infrared spectroscopy of small polycyclic aromatic hydrocarbons, *Physical Chemistry Chemical Physics*, 2006, **8**, 3707–3714.
- [83] S. C. Capelli *et al.*, Molecular Motion in Crystalline Naphthalene: Analysis of Multi-Temperature X-Ray and Neutron Diffraction Data, *The Journal of Physical Chemistry A*, 2006, **110**, 11695–11703.
- [84] I. Harada and T. Shimanouchi, Normal Vibrations and Intermolecular Forces of Crystalline Benzene and Naphthalene, *The Journal of Chemical Physics*, 1966, **44**, 2016–2028.
- [85] G. Taddei *et al.*, The calculation of the normal coordinates of molecular crystals using pairwise potential functions, *Chemical Physics Letters*, 1972, **13**, 136–139.
- [86] P. Hermet *et al.*, Far- and Mid-Infrared of Crystalline 2,2'-Bithiophene: Ab Initio Analysis and Comparison with Infrared Response, *The Journal of Physical Chemistry A*, 2005, **109**, 1684–1691.
- [87] A. M. Fedor *et al.*, The terahertz spectrum and quantum chemical assignment of 2,2'-bithiophene in cyclohexane, *Vibrational Spectroscopy*, 2009, **49**, 124–132.
- [88] G. Wang *et al.*, Vibrational spectra of ketamine hydrochloride and 3, 4-methylenedioxymethamphetamine in terahertz range, *Journal of Applied Physics*, 2007, **102**, 013106.
- [89] D. G. Allis *et al.*, The solid-state terahertz spectrum of MDMA (Ecstasy)—A unique test for molecular modeling assignments, *Chemical Physics Letters*, 2008, **463**, 353–356.
- [90] P. Hakey *et al.*, Density Functional Dependence in the Theoretical Analysis of the Terahertz Spectrum of the Illicit Drug MDMA (Ecstasy), *Sensors*, 2010, **10**, 478–484.
- [91] G. Grosso and G. P. Parravicini, *Solid State Physics*, Academic Press, 1st edn., 2000.
- [92] C. Cohen-Tannoudji *et al.*, *Quantum Mechanics*, John Wiley & Sons, 1991.
- [93] W. Ludwig, *Recent developments in lattice theory*, Springer, 1967.
- [94] A. J. Pertsin and A. I. Kitaigorodskii, *The atom-atom potential method: applications to organic molecular solids*, Springer-Verlag, 1987, vol. 43.

- [95] E. W. Kellermann, Theory of the Vibrations of the Sodium Chloride Lattice, *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 1940, **238**, 513–548.
- [96] G. Raunio and S. Rolandson, Lattice Dynamics of NaCl, KCl, RbCl, and RbF, *Physical Review B*, 1970, **2**, 2098–2103.
- [97] M. P. Verma and B. Dayal, Lattice Dynamics of MgO, *Physica Status Solidi (B)*, 1967, **19**, 751–756.
- [98] G. Raunio *et al.*, Phonon Dispersion Relations in NaCl, *Physical Review*, 1969, **178**, 1496–1501.
- [99] A. M. Karo and J. R. Hardy, Precise Vibrational Frequency Distributions and the Second-Order Raman Spectrum and Specific Heat of NaCl, *Physical Review*, 1966, **141**, 696–710.
- [100] U. Schröder, A new model for lattice dynamics (“breathing shell model”), *Solid State Communications*, 1966, **4**, 347–349.
- [101] W. Dyck, Lattice dynamics of alkali hydrides and euterides with the NaCl type structure, *Journal of Physics C: Solid State Physics*, 1981, **14**, 4193.
- [102] *List of quantum chemistry and solid-state physics software*, online [en.wikipedia.org/wiki/List\\_of\\_quantum\\_chemistry\\_and\\_solid\\_state\\_physics\\_software](http://en.wikipedia.org/wiki/List_of_quantum_chemistry_and_solid_state_physics_software), Last consulted: January 2012.
- [103]  $\Psi_k$ . *Available Electronic Structure Codes*, online [psi-k.org/codes.shtml](http://psi-k.org/codes.shtml), Last consulted: January 2012.
- [104] S. Baroni *et al.*, Green’s-function approach to linear response in solids, *Physical Review Letters*, 1987, **58**, 1861–1864.
- [105] J. D. Gale, *General Utility Lattice Program Version 3.4*, [http://projects.ivec.org/gulp/help/gulp3.4\\_manual.pdf](http://projects.ivec.org/gulp/help/gulp3.4_manual.pdf).
- [106] D. Williams, Nonbonded Potential Parameters Derived from Crystalline Aromatic Hydrocarbons, *The Journal of Chemical Physics*, 1966, **45**, 3770–3779.
- [107] S. R. Cox *et al.*, Nonbonded potential function models for crystalline oxohydrocarbons, *Acta Crystallographica Section A*, 1981, **37**, 293–301.
- [108] D. E. Williams and S. R. Cox, Nonbonded potentials for azahydrocarbons: the importance of the Coulombic interaction, *Acta Crystallographica Section B*, 1984, **40**, 404–417.

- [109] D. E. Williams and D. J. Houpt, Fluorine nonbonded potential parameters derived from crystalline perfluorocarbons, *Acta Crystallographica Section B*, 1986, **42**, 286–295.
- [110] R. Righini *et al.*, Lattice dynamics of crystalline ammonia and deuterio-ammonia, *The Journal of Chemical Physics*, 1978, **33**, 345–353.
- [111] G. M. Day and S. L. Price, A nonempirical anisotropic atom-atom model potential for chlorobenzene crystals, *Journal of the American Chemical Society*, 2003, **125**, 16434–16443.
- [112] S. H. Walmsley, in *Lattice dynamics and intermolecular forces*, Academic Press, 1975, vol. 55, p. 82.
- [113] A. J. Stone and M. Alderton, Distributed multipole analysis: methods and applications, *Molecular Physics*, 1985, **56**, 1047–1064.
- [114] M. J. Frisch *et al.*, *Gaussian 03, Revision D.01*, Gaussian, Inc., Wallingford, CT, 2004.
- [115] A. D. Buckingham *et al.*, Electrostatic predictions of shapes and properties of Van der Waals molecules, *International Reviews in Physical Chemistry*, 1986, **5**, 107–114.
- [116] P. D. DeCicco and F. A. Johnson, The Quantum Theory of Lattice Dynamics. IV, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 1969, **310**, 111–119.
- [117] P. Hohenberg and W. Kohn, Inhomogeneous Electron Gas, *Physical Review*, 1964, **136**, B864–B871.
- [118] W. Kohn and L. J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, *Physical Review*, 1965, **140**, A1133–A1138.
- [119] R. G. Parr and W. Yang, *Density-functional theory of atoms and molecules*, Oxford science publications, 1st edn., 1989.
- [120] P. A. M. Dirac, Note on Exchange Phenomena in the Thomas Atom, *Mathematical Proceedings of the Cambridge Philosophical Society*, 1930, **26**, 376–385.
- [121] P. Nozières and D. Pines, Correlation Energy of a Free Electron Gas, *Physical Review*, 1958, **111**, 442–454.
- [122] M. Gell-Mann and K. A. Brueckner, Correlation Energy of an Electron Gas at High Density, *Physical Review*, 1957, **106**, 364–368.

- [123] W. J. Carr and A. A. Maradudin, Ground-State Energy of a High-Density Electron Gas, *Physical Review*, 1964, **133**, A371–A374.
- [124] D. M. Ceperley and B. J. Alder, Ground State of the Electron Gas by a Stochastic Method, *Physical Review Letters*, 1980, **45**, 566–569.
- [125] J. P. Perdew and A. Zunger, Self-interaction correction to density-functional approximations for many-electron systems, *Physical Review B*, 1981, **23**, 5048–5079.
- [126] S. H. Vosko *et al.*, Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis, *Canadian Journal of Physics*, 1980, **58**, 1200–1211.
- [127] J. P. Perdew *et al.*, in *Uniform Density Limit of Exchange-Correlation Energy Functionals*, ACS, 2007, ch. 3, pp. 13–25.
- [128] A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, *Physical Review A*, 1988, **38**, 3098–3100.
- [129] C. Fiolhais *et al.*, *A primer in density functional theory*, Springer, 2003.
- [130] J. P. Perdew *et al.*, Generalized Gradient Approximation Made Simple, *Physical Review Letters*, 1996, **77**, 3865–3868.
- [131] P. J. Stephens *et al.*, Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields, *The Journal of Physical Chemistry*, 1994, **98**, 11623–11627.
- [132] P. J. Stephens *et al.*, in *Comparison of Local, Nonlocal, and Hybrid Density Functionals Using Vibrational Absorption and Circular Dichroism Spectroscopy*, ACS, 1996, ch. 8, pp. 105–113.
- [133] F. O. Kannemann and A. D. Becke, Van der Waals Interactions in Density-Functional Theory: Rare-Gas Diatomics, *Journal of Chemical Theory and Computation*, 2009, **5**, 719–727.
- [134] F. Shimojo *et al.*, Density functional study of 1,3,5-trinitro-1,3,5-triazine molecular crystal with van der Waals interactions, *The Journal of Chemical Physics*, 2010, **132**, 094106.
- [135] S. G. Boyd and K. J. Boyd, A computational analysis of the interaction of lattice and intramolecular vibrational modes in crystalline  $\alpha$ -RDX, *The Journal of Chemical Physics*, 2008, **129**, 134502.

- [136] S. Grimme *et al.*, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, *The Journal of Chemical Physics*, 2010, **132**, 154104.
- [137] C. Herring, A New Method for Calculating Wave Functions in Crystals, *Physical Review*, 1940, **57**, 1169–1177.
- [138] N. Ashcroft, Electron-ion pseudopotentials in metals, *Physics Letters*, 1966, **23**, 48–50.
- [139] D. R. Hamann *et al.*, Norm-conserving pseudopotentials, *Physical Review Letters*, 1979, **43**, 1494–1497.
- [140] D. R. Hamann, Generalized norm-conserving pseudopotentials, *Physical Review B*, 1989, **40**, 2980–2984.
- [141] D. Vanderbilt, Soft self-consistent pseudopotentials in a generalized eigenvalue formalism, *Physical Review B*, 1990, **41**, 7892–7895.
- [142] D. Vanderbilt, Optimally smooth norm-conserving pseudopotentials, *Physical Review B*, 1985, **32**, 8412–8415.
- [143] A. R. Leach, *Molecular modelling: principles and applications*, Prentice Hall, 2001.
- [144] D. J. Willock *et al.*, The relaxation of molecular crystal structures using a distributed multipole electrostatic model, *Journal of Computational Chemistry*, 1995, **16**, 628–647.
- [145] S. L. Price *et al.*, Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials, *Physical Chemistry Chemical Physics*, 2010, **12**, 8478–8490.
- [146] G. M. Sheldrick, A short history of SHELX, *Acta Crystallographica Section A*, 2008, **64**, 112–122.
- [147] W. J. Stevens *et al.*, Compact effective potentials and efficient shared-exponent basis sets for the first- and second-row atoms, *The Journal of Chemical Physics*, 1984, **81**, 6026–6033.
- [148] W. J. Stevens *et al.*, Relativistic compact effective potentials and efficient, shared-exponent basis sets for the third-, fourth- and fifth-row atoms, *Canadian Journal of Chemistry*, 1992, **70**, 612–630.
- [149] T. R. Cundari and W. J. Stevens, Effective core potential methods for the lanthanides, *The Journal of Chemical Physics*, 1993, **98**, 5555–5565.

- [150] R. Ditchfield *et al.*, Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules, *The Journal of Chemical Physics*, 1971, **54**, 724–728.
- [151] W. J. Hehre *et al.*, Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules, *The Journal of Chemical Physics*, 1972, **56**, 2257–2261.
- [152] P. C. Hariharan and J. A. Pople, The influence of polarization functions on molecular orbital hydrogenation energies, *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 1973, **28**, 213–222.
- [153] P. Hariharan and J. Pople, Accuracy of  $AH_n$  equilibrium geometries by single determinant molecular orbital theory, *Molecular Physics*, 1974, **27**, 209–214.
- [154] A. J. Stone, Distributed Multipole Analysis: Stability for Large Basis Sets, *Journal of Chemical Theory and Computation*, 2005, **1**, 1128–1132.
- [155] D. S. Coombes *et al.*, Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study, *The Journal of Physical Chemistry*, 1996, **100**, 7352–7360.
- [156] T. Beyer and S. L. Price, Dimer or Catemer? Low-Energy Crystal Packings for Small Carboxylic Acids, *The Journal of Physical Chemistry B*, 2000, **104**, 2647–2655.
- [157] D. Williams, Improved intermolecular force field for crystalline hydrocarbons containing four- or three-coordinated carbon, *Journal of Molecular Structure*, 1999, **485–486**, 321 – 347.
- [158] D. E. Williams, Improved intermolecular force field for crystalline oxohydrocarbons including O–H ··· O hydrogen bonding, *Journal of Computational Chemistry*, 2001, **22**, 1–20.
- [159] D. E. Williams, Improved intermolecular force field for molecules containing H, C, N, and O atoms, with applications to nucleoside and peptide crystals, *Journal of Computational Chemistry*, 2001, **22**, 1154–1166.
- [160] C. Kittel, *Introduction to solid state physics*, Wiley, 2005.
- [161] M. Churchill, Carbon-hydrogen and nitrogen-hydrogen distances assumed in, and determined from, recent x-ray diffraction studies on inorganic complexes, *Inorganic Chemistry*, 1973, **12**, 1213–1214.

- [162] G. Filippini and A. Gavezzotti, Empirical intermolecular potentials for organic crystals: the '6-exp' approximation revisited, *Acta Crystallographica*, 1993, **B49**, 868–880.
- [163] G. M. Day, *Ph.D. thesis*, University College London, 2003.
- [164] R. M. Hanson, *Jmol* – a paradigm shift in crystallographic visualization, *Journal of Applied Crystallography*, 2010, **43**, 1250–1260.
- [165] S. J. Clark *et al.*, First principles methods using CASTEP, *Zeitschrift für Kristallographie*, 2005, **220**, 567–570.
- [166] F. Ortman *et al.*, Semiempirical van der Waals correction to the density functional description of solids and molecular structures, *Phys. Rev. B*, 2006, **73**, 205101.
- [167] S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction, *Journal of Computational Chemistry*, 2006, **27**, 1787–1799.
- [168] P. Jurečka *et al.*, Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations, *Journal of Computational Chemistry*, 2007, **28**, 555–569.
- [169] A. Tkatchenko and M. Scheffler, Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data, *Phys. Rev. Lett.*, 2009, **102**, 073005.
- [170] A. Stone, *The theory of intermolecular forces*, Clarendon Press, 1997.
- [171] Q. Wu and W. Yang, Empirical correction to density functional theory for van der Waals interactions, *The Journal of Chemical Physics*, 2002, **116**, 515–524.
- [172] M. A. Spackman, Molecular electric moments from x-ray diffraction data, *Chemical Reviews*, 1992, **92**, 1769–1797.
- [173] C. A. Coulson and D. Eisenberg, Interactions of H<sub>2</sub>O Molecules in Ice. I. The Dipole Moment of an H<sub>2</sub>O Molecule in Ice, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 1966, **291**, 445–453.
- [174] B. Chen *et al.*, Hydrogen Bonding in Water, *Physical Review Letters*, 2003, **91**, 215503.
- [175] G. G. Hall and C. M. Smith, Fitting electron densities of molecules, *International Journal of Quantum Chemistry*, 1984, **25**, 881–890.

- [176] C. M. Smith and G. G. Hall, The approximation of electron densities, *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 1986, **69**, 63–69.
- [177] R. S. Mulliken, Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I, *The Journal of Chemical Physics*, 1955, **23**, 1833–1840.
- [178] F. A. Momany, Determination of partial atomic charges from ab initio molecular electrostatic potentials. Application to formamide, methanol, and formic acid, *The Journal of Physical Chemistry*, 1978, **82**, 592–601.
- [179] C. M. Breneman and K. B. Wiberg, Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis, *Journal of Computational Chemistry*, 1990, **11**, 361–373.
- [180] R. Li *et al.*, A study into the effect of subtle structural details and disorder on the terahertz spectrum of crystalline benzoic acid, *Physical Chemistry Chemical Physics*, 2010, **12**, 5329–5340.
- [181] M. D. King and T. M. Korter, Effect of Waters of Crystallization on Terahertz Spectra: Anhydrous Oxalic Acid and Its Dihydrate, *The Journal of Physical Chemistry A*, 2010, **114**, 7127–7138.
- [182] J. van de Streek, Searching the Cambridge Structural Database for the ‘best’ representative of each unique polymorph, *Acta Crystallographica Section B*, 2006, **62**, 567–579.
- [183] J. van de Streek, All series of multiple solvates (including hydrates) from the Cambridge Structural Database, *CrystEngComm*, 2007, **9**, 350–352.
- [184] M.-A. Perrin *et al.*, Crystal structure determination of the elusive paracetamol Form III, *Chemical Communications*, 2009, 3181–3183.
- [185] A. Parkin *et al.*, Paracetamol monohydrate at 150K, *Acta Crystallographica Section E*, 2002, **58**, o1345–o1347.
- [186] P. A. McGregor *et al.*, Preparation and crystal structure of a trihydrate of paracetamol, *Journal of Pharmaceutical Sciences*, 2002, **91**, 1308–1311.
- [187] T. Beyer *et al.*, The Prediction, Morphology, and Mechanical Properties of the Polymorphs of Paracetamol, *Journal of the American Chemical Society*, 2001, **123**, 5086–5094.
- [188] *Medline Plus: Theophylline*, online: <http://www.nlm.nih.gov/medlineplus/druginfo/meds/a681006.html>, 2010.

- [189] G. Mengozzi *et al.*, Comparative bioavailability of two sustained-release theophylline formulations in the dog, *Pharmacological Research*, 1998, **38**, 481–485.
- [190] N. V. Phadnis and R. Suryanarayanan, Polymorphism in anhydrous theophylline – implications on the dissolution rate of theophylline tablets, *Journal of Pharmaceutical Sciences*, 1997, **86**, 1256–1263.
- [191] M. Eddlestone, private communication.
- [192] L. Seton *et al.*, Solid State Forms of Theophylline: Presenting a New Anhydrous Polymorph, *Crystal Growth & Design*, 2010, **10**, 3879–3886.
- [193] C. Roy *et al.*, Theophylline formulation by supercritical anti-solvents, *International Journal of Pharmaceutics*, 2007, **343**, 79–89.
- [194] Y. Ebisuzaki *et al.*, Methylxanthines. I. Anhydrous Theophylline, *Acta Crystallographica Section C*, 1997, **53**, 777–779.
- [195] C. Sun *et al.*, Theophylline monohydrate, *Acta Crystallographica Section E*, 2002, **58**, o368–o370.
- [196] M. Otsuka *et al.*, Physicochemical Stability of Nitrofurantoin Anhydrate and Monohydrate Under Various Temperature and Humidity Conditions, *Pharmaceutical Research*, 1991, **8**, 1066–1068.
- [197] M. A. Dalby, Antiepileptic and Psychotropic Effect of Carbamazepine (Tegretol) in the Treatment of Psychomotor Epilepsy, *Epilepsia*, 1971, **12**, 325–334.
- [198] M. M. Robertson and M. R. Trimble, Depressive Illness in Patients with Epilepsy: A Review, *Epilepsia*, 1983, **24**, S109–S116.
- [199] T. Okuma and A. Kishimoto, A history of investigation on the mood stabilizing effect of carbamazepine in Japan, *Psychiatry and Clinical Neurosciences*, 1998, **52**, 3–12.
- [200] K. Kipouros *et al.*, Simultaneous quantification of carbamazepine crystal forms in ternary mixtures (I, III, and IV) by diffuse reflectance FTIR spectroscopy (DRIFTS) and multivariate calibration, *Journal of Pharmaceutical Sciences*, 2006, **95**, 2419–2431.
- [201] J.-B. Arlin *et al.*, A strategy for producing predicted polymorphs: catemeric carbamazepine form V, *Chemical Communications*, 2011, **47**, 7074–7076.
- [202] A. J. Cruz Cabeza *et al.*, Solvent inclusion in form II carbamazepine, *Chemical Communications*, 2007, 1600–1602.

- [203] C. C. Wilson *et al.*, A single-crystal neutron diffraction study of the temperature dependence of hydrogen-atom disorder in benzoic acid dimers, *Journal of the Chemical Society, Faraday Transactions*, 1996, **92**, 5051–5057.
- [204] M. Takahashi *et al.*, Interpretation of temperature-dependent low frequency vibrational spectrum of solid-state benzoic acid dimer, *Chemical Physics Letters*, 2009, **479**, 211–217.
- [205] C. C. Wilson, Neutron diffraction of p-hydroxyacetanilide (Paracetamol): libration or disorder of the methyl group at 100 K, *Journal of Molecular Structure*, 1997, **405**, 207–217.
- [206] E. W. Pienaar *et al.*, Polymorphs of nitrofurantoin. I. Preparation and X-ray crystal structures of two monohydrated forms of nitrofurantoin, *Journal of Chemical Crystallography*, 1993, **23**, 739–744.
- [207] C. Lefebvre *et al.*, Polymorphic Transitions of Carbamazepine During Grinding and Compression, *Drug Development and Industrial Pharmacy*, 1986, **12**, 1913–1927.
- [208] W. Liu *et al.*, Solubility of Carbamazepine (Form III) in Different Solvents from (275 to 343) K, *Journal of Chemical & Engineering Data*, 2008, **53**, 2204–2206.
- [209] J. Aaltonen *et al.*, Hyphenated spectroscopy as a polymorph screening tool, *Journal of Pharmaceutical and Biomedical Analysis*, 2007, **44**, 477–483.
- [210] 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*, 2005, **109**, 15531–15541.
- [211] 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*, 2010, **66**, 544–558.
- [212] M. D. King *et al.*, Application of London-type dispersion corrections to the solid-state density functional theory simulation of the terahertz spectra of crystalline pharmaceuticals, *Physical Chemistry Chemical Physics*, 2011, **13**, 4250–4259.
- [213] R. Dovesi *et al.*, CRYSTAL: a computational tool for the ab-initio study of the electronic properties of crystals, *Zeitschrift für Kristallographie*, 2005, **220**, 571–573.
- [214] R. Dovesi *et al.*, *CRYSTAL09 User's Manual*, 2009, University of Torino.

- [215] H. Zhang *et al.*, A dynamics study of the A-chain of ricin by terahertz vibrational calculation and normal modes analysis, *Journal of Molecular Graphics and Modelling*, 2009, **27**, 655–663.
- [216] R. K. Khankari and D. J. Grant, Pharmaceutical hydrates, *Thermochimica Acta*, 1995, **248**, 61–79.
- [217] H. Brittain, in *Polymorphism in pharmaceutical solids*, M. Dekker, 1999, pp. 481–509.
- [218] M. W. Wong and A. G. Mitchell, Physicochemical characterization of a phase change produced during the wet granulation of chlorpromazine hydrochloride and its effects on tableting, *International Journal of Pharmaceutics*, 1992, **88**, 261–273.
- [219] P. York, Solid-state properties of powders in the formulation and processing of solid dosage forms, *International Journal of Pharmaceutics*, 1983, **14**, 1–28.
- [220] H.-B. Liu *et al.*, Characterization of anhydrous and hydrated pharmaceutical materials with THz time-domain spectroscopy, *Journal of Pharmaceutical Sciences*, 2007, **96**, 927–934.
- [221] H.-B. Liu and X.-C. Zhang, Dehydration kinetics of D-glucose monohydrate studied using THz time-domain spectroscopy, *Chemical Physics Letters*, 2006, **429**, 229–233.
- [222] J. K. Gregory *et al.*, The Water Dipole Moment in Water Clusters, *Science*, 1997, **275**, 814–817.
- [223] M. A. Spackman *et al.*, Dipole Moment Enhancement in Molecular Crystals from X-ray Diffraction Data, *ChemPhysChem*, 2007, **8**, 2051–2063.
- [224] M. A. Spackman *et al.*, The use of dipole lattice sums to estimate electric fields and dipole moment enhancement in molecular crystals, *Chemical Physics Letters*, 2007, **443**, 87–91.
- [225] D. Dunmur, The local electric field in anisotropic molecular crystals, *Molecular Physics*, 1972, **23**, 109–115.
- [226] E. Arunan *et al.*, Definition of the hydrogen bond (IUPAC Recommendations 2011), *Pure and Applied Chemistry*, 2011, **83**, 1637–1641.
- [227] E. Arunan *et al.*, Defining the hydrogen bond: An account (IUPAC Technical Report), *Pure and Applied Chemistry*, 2011, **83**, 1619–1636.

- [228] J. J. Dannenberg *et al.*, Are Hydrogen Bonds Covalent or Electrostatic? A Molecular Orbital Comparison of Molecules in Electric Fields and H-Bonding Environments, *The Journal of Physical Chemistry A*, 1999, **103**, 7083–7086.
- [229] S. L. Huth *et al.*, *N*-(4-fluorophenyl)acetamide, 2008, <http://ecrystals.chem.soton.ac.uk/548/>.
- [230] S. L. Huth *et al.*, *N*-(4-methylphenyl)acetamide, 2008, <http://ecrystals.chem.soton.ac.uk/544/>.
- [231] *eCrystals* – University of Southampton, online: <http://ecrystals.chem.soton.ac.uk/>, Accessed on October 2011.
- [232] G. M. Day and S. L. Price, A Nonempirical Anisotropic Atom–Atom Model Potential for Chlorobenzene Crystals, *Journal of the American Chemical Society*, 2003, **125**, 16434–16443.
- [233] P. Naumov *et al.*, Direct Observation of Aminyl Radical Intermediate during Single-Crystal to Single-Crystal Photoinduced Orton Rearrangement, *The Journal of Physical Chemistry B*, 2007, **111**, 10373–10378.
- [234] D. Rapaport, *The Art of Molecular Dynamics Simulation*, Cambridge University Press, 2004.
- [235] P. Balbuena and J. Seminario, *Molecular dynamics: from classical to quantum methods*, Elsevier, 1999.
- [236] A. M. Reilly *et al.*, The Phonon Spectrum of Phase-I Ammonia: Reassignment of Lattice Mode Symmetries from Combined Molecular and Lattice Dynamics Calculations, *The Journal of Physical Chemistry A*, 2008, **112**, 1322–1329.