Far infrared/Tera-Hertz spectroscopy in the gas phase: experiments and theory

Spectroscopie infrarouge en phase gazeuse dans le domaine de l’infrarouge lointain/Tera-Hertz: expériences et théorie

Thèse de doctorat de l'Université Paris-Saclay préparée à l'université d’Evry Val d’Essonne

École doctorale n°571 : Sciences chimiques : Molécules, Matériaux, Instrumentation et Biosystèmes
Spécialité de doctorat: Chimie

Thèse présentée et soutenue à Évry, le 08/12/2017 par

Jérôme Mahé

Composition du Jury:

Mr. David Lauvergnat
Directeur de recherche à l'Université Paris-Sud Orsay, France
Président

Mme. Sandra Luber
Professeure à l'Université de Zurich, Suisse
Rapporteure

Mr. Knut Asmis
Professeur à l'Université de Leipzig, Allemagne
Rapporteur

Mme. Anouk Rijks
Maître de conférences à l’Université Radboud de Nijmegen, Pays-Bas
Examinatrice

Mr. Michel Mons
Directeur de recherche au CEA Saclay, France
Examinateur

Mme. Marie-Pierre Gaigeot
Professeure à l'Université d'Evry Val d’Essonne, France
Directrice de thèse
## Contents

1 Introduction.  
1.1 Infrared spectroscopy for conformational assignment .......................... 10  
1.2 Infrared spectroscopy in the gas phase and alternative strategies for gas phase conformational assignment .................................................. 17  
1.3 Far infrared/Tera-Hertz experimental and theoretical spectroscopy .......... 25  
1.4 Scope of this thesis ............................................................................. 26  

2 Brief description of the systems investigated. ........................................... 29  
2.1 Peptidic systems ............................................................................. 29  
2.1.1 Ac-Phe-AA-NH₂ dipeptide series .................................................. 34  
2.1.2 Ac-Phe-OMe and (Ac-Phe-OMe)₂ .................................................. 37  
2.1.3 Z-(Ala)₆-NH₂ ............................................................................. 40  
2.2 Phenol derivatives .......................................................................... 42  
2.2.1 First series: phenol, catechol, saligenin, salicylic acid and nitrophenol series ...................................................................................... 42  
2.2.2 Second series: saligenin and its water clusters. ............................... 42  
2.2.3 Third series: phenol, catechol, saligenin and phenol-water complex. .. 44  
2.3 Base pairs and analogues ................................................................. 45  

3 IR-UV ion dip spectroscopy. .................................................................. 51  
3.1 Direct absorption spectroscopy, action spectroscopy and alternative methods to probe vibrational levels ...................................................... 53  
3.2 IR-UV ion dip spectroscopy principles .............................................. 54  
3.3 Production of cold and isolated neutral species in the gas phase .......... 57  
3.3.1 Heating source and desorption laser .............................................. 58  
3.3.2 Molecular beam ........................................................................ 59  
3.4 Detection of the ionised systems - Reflectron time-of-flight mass spectrometer .................. 60  
3.5 REMPI (Resonance Enhanced Multi-Photon Ionization ..................... 61  
3.6 Far Infrared light production: the Free Electron Lasers FELIX .......... 65  

4 DFT-MD theoretical method for anharmonic vibrational spectroscopy (and alternatives). ................................................................. 69  
4.1 Electronic representation: DFT (Density Functional Theory) .............. 71  
4.1.1 Basic definitions ...................................................................... 71
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.2 Brief history of the DFT methodology and presentation of the BLYP-D3 density functional, and beyond.</td>
<td>73</td>
</tr>
<tr>
<td>4.1.3 How to build the wavefunction/electronic density in the GPW Gaussian Plane Wave scheme</td>
<td>79</td>
</tr>
<tr>
<td>4.1.4 Kohn-Sham DFT in the CP2K package</td>
<td>84</td>
</tr>
<tr>
<td>4.1.5 SCF procedure: optimisation of the wavefunction</td>
<td>86</td>
</tr>
<tr>
<td>4.1.6 Specific characteristics for hybrid functionals in CP2K</td>
<td>88</td>
</tr>
<tr>
<td>4.2 DFT-MD (Density Functional Theory based Molecular Dynamics)</td>
<td>89</td>
</tr>
<tr>
<td>4.3 Calculation of an infrared spectrum from molecular dynamics</td>
<td>91</td>
</tr>
<tr>
<td>4.4 Geometry optimisation</td>
<td>95</td>
</tr>
<tr>
<td>4.5 Harmonic approximation for infrared spectroscopy - Vibrational Analysis in the Gaussian package</td>
<td>96</td>
</tr>
<tr>
<td>4.6 VPT2 anharmonic spectroscopy in the Gaussian package</td>
<td>97</td>
</tr>
<tr>
<td>4.7 Far infrared spectroscopy of the Ac-Phe-Pro-NH₂ dipeptide.</td>
<td>101</td>
</tr>
<tr>
<td>4.8 Mapping gas phase dipeptide motions in the far-infrared domain.</td>
<td>123</td>
</tr>
<tr>
<td>4.9 Anharmonic, dynamical and functional level effects in far-infrared spectroscopy. Test case on phenol derivatives.</td>
<td>149</td>
</tr>
<tr>
<td>4.10 Fingerprints of Inter- and Intramolecular Hydrogen Bonding in Saligenin-Water Clusters Revealed by Mid- and Far-Infrared Spectroscopy.</td>
<td>173</td>
</tr>
<tr>
<td>4.11 Mapping vibrational modes in the far-infrared/Tera-Hertz domain.</td>
<td>205</td>
</tr>
<tr>
<td>4.11.1 ICDOS: strengths and limitations.</td>
<td>207</td>
</tr>
<tr>
<td>4.11.2 Localised modes</td>
<td>209</td>
</tr>
<tr>
<td>4.11.3 Fingerprint range</td>
<td>211</td>
</tr>
<tr>
<td>4.11.4 Waggings</td>
<td>213</td>
</tr>
<tr>
<td>4.11.5 Delocalised modes</td>
<td>213</td>
</tr>
<tr>
<td>4.11.6 Conclusion</td>
<td>214</td>
</tr>
<tr>
<td>4.11.7 Isotopic substitution in experiments</td>
<td>216</td>
</tr>
<tr>
<td>4.11.8 Mapping local wagging modes</td>
<td>220</td>
</tr>
<tr>
<td>4.11.9 C-H phenyl ring waggings (690-760 cm⁻¹)</td>
<td>220</td>
</tr>
<tr>
<td>4.11.10 N-H waggings (400-900 cm⁻¹)</td>
<td>227</td>
</tr>
<tr>
<td>4.11.11 NH₂ waggings (230-860 cm⁻¹)</td>
<td>234</td>
</tr>
<tr>
<td>4.11.12 OH waggings (200-700 cm⁻¹)</td>
<td>240</td>
</tr>
<tr>
<td>4.11.13 CH₃ rotational modes (&lt;300 cm⁻¹)</td>
<td>242</td>
</tr>
<tr>
<td>4.11.14 Mapping delocalised modes</td>
<td>245</td>
</tr>
<tr>
<td>4.11.15 Peptide backbone and lateral chain motions (&lt;400 cm⁻¹)</td>
<td>245</td>
</tr>
<tr>
<td>4.11.16 Hydrogen bond signatures (&lt;below 250 cm⁻¹)</td>
<td>247</td>
</tr>
<tr>
<td>4.11.17 Alternative analyses</td>
<td>251</td>
</tr>
</tbody>
</table>
### 10 Which method could be "qualified as better" in order to calculate far infrared/THz spectra?

10.1 Anharmonic infrared spectra calculated from DFT-MD/BLYP-D3 .......................... 259
   10.1.1 Ac-Phe-AA-NH$_2$ series .................................................. 259
   10.1.2 Phenol derivatives: Phenol, Catechol, Saligenin, Phenol-Water ......................... 260
   10.1.3 DNA base pairs .............................................................. 262
10.2 Basis set effects ........................................................................ 264
10.3 Harmonic BLYP-D3 infrared spectra .............................................. 266
   10.3.1 Dipeptide series ................................................................. 266
   10.3.2 Ac-Phe-OMe ....................................................................... 268
   10.3.3 Phenol derivatives .............................................................. 269
   10.3.4 DNA base pairs ................................................................. 272
   10.3.5 Summary of what we have learned up to now .................................................. 274
10.4 Anharmonic infrared spectra calculated from VPT2/BLYP-D3 ......................... 276
   10.4.1 Case of Ac-Phe-Ser-NH$_2$ dipeptide and Ac-Phe-OMe peptide model ................ 276
   10.4.2 Case of phenol derivatives .................................................... 280
   10.4.3 Case of DNA base pairs ....................................................... 282
   10.4.4 Summary ............................................................................ 287
10.5 Anharmonic DFT-MD/B3LYP-D3: test case for four phenol derivatives ............... 287
10.6 Conclusions and outlook ................................................................ 290

### 11 A final illustration of the importance of far infrared/tera-hertz spectroscopy: conformational assignment of the (Ac-Phe-OMe)$_2$ dimer.

11.1 Conformational assignment of the monomer: Ac-Phe-OMe ................................. 297
11.2 Conformational assignment of the dimer: (Ac-Phe-OMe)$_2$ $\beta$-sheet .................... 303
11.3 Comparisons of vibrational modes between the monomer and the dimer .......... 306
11.4 Conclusion ............................................................................... 308

### 12 Conclusions and outlook

Bibliography ......................................................................................... 320
Remerciements/Acknowledgements ...................................................... 332
Resumé ............................................................................................... 334
Abstract ............................................................................................ 336
Chapter 1

Introduction.

In 1800, Sir William Herschel, a British astronomer analysed sunlight passing through a prism, and detected "heat rays" beyond the red of the visible rainbow\(^1\). He was studying the "temperature" of different colors by moving a thermometer through light splitted by a prism. Discovering a "rainbow of colors" in the sunlight at shorter frequencies than those visible to the naked eye, he named the infrared region (IR, latin for "beneath red") with far-reaching implications, especially the creation of the infrared spectroscopy.

For most of humanity time, the visible light was the only known part of the electromagnetic spectrum. The electromagnetic spectrum as we know it now is presented in figure 1.1. Each frequency corresponds to a unique wavelength and a unique photon energy.

![Figure 1.1: Schematic presentation of the electromagnetic spectrum. Reproduced from the internet website: "http://www2.lbl.gov/MicroWorlds/ALSTool/EMSpec/EMSpec2.html".](image)

The electromagnetic spectrum extends from the low frequencies (long wavelengths) used for radio communication to gamma radiation at short-wavelengths (high-frequencies), thereby covering wavelengths from thousands of kilometres down to a fraction of the size of an atom. Visible light lies toward the shorter end, with wavelengths from 400 to 700 nm (750-430 THz). The limit for long wavelengths is the size of the universe itself, while it is thought that the short
The infrared range is located between 700 nm and 1 mm (430-0.3 THz, 14300-10 cm\(^{-1}\)) with the far infrared range of interest in this work being between 800 and 10 cm\(^{-1}\) (24-0.3 THz). An interesting characteristic of the infrared light is that the frequencies of lightwaves in this region match the frequencies of characteristic vibrations of molecular bonds. Thus, a given vibrational mode resonantly absorbs infrared light at a specific frequency that induces a variation in the amplitude of vibrations of the system in the corresponding mode. Shining a bright infrared rainbow of light at a molecular sample and collecting the light after its interaction with the sample reveals that the infrared light at some frequencies is diminished in brightness. This means that photons at these frequencies have been absorbed so as to stimulate particular molecular vibrational modes in the sample. This technique is called infrared spectroscopy and for any given sample, the complex patterns of diminished and undiminished intensities of the infrared light as a function of the frequency is called the characteristic infrared spectrum of the material. Thanks to the development of electronics in the early 20th century, the first infrared spectrometer fully automated has been developed in 1937 by Erwin Lehrer\(^2\). The first commercial spectrometer, however, would only be released in the 1960s.

A graph generated by a spectrometer is similar to our fingerprints. As there are no two people with identical fingerprints, there are also no two samples with identical spectra and, as fingerprints are used to identify a person, a spectrum is also used to identify a material. Nowadays the main application of the infrared spectroscopy technique in the industry or in the analytic world is to identify what chemical functions are present in a material and for this aim tables of infrared absorption patterns have been built, especially for the mid infrared spectral range (800-4000 cm\(^{-1}\), 24-120 THz), see table 1.2. One can for example distinguish between a C=O function involved in an amide bond (1690-1630 cm\(^{-1}\) for its typical vibration) from a C=O function in an aldehyde (1740-1690 cm\(^{-1}\) for its typical vibration). Vibrational infrared spectroscopy is widely used on a daily basis in organic or analytical chemistry because it offers huge advantages, i.e. a large application range from small soluble proteins to large membrane proteins, a high time resolution down to 1 \(\mu\)s with moderate effort, often a short measuring time, low amount of sample required (typically 10-100 \(\mu\)g), at relatively low cost (top class spectrometers cost \(\approx 40000\) \(\epsilon\)).

Using more advanced infrared spectroscopic methods, that will be described hereafter we can nowadays go much further than just simply identifying chemical functions in a given material: one can now unravel the three dimensional structure of a given molecular system. Let us go deeper into this point, and especially ask the question of what we can learn from infrared vibrational spectra to make a link to three dimensional structures.
<table>
<thead>
<tr>
<th><strong>Functional Group</strong></th>
<th><strong>Characteristic Absorption(s) (cm⁻¹)</strong></th>
<th><strong>Notes</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alkyl C-H Stretch</td>
<td>2950 - 2850 (m or s)</td>
<td>Alkane C-H bonds are fairly ubiquitous and therefore usually less useful in determining structure.</td>
</tr>
<tr>
<td>Alkenyl C-H Stretch</td>
<td>3100 - 3010 (m)</td>
<td>Absorption peaks above 3000 cm⁻¹ are frequently diagnostic of unsaturation</td>
</tr>
<tr>
<td></td>
<td>1680 - 1620 (v)</td>
<td></td>
</tr>
<tr>
<td>Alkynyl C-H Stretch</td>
<td>~3300 (s)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2260 - 2100 (v)</td>
<td></td>
</tr>
<tr>
<td>Aromatic C-H Stretch</td>
<td>~3030 (v)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>860 - 680 (s)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1700 - 1500 (m,m)</td>
<td></td>
</tr>
<tr>
<td>Alcohol/Phenol O-H Stretch</td>
<td>3550 - 3200 (broad, s)</td>
<td></td>
</tr>
<tr>
<td>Carboxylic Acid O-H Stretch</td>
<td>3000 - 2500 (broad, v)</td>
<td></td>
</tr>
<tr>
<td>Amine N-H Stretch</td>
<td>3500 - 3300 (m)</td>
<td>Primary amines produce two N-H stretch absorptions, secondary amides only one, and tertiary none.</td>
</tr>
<tr>
<td>Nitrile C≡N Stretch</td>
<td>2260 - 2220 (m)</td>
<td></td>
</tr>
<tr>
<td>Aldehyde C=O Stretch</td>
<td>1740 - 1690 (s)</td>
<td>The carbonyl stretching absorption is one of the strongest IR absorptions, and is very useful in structure determination as one can determine both the number of carbonyl groups (assuming peaks do not overlap) but also an estimation of which types.</td>
</tr>
<tr>
<td>Ketone C=O Stretch</td>
<td>1750 - 1680 (s)</td>
<td></td>
</tr>
<tr>
<td>Ester C=O Stretch</td>
<td>1750 - 1735 (s)</td>
<td></td>
</tr>
<tr>
<td>Carboxylic Acid C=O Stretch</td>
<td>1780 - 1710 (s)</td>
<td></td>
</tr>
<tr>
<td>Amide C=O Stretch</td>
<td>1690 - 1630 (s)</td>
<td></td>
</tr>
<tr>
<td>Amide N-H Stretch</td>
<td>3700 - 3500 (m)</td>
<td>As with amines, an amide produces zero to two N-H absorptions depending on its type.</td>
</tr>
</tbody>
</table>

Figure 1.2: Table of mid-infrared absorptions. Reproduced from the internet website: "https://webspectra.chem.ucla.edu/irtable.html"
1.1 Infrared spectroscopy for conformational assignment

Infrared (IR) vibrational spectroscopy is nowadays one of the main analytical tools widely used to unravel structures and isomers, for molecules and clusters in the gas phase, for liquids and (bio)-molecules immersed in liquids, for solids, solid/liquid and liquid/air interfaces. Infrared spectroscopy is a generic term that may directly refer to transmission spectroscopy but various infrared spectroscopy techniques are applied in the condensed phase, let us for instance cite ATR spectroscopy\(^3\),\(^4\) (Infrared Attenuated Total Reflection Spectroscopy) nowadays the preferred choice to study proteins by vibrational spectroscopy, IRRAS\(^5\) (Infrared Reflection Absorption Spectroscopy), AFM-IR\(^6\) (Atomic Force Microscopy - InfraRed). One can also cite the SFG\(^7\) (Sum Frequency Generation), SHG\(^8\) (Second Harmonic Generation) and SEIRAS\(^9\) (Surface-Enhanced InfraRed Absorption Spectroscopy) spectroscopies that can be applied at surfaces and interfaces. These methods are not suited for the gas phase which is the phase of interest in this thesis. Gas phase methods, mainly action spectroscopy, i.e. IR-UV ion dip, IR-PD, IR-MPD, will be described in section 1.2.

It is now well established that there is a direct relationship between the frequency of the vibrational modes and the three-dimensional structure of the molecular system. In other words, an infrared spectrum generated by a spectrometer is actually similar to our (human) fingerprints. As there are no two people with identical fingerprints, there are also no two samples with identical spectra and, as fingerprints are used to identify a person, a spectrum is also used to identify a material. The paradigm is that it is possible to associate each three dimensional structure (or conformation) that can be adopted by a given material/molecular system to a unique vibrational spectrum. In other words, it is possible to extract information about the conformation of the system from the infrared spectrum alone. In the following I will use both the term vibrational spectroscopy and infrared spectroscopy as equivalent. This is not correct of course. Infrared spectroscopy is a specific vibrational spectroscopy in which the dipole moment of the molecular system interacts with electromagnetic radiations, thus giving rise to specific vibrational selection rules. Only certain vibrations will be active in infrared spectroscopy.

The determination of three dimensional structures of molecules, assemblies of molecules or materials in general is crucial in several domains. Let me cite only two examples here:

- In biology, the function of a biological molecule is known to be intimately connected to its three-dimensional structure. An enormous amount of effort is therefore put to determine the latter with the hope of understanding, and perhaps intervening into the former. This will be illustrated with the examples of proteins and DNA below.

- In industry, catalysis at surfaces is at the heart of many processes. Characterising the structure and the chemistry of the surface, as well as how the catalyst interacts with it, is crucial for developing efficient catalysts. One can cite as example the oxide alumina surface in contact with liquid water which associated to other molecular catalysts serve for the cracking of petrol in the petroleum industry.

Knowing the structure is therefore the first step into understanding the material and possibly designing modifications for improving its function. This is one main goal in analytical
chemistry nowadays.

Basically all polar bonds contribute to the infrared absorption. This is at the same time the crux and the strength of infrared spectroscopy. A strength, because nearly all molecular systems absorb infrared radiations. A crux, because the spectrum of large molecules (like proteins for example) is thus composed of many overlapping bands with the consequence that much structural information is hidden under broad, featureless absorption bands.

This is called spectral congestion, happening when many vibrational modes absorb simultaneously at frequencies that are close to each other. Spectral congestion can also occur because of many conformations of the same system co-existing in the experimental conditions, because of temperature inducing conformational changes and hot bands, also because of temperature inducing broadening of the vibrational bands. In the condensed phase, water surrounding a biomolecule for instance or water in contact with a surface is also a source of broadening the vibrational spectra and therefore a source of spectral congestion.

In these conditions, spectral congestion can make it hard to make a unique link between the vibrational spectrum and the structure of the system. All the structural information are in the vibrational spectrum but it is difficult to extract them. In order to get that information, one can propose two paths.

One path is to remove part of the complexity of the molecular system by removing for instance the surrounding solvent. The molecular system of interest is hence analysed in the gas phase. This is of course not removing the spectral congestion arising from many oscillators absorbing photons at very close frequencies within the gas phase system, which remains crucial for large molecules like proteins for instance. To go beyond that limitation, one can reduce the size of the system investigated and hope that hence the complexity in analysing the spectral signatures will be reduced. In this reductionist approach one has to make sure that the (reduced) "building blocks" analysed are still maintaining the main relevant structural architectures that are present in the initial large molecule. In this way, the structural characterization of the building blocks by vibrational spectroscopy gives us an understanding of the structure of the larger (targetted) molecule. It also gives us spectroscopic fingerprints to further use as recognition patterns in the spectra of the larger molecule.

The second path into getting an understanding of the vibrational signatures and make a direct link with 3D structures is to use theory and simulations to calculate infrared spectra of the system of interest for different potential structure candidates, and then search for a match between experiment and calculation. It might not be possible to do this calculation directly on the large and complex system of interest, and again one might have to go for building blocks instead.

These paths are taken in my thesis work. We are interested in characterizing the structure of biomolecules, proteins and DNA in our case, in order to possibly provide some understanding into their functions. To that end, building blocks of these complex and large objects are going to be spectroscopically characterized in the gas phase. IR spectroscopic experiments are going to be conducted. Theoretical calculations of the infrared spectra of these building blocks
are going to be calculated, analysed, and these analyses are going to help us unravel fingerprint spectroscopic signatures, recognize these fingerprints, understand experimental signatures for the building blocks and for the larger targetted (bio)molecule, and make the link to three-dimensional structures.

For these experiments and calculations to be useful and relevant, one has to carefully choose the building blocks investigated so that the spectroscopic-structure relationships hence gained can be useful for characterizing (without ambiguity) the 3D structures of the large proteins and DNA assemblies in their natural environment. Before discussing the building blocks investigated in my work (see also chapter 2 for more details), let me present some fundamental data on proteins and DNA. It will be then easier to explain the choices made in my work. Protein and DNA structures are illustrated in figures 1.3-1.4.

Proteins play an important role in biology\textsuperscript{10–12}. A protein is a polymer made of amino acids and only 21 amino acids are being used in the synthesis of proteins in nature. Proteins will be discussed in more details in chapter 2. One can cite as example of proteins the well known haemoglobin (present in our blood, responsible for the transportation of oxygen and carbon dioxide through our bodies). Our goal through vibrational spectroscopy is to unravel secondary and tertiary structures of proteins, see figure 1.3. Proteins are now known to be structurally organised into well-defined $\alpha$-helices or $\beta$-sheets (figure 1.4), either purely into these architectures or as a mixing of these. Some proteins are also not so well organised and are known as random coils. The general view is that $\alpha$-helices and $\beta$-sheets structures are built on “well designed/well known” intermolecular hydrogen bonds between amide N-H and carbonyl C=O groups that belong to the protein backbone scaffold, see top fig. 1.4. From a long distance, it thus seems that everything is known on these structures. However, a more careful analysis of the proteins structures recorded in the PDB (\textit{Protein Data Bank}) shows that these structures are more subtle, there are especially some subtle motifs based on specific hydrogen bonds hence forming local nests of H-Bonds that actively participate to the final three dimensional structures. They presumably are an important part of the final protein function. These motifs\textsuperscript{13} are typically called C5 found in $\beta$ strands, C7 found in $\gamma$-turns, C10 in $\beta$-turns and 3\textsubscript{10} helices, C13 in $\alpha$-helices and $\beta$-sheets, with Cn meaning a hydrogen bond ring/nest of size n (n stands for the number of atoms present in the ring formed by the whole hydrogen bond network). Larger H-Bonded rings also exist, i.e. C17, C20. It is pivotal to recognise and characterise these motifs within the whole protein scaffold, and this is where vibrational spectroscopy might be very useful.

In DNA (\textit{DeoxyriboNucleic Acid}), especially in the well known double helix seen at the bottom of figure 1.4, strands made of sugars and phosphate groups are linked through base pairs being hydrogen bonded. There are four DNA bases, cytosine, thymine, guanine and adenine, see figure 1.4. The hydrogen bond patterns between the bases are of importance to characterise and recognise spectroscopically as they are not only the main elements for the double strand scaffold but they are also possibly involved in accommodating metals and organic molecules within the whole architecture, among other functions for the replication of the genome. This can possibly be a source of disturbance of the whole DNA scaffold, altering the function of the

One can see from these short examples that the recognition and characterisation of specific motifs within the whole protein and DNA scaffolds can be of importance, not only to understand the whole 3D structure adopted by these molecules but also to understand the structure of subtle local hydrogen bond nests and networks that might be of relevance for the function of the biomolecule and might be the location for specific interactions with external drugs. Any
CHAPTER 1. INTRODUCTION.

Figure 1.4: **Top:** The $\alpha$-helix and $\beta$-pleated sheet form because of hydrogen bonding between carbonyl and amino groups in the peptide backbone. Certain amino acids have a propensity to form an $\alpha$-helix, while others have a propensity to form a $\beta$-pleated sheet. Reproduced from the internet website: "https://www.boundless.com/biology/textbooks/boundless-biology-textbook/biological-macromolecules-3/proteins-56/protein-structure-304-11437/". **Bottom:** Schematic representation of the DNA double helix (Reproduced from the internet website: "https://www.quora.com/What-makes-up-the-rungs-of-a-DNA-ladder").
structural disturbance on these motifs might also be at the origin of specific diseases. Spectroscopic characterization is certainly one way to get such information on these structures, as will hopefully be demonstrated in my thesis work.

In the next pages, I would like to illustrate what is done to unravel three dimensional structures of proteins in the condensed phase, especially in liquid water. I would especially like to illustrate the importance of the spectroscopy of building blocks. Hereafter I will concentrate only on proteins, but my arguments are easily transferable to DNA.

In the liquid phase, characterization of three dimensional structures of proteins by infrared spectroscopy is done mostly through analyses of the Amide I mode/band ($\approx 1600$-$1800$ cm$^{-1}$). The amide I vibration, absorbing near 1650 cm$^{-1}$, arises mainly from the C=O stretching vibrations. The absolute position and shape of the associated band has been demonstrated to give information on secondary structures of proteins, unraveling a $\alpha$-helix, a $\beta$-sheet structure, a $\beta$-turn, or more irregular structures. The most common approach to determine the secondary structure is the decomposition of the amide I massif into identified contributions of biological motifs for which the frequency is somehow well identified, using deconvolution as presented in figure 1.5.

![Figure 1.5](image)

Figure 1.5: Fourier transform infrared (FTIR) spectrum of a typical protein illustrating the amide I and amide II bands at approximately 1650 cm$^{-1}$ and approximately 1540 cm$^{-1}$ respectively. The inset depicts an expanded view of the amide I band, which can be deconvoluted into its secondary structure components. (Reproduced from the paper "FTIR spectroscopic imaging of protein aggregation in living cells" by Miller et al.).

From such deconvolution, one can obtain the composition of the protein in terms of percentage of these generic motifs ($\alpha$-helices, $\beta$-sheets, ...). While this is already an important information/knowledge, the location of these secondary motifs in the whole protein structure/scaffold is however not straightforward to obtain, if not impossible, using the infrared spectroscopy alone (at least with the deconvolution presented here). Any deviation (small or large) from these general motifs is impossible to quantify from such analysis. Therefore the
unravelling of details within the conformation can not be expected from such simple analysis of the infrared signatures, and especially the subtle hydrogen bond patterns/nests/networks discussed above are not directly accessible from such simple deconvolution. Of course, one could maybe get more detailed knowledge into the structure for instance by isotopic substitution, but this requires some efforts. Note also that such method may induce unwanted modifications of the 3 dimensional structure.

One can see from the discussion above that 3D structure recognition of proteins, here illustrated in liquid water, can be powerful and extremely limited at the same time, using IR spectroscopy fingerprints. To go beyond the initial knowledge of a $\alpha$-helix, $\beta$-sheet, random coil organisation and/or a rough % of each of these into the whole protein scaffold and get instead a more detailed knowledge about the structure like the localisation of these generic motifs into the whole structure, possible interactions between $\alpha$ and $\beta$ parts of the protein, connection between these parts for instance in the form of turns, subtle local hydrogen bond networks that might be at the origin of a local change in the whole structure, etc, requires more advanced IR spectroscopy and more advances analysis features.

This is where IR spectroscopy of individual building block motifs enters into play, and as already said before, removing the solvent from this strategy is an excellent idea in order to reduce the complexity of the whole analyses. If we can characterise the spectral signatures of the essential building blocks of proteins (and DNA) by gas phase infrared spectroscopy, we are in a position to provide new information to be used into the whole deconvolution strategy of the spectra of the complex proteins.

This is the strategy followed in my thesis work, experimentally and theoretically. As will be seen throughout my work, theory is an essential part of the strategy.

Let me finish this part by a bit of digression before discussing in details the gas phase infrared spectroscopy. Infrared spectroscopy is a wonderful tool for unravelling 3D structures, but of course other techniques exist with the same aim. To be honest, it is the combination of results arising from all these methods that provide a more complete and more definitive view on biomolecules structures. Let me briefly discuss some of these methods.

- X-ray crystallography has been a central methodology for unravelling the three dimensional structure of proteins and DNA allowing one to obtain structures at the atomic resolution. The main limitation of this methodology is the requirement of crystallised samples, which means that the protein is analysed without its natural bioenvironment and may also crystallise into a different form (than the one in the cell) due to the change of its environment. Note also that this method is blind to flexible parts of the protein where conformational transitions occur, and that there are plenty of examples of proteins difficult or impossible to crystallise. A large amount of biological components remain therefore inaccessible for X-ray analyses.

- Another technique, the cryo-EM (Cryo-Electron-Microscopy), which got the 2017 Nobel prize in chemistry, is expanding in structural biology. It promises better resolution than the X-ray crystallography with no need of crystallised samples, which allows the investigation of all proteins that are impossible to crystallise with the current technologies.
1.1. INFRARED SPECTROSCOPY IN THE GAS PHASE

The samples only need to be frozen. This is of course not directly comparable to the in vivo conditions but the whole method is a huge step forward in characterising 3D structures of biomolecules. This technology is currently overcoming X-ray crystallography and becomes the dominant methodology.

- Among the most important methods, one can also cite the NMR method\(^\text{17}\) (Nuclear Magnetic Resonance) allowing one to investigate liquid solutions containing the biomolecules of interest. This can mimic somehow part of the biological environment. NMR offers similar advantages than infrared spectroscopy (like studying conformational changes). But NMR also suffers from similar limitations as the infrared spectroscopy with a loss of information for large molecules. This can however be improved by isotopic substitution (e.g. one can choose the location of \(^{13}\text{C}\) for C-13 NMR).

- The UV circular dichroism technique\(^\text{18}\) provides similar information as the infrared spectroscopy, unraveling the secondary structures of proteins, but the exact three dimensional conformation is still not accessible (i.e. the locations of the identified secondary structures in the whole protein remain unknown).

1.2 Infrared spectroscopy in the gas phase and alternative strategies for gas phase conformational assignment

As discussed above, one way to simplify vibrational spectra is to investigate molecules in gas phase conditions. Among the methods described above for the condensed phase only the vibrational spectroscopy can be transposed into the gas phase, but the infrared methods working for condensed phase (i.e. ATR\(^\text{3,4}\), IRRAS\(^\text{5}\), AFM-IR\(^\text{6}\), ...) are not adapted anymore. We will present in the following, infrared spectroscopic experiments for the gas phase as well as other methodologies able to provide conformational assignments.

Molecular spectroscopists began to remove biological molecules from their natural environment in the mid-1980s, in order to interrogate them in their isolated state. Early infrared spectroscopic experiments have been mostly developed by Levi’s group\(^\text{19–25}\). Fair enough, the chemical and biological environments are missing but we gain in the comprehension of the vibration modes as well as the relationship vibrational modes ↔ structures. As already said gas phase is also expected to decrease the complexity of vibrational features mainly due to the smaller size of the systems investigated (less chemical functions interrogated).

Gas phase spectroscopy is mostly done by vibrational action spectroscopy that can be roughly divided into experiments performed at low (and extremely low) temperatures and experiments performed at finite temperature. A further division concerns neutral and charged molecules (and clusters) being investigated, in relation to the initial stage of production of the molecules. In our case we study neutral species at low temperature using a desorption laser to produce molecules in the gas phase and a molecular beam to cool them down. These two technologies are presented in section 3.3 in chapter 3. Molecular beams produce cold molecules
in high vacuum. Because of low temperature, the complexity of the spectrum will be mechan-
ically decreased, showing thinner bands and far less bands overlaps because of the reduction
of the amount of conformations, disappearance of hot bands as well as the diminution of the
rotational broadening will also be seen. The current paradigm is that molecular beams produce
the lowest free energy conformer.

We present now experiments for gas phase spectroscopy. In these experiments, the density
of absorbing species is always extremely small compared with the one encountered in liquids
or solids (this is even more true in molecular beam conditions). Except in few cases, direct
monitoring of absorption is therefore usually impossible in the gas phase.

If one wants to stick to direct absorption spectroscopy, one can increase the path length of
the infrared beamlight. This solution is offered for example by the FTIR spectroscopy (Fourier
Transform InfraRed). Fourier-transform spectroscopy records infrared spectra over a wide spec-
tral range without the need of broadly tunable sources thanks to the use of a Michelson interferometer that selects a chosen wavelength. This solution is used for instance by Jean-Marie
Flaud or Pascale Roy in France, with a path length over 100 m. But even with pressure
conditions that correspond to not advanced vacuum and using an intense source of light such
as the beamline AILES at the SOLEIL synchrotron (Saclay, France), such experimental setup
can only address molecular systems of small size (typically lower than 20 atoms).

Cavity ring-down spectroscopy is a methodology sharing the same spirit than the FTIR
spectroscopy above. A laser beam is injected into an optical resonator formed by a pair of high
reflectivity (≃99%) mirrors and bounces back and fourth. Only a small fraction is transmitted
through the exit mirror. It provides a multi-path scheme equivalent to an absorption-cell of
several hundred metres or more and allows very sensitive detection.

The most common approach to measure gas phase infrared spectra is to use action spec-
troscopy, where the consequence of the infrared photon(s) absorption by the molecular system
is measured. Usually, a mass spectrometer (much more sensible than the current infrared
detector) is used to probe the consequences of the photon absorption.

We present briefly below gas phase experiments based on action spectroscopy, i.e. InfraRed
- Photo Dissociation (IR-PD), InfraRed Multi Photon Dissociation (IRMPD) and IR-UV ion dip
spectroscopy.

- InfraRed - Photo Dissociation (IR-PD) monitors the infrared-induced fragmentation of a
molecular system as a function of the infrared radiation frequency. In order to record the
infrared spectrum of a molecular system A, it is possible to weakly attach a messenger m (e.g.
a rare gas atom or a molecule) and then probe the weakly bound A···m system. When this
system absorbs a photon with energy $h\nu$ corresponding to the excitation of one intramolecular
vibration in the A system, the rapid redistribution of energy among all vibrational degrees of
freedom is followed by the breaking of the weakest bond. If the A···m binding energy is less
than the infrared photon energy $h\nu$ (as it is in practice), infrared absorption is monitored by
observing the decrease of the A···m mass spectrometer signal corresponding to the destruction
of the A···m complex producing the fragments A and m. The presence of the messenger m
should, in principle, be as less perturbative as possible on the A system. A rare gas atom
is generally chosen since it has a sufficient polarisability to attach to the A system and its presence only induces rather small spectral shifts and does not change conformations (at least changes in the conformation are not expected). Argon is an atom widely employed in IR-PD experiments.

IR-PD has been developed and used by the groups of Profs Lisy\textsuperscript{28,29} (USA), Johnson\textsuperscript{30} (USA), Asmis\textsuperscript{31} (Germany), Duncan\textsuperscript{32,33} (USA), to cite a few.

• InfraRed - Multi Photon Dissociation (IR-MPD) uses multiple absorption of photons for the vibrational activation of polyatomic systems above fragmentation thresholds.\textsuperscript{34} IVR (Intramolecular Vibrational energy Redistribution) is responsible for the breaking of the weakest bonds in molecular systems whichever intramolecular vibrational frequencies are initially excited. Those molecular systems possess a set of vibrational modes with frequencies $\nu_n$, each mode being characterised by a quantum number $n_i$. Although the molecular systems may be at temperature of several hundred degrees Kelvin (when no molecular beam is used), only the fundamental level $n_i=0$ of each mode has a significant population. In IR-MPD, molecular systems are confined in a small spatial region overlapping with an infrared beam. This allows long interaction times up to seconds and thus the possibility of a large number of sequential infrared photon absorptions. Absorption takes place resonantly whenever the frequency of the infrared beam illuminating the molecules coincides with the frequency of a vibrational transition starting from the fundamental level $n_i=0$ towards the level $n_i=1$. Following the initial absorption step, the systems have acquired an internal energy $E_i=h\nu_i$. Inbetween two infrared excitation pulses, IVR takes place and the energy $E_i$ is redistributed over all vibrational degrees of freedom. It is usually believed that the IVR is complete, so that the absorbing vibrational mode goes back to its fundamental level, before the subsequent photon absorption. A second absorption at the initial pump frequency $n_i=0 \rightarrow n_i=1$ can thus take place and the system internal energy again increases by a second infrared photon energy $E_i=h\nu_i$. This process is repeated until the acquired internal energy exceeds a certain fragmentation energy threshold. Monitoring the fragmentation provides the signature of the ion absorption and therefore provides a vibrational infrared spectrum. Some experimental bands can differ from those measured with direct absorption spectroscopy in terms of frequencies and intensities and this can be attributed to two different reasons. The IVR process taking place between two photon absorptions is more or less efficient whether a "local" mode or a mode involving motions spread over the molecular system is initially excited. It can also turn out that following the first photon absorption, a chemical transformation such as a tautomerization, can occur.

This method has been developed and used by groups of Profs Maitre\textsuperscript{34,35} (France), Oomens\textsuperscript{34,36} (The Netherlands), Von Helden\textsuperscript{37} (Germany), Fielicke\textsuperscript{38} (Germany), Compagnon\textsuperscript{39} (France), to cite a few.

• In the IR UV ion dip spectroscopy, a UV laser will now be used at the $S_0 \rightarrow S_1$ electronic transition of the selected molecular conformation. A chromophore (typically an aromatic ring) is therefore needed for an efficient absorption of UV photons. Ionisation of the molecule via a two photon scheme will be performed, and the ion current will be measured with a mass spectrometer. The ion current (signal coming from the mass spectrometer) is constant if no other process is applied. In the IR-UV scheme an infrared source of light irradiates the gas phase
population before the UV irradiation is done. At this stage, no conformer selection (by UV) has been done. The infrared light will be scanned over the whole range where we want to plot the action spectrum. Whenever the infrared frequency will make the infrared photon resonant with a vibrational transition of the molecular system investigated, a vibrational excitation will occur due to the absorption of the infrared photon. Once the conformation is vibrationally excited, it looses its resonance with the pre-chosen UV photons, which stops the two photon scheme for ionisation. A depletion in the measured ion current is consequently observed. The ion current plotted as a function of the infrared frequency is the action spectrum (in IR-UV ion dip spectroscopy), the depletions (dips) correspond to the absorption of the IR photons.

IR-UV ion dip spectroscopy has been developed and used by groups of Profs Rizzo (Switzerland), Mons (France), Rijs (The Netherlands), DeVries (USA), Gerhards (Germany), to name a few. In collaboration with Dr A.M. Rijs, this is the spectroscopy applied in our work. We will see later that we employ this spectroscopy in a new vibrational domain, the far infrared/Tera-Hertz one, that Dr. Anouk Rijs has contributed to develop in the gas phase community. This method is fully described in chapter 3.

Note that other methods exist allowing conformational assignment based on the study of vibrational modes but not based on the infrared spectroscopy:

- **Raman spectroscopy**, but this method is challenging in the gas phase and therefore limited to small size systems. Note that the rules of selection are different from infrared and therefore non active modes in infrared spectroscopy could be probed here and vice versa.

- **LIF (Laser Induced Fluorescence)** can probe vibrational modes and therefore provide conformational assignment by monitoring the fluorescence of a system as a function of the excitation wavelength.

Further methodologies allowing conformational assignment in the gas phase exists, either based on rotational spectroscopy or based on other principles than spectroscopy:

- **Microwave spectroscopy** has since a long time been a powerful method for precise determination of gas phase structures of molecules in the gas phase. An absorption frequency measurement provides an experimental value of I, the moment of inertia of the molecular system characterised. This value can then be compared with possible predicted values. Corresponding to different molecular geometries obtained from structure calculations, the match between calculation and experiment provides the right structure of the system. Unfortunately this method is limited in terms of system size that can be investigated, ≃20 atoms can be done routinely nowadays. Note that systems of this size can be conformationally assigned without too much problems using infrared spectroscopy. It is therefore not so clear whether microwave spectroscopy is so relevant to unravel three dimensional structures of the larger systems we want to investigate. Only recently (2013), the simplest capped amino acids have been documented using microwave spectroscopic techniques.

- **Collisional induced dissociation (CID) using tandem mass spectrometry**. The principle is that a mass spectrum of the fragments obtained after collision will be dependent on the
1.2. INFRARED SPECTROSCOPY IN THE GAS PHASE

initial conformation of the system (before collision). It is therefore possible to extract information about the conformation of the system from the mass spectrum obtained by CID. This is typically done by coupling experimental CID with molecular dynamics simulating the collision.

Nowadays, action spectroscopies are probably the best tools available for conformational assignment of gas phase entities (molecules, clusters) and they will be used in the context of this thesis. The gas phase infrared action spectra presented in this thesis are measured using IR-UV ion dip spectroscopy. This method is presented in details in chapter 3.

It is one thing to be able to measure experimental spectra of gas phase molecules, it is another issue to be able to interpret the spectral signatures in terms of vibrational motions and relate the spectra to specific 3D conformations. This is where theoretical calculations of vibrational spectra are essential. As discussed above for condensed phase, infrared spectroscopy is a method of choice to identify chemical functions and general structural organisations but for smaller systems isolated in the gas phase, one can go further and unravel the “exact” three dimensional structure of the molecular system. In practice, gas phase experimental spectra are compared with calculated spectra for different possible conformations and the theoretical spectrum that provides the match with the spectrum measured experimentally corresponds to the structure present in the experimental conditions. In this context, we need reliable theoretical spectroscopic methods.

While only force field or coarse-grain models can deal with entire proteins in condensed phase, one can easily use an electronic representation as DFT (Density Functional Theory) for systems up to hundreds of atoms (when its environment has been removed). This is the representation adopted in our work. In our work and as described in chapter 4, DFT is coupled with MD, i.e. we perform DFT-MD simulations. This is an advanced methodology that takes into account anharmonicities of the potential energy surfaces, of the dipole surfaces and that takes into account modes couplings into the infrared spectrum calculation (by opposition to the harmonic method for which the description of the potential and of the dipole is restricted to a harmonic function). Alternative methods to calculate vibrational spectra, i.e. the harmonic approximation and VPT2 anharmonic spectroscopy are also presented in chapter 4.

Very diverse systems have been investigated in the gas phase for conformational characterisation in the last two decades, with a special emphasis given on biomolecular models. A good view is provided in two recent reviews from Baldauf et al. (2015) and from Rizzo et al. (2009) as well as in the very recent book from Anouk Rijs and Jos Oomens (2015): "Gas-Phase IR Spectroscopy and Structure of Biological Molecules" or in the slightly older book from Jean-Pierre Schermann: "Spectroscopy and Modeling of Biomolecular Building Blocks". These gas phase investigations cover a broad range of systems of interest, i.e. peptides, saccharides, water clusters, DNA bases, Metal-Water clusters, metal clusters. All these works, as my work here, are based on the reductionist approach already discussed, i.e. which consists in characterizing the structures of smaller motifs/building blocks in the gas phase.

In the context of this thesis, we worked on gas phase building blocks of proteins and
DNA, especially peptides (building blocks of proteins) and base pairs (building blocks of DNA intermolecular-strands). We also worked on Phenol derivatives molecules and their complexes with water molecules, as these molecules can also be considered as model systems for DNA bases. Chapter 2 gathers background information about the systems investigated in my thesis. As will become more clear in this latter chapter, we will spectroscopically characterise these building blocks and their specific motifs of interest for the larger protein and DNA structures. All the experimental measurements have been done by Dr. Sander Jæqx, Daniël Bakker, Dr. Faady Siouri and myself, in collaboration with the Rijs and de Vries’s groups (groups from the FELIX laboratory, Radbout University, Nijmegen, The Netherlands, and from the Department of Chemistry & Biochemistry, University of California, Santa Barbara, USA).

Let us provide below some introductions to our experiments for peptides and DNA base pairs, while we will let the reader refer to the references cited above for other systems. In the discussions below I will return to the ideas of motifs/building blocks and spectral congestion, and show that a new spectral domain has to be investigated. This is where our experiments are new to the community.

- Gas phase peptides of various sizes have been investigated in the 1000-4000 cm\(^{-1}\) spectral range within the past two decades. One of the first goals of spectroscopic studies of isolated neutral peptides was to check whether gas phase systems were spontaneously capable of forming the secondary structures typically observed in the condensed phase\(^{13}\), i.e. using sole intramolecular interactions as driving forces. All features that can be seen for condensed phase proteins have been indeed observed in isolated neutral peptides in the 2000s, i.e. C5 hydrogen bond networks in \(\beta\) strands\(^{61-65}\), C7 in \(\gamma\)-turns\(^{61-65}\), C10 in \(\beta\)-turns\(^{61-65}\) and \(3_{10}\) helices\(^{66,67}\), C13 in \(\alpha\)-helices\(^{68}\) and \(\beta\)-sheets\(^{69-74}\) (n stands for the number of atoms present in the ring formed by the whole hydrogen bond network/nest).

- While nucleobases (adenine, thymine, cytosine, guanine) have been extensively investigated in the gas phase in the 2000s\(^{75}\), the amount of publications about base pairs in the gas phase is much more limited\(^{76-81}\). Work on hydrated base pairs can also be found in literature\(^{82-85}\) but at my best knowledge no gas phase studies of base pairs linked to the deoxyribose (sugar) moiety have been reported yet.

Among these studies, numerous have found that even for gas phase systems considered as rather smallish, spectral congestion may arise, thus loosing the conformational details hidden underneath the IR features. One can cite the example of the DSIP\(^{86}\) (\textit{Delta sleep-inducing peptide}, \(\approx 110\) atoms). Figure 1.6 presents the experimental spectrum of the DSIP peptide measured by Bakker et al.\(^{86}\) as well as two helical and stretched optimised structures and their associated theoretical harmonic vibrational spectra. The infrared spectrum recorded in the range 600-1800 cm\(^{-1}\) is clearly not conformational selective enough to distinguish between these two structures. If we focus on the two theoretical spectra, we observe similar signatures in the range 1400-2200 cm\(^{-1}\) which means that this range can not be used to distinguish between the two structures. Still focusing on the two theoretical spectra, we start to observe differences below 1400 cm\(^{-1}\) (in the fingerprint region) but unfortunately the experimental spectrum is not resolved enough to allow one to go further.
As seen above, it appears challenging to assign a conformation for a system of the size of the DSIP system (∼110 atoms). Other examples in the literature show the difficulty to unambiguously assign a structure for even smaller molecular systems when there are subtle differences between the lowest free energy stable geometries. We will illustrate this with the example of the tryptophan-glycin-glycin system\textsuperscript{86,87} (∼60 atoms). Two tryptophan-glycin-glycin conformations co-exist in the gas phase and their conformational assignment is presented in figure 1.7. The ‘a’ measured spectrum can easily be assigned to the A conformation (calculated structure and spectrum) because we observe a good match between the experimental spectrum labelled ‘a’ and the calculated spectrum of the calculated structure A. The study can not conclude on the assignment for the spectrum labelled b: one can not distinguish between structures D and G which infrared spectra both fit the experimental spectrum. Nothing in the experimental spectrum presented could unfortunately allow a definitive conclusion on whether structure D or G is present in the experimental conditions.
CHAPTER 1. INTRODUCTION.

Figure 1.7: On the left panel, calculated structures of the Trp-Gly-Gly tripeptide at the B3LYP/6-31G(d,p) level of theory. Relative electronic energies (in cm$^{-1}$) are given in parentheses. On the right panel, gas phase experimental infrared spectra (top two traces) and calculated spectra of Trp-Gly-Gly in the spectral region 1000-2000 cm$^{-1}$. The harmonic frequencies are calculated at the B3LYP/6-31G(d,p) level of theory and are scaled by 0.964. The assigned structures are indicated by the appropriate conformer labels. This figure has been copied from the article: "Folding structures of isolated peptides as revealed by gas-phase mid-infrared spectroscopy".

To go beyond spectral congestion observed in the 1000-4000 cm$^{-1}$ spectral range that sometimes prevent the conformational assignment as in the studies illustrated above, vibrational action spectroscopy can be coupled with ion mobility or isotopic substitution, that hence provide extra information.

We provide in this thesis work an alternative strategy by probing the far infrared spectral domain (10-800 cm$^{-1}$ or 24-0.3 THz). Although the whole IR spectrum (10-4000 cm$^{-1}$) can in principle be probed, the 1000-2000 cm$^{-1}$ mid-IR (30-60 THz) and the 3000-4000 cm$^{-1}$ (90-120 THz) spectral domains have certainly been the most employed in the literature. To date far IR/THz vibrational spectroscopy has been mostly developed and employed to characterise crystals, semiconductors, liquids and biomolecules in the condensed phase.

We therefore propose in the context of this thesis to investigate molecular systems produced in the gas phase that still suffer from spectral congestion, without adding up one more dimension to vibrational spectroscopy (i.e. ion mobility, isotopic substitution), but instead by exploring the range of the far infrared (< 800 cm$^{-1}$, <24 THz) signatures. Our main aim is to see whether the far infrared signatures can be more conformational selective than the more
1.2. FAR INFRARED SPECTROSCOPY

traditional infrared domain (1000-4000 cm$^{-1}$, 30-120 THz).

1.3 Far infrared/Tera-Hertz experimental and theoretical spectroscopy

Development and applications of far infrared spectroscopy in molecular beam conditions has been prevented for a long time because of the lack of intense far infrared sources. The arrival of Free Electron Lasers (FEL) producing far infrared light gave the "green light" for far infrared molecular beam spectroscopy in the gas phase. These far infrared facilities can be found at the FHI (Fritz Haber Institute) in Berlin-Germany, and at the Radboud University in Nijmegen in The Netherlands (FELIX laboratory). All the experimental spectra presented in this thesis have been measured at the FELIX laboratory. The Free electron laser principles are presented in section 3.6 of chapter 3. As discussed above, we mainly study biological building blocks in this thesis but other equivalent set-ups have been developed at the same time for probing gas phase molecular systems in the far-IR/THz domain and dedicated to other types of systems. For instance, investigations on water clusters$^{96,97}$ and metal clusters$^{98–104}$ are performed elsewhere, mainly by Asmis and Fielicke's groups (at the FHI in Germany).

My thesis work is based on the initial demonstration by the groups of Dr A. M. Rijs at FELIX laboratory and Prof. MP Gaigeot in Evry that far IR/THz spectra can indeed be extremely powerful in the identification of 3D structures of peptides. In their seminal work$^{105}$, they showed that far infrared spectroscopy could differentiate the very subtle difference in two peptides structures of Ac-Phe-Gly-NH$_2$ that were just differing by axial and equatorial forms of the $\gamma$-turn (i.e. a very subtle difference in the hydrogen bond network of the $\gamma$-turn). Such subtle difference was out of reach of the spectroscopic signatures recorded in the mid-infrared$^{105,106}$. For this system, the far infrared range was shown far more efficient to distinguish between very similar three dimensional conformations. The far infrared spectrum was also found extremely well-resolved, with very thin peaks, in which spectral congestion could basically not be seen. One of the aims of my thesis is to check whether the selectivity between conformations remains more efficient using the far infrared/THz spectral signatures for systems that suffer of spectral congestion in the range 800-4000 cm$^{-1}$.

The discussions above are the starting point of my thesis work leaving us with the main following questions:

- Are gas phase far infrared/THz spectra systematically well resolved, with well-defined thin peaks, do we basically avoid spectral congestion?

- What are the vibrational modes probed in the far infrared/THz domain? Are these modes harmonic or anharmonic?

- What is the correct theoretical tool to calculate far-infrared/THz spectra and how to possibly further improve the quality of theoretical spectra in this spectral domain?
• How far can we go in terms of conformational assignment using the far infrared/THz vibrational signatures? Is this range really more conformational selective, and why?

1.4 Scope of this thesis

Chapters 2, 3 and 4 of this manuscript respectively present the systems investigated in this thesis (in relation with the motifs discussed in this chapter for proteins and DNA scaffolds), the IR-UV ion dip vibrational spectroscopy in the far-IR/THz domain (our experimental tool) and the ab initio DFT-MD (DFT-based molecular dynamics, DFT being the Density Functional Theory) theoretical method for anharmonic vibrational spectroscopy in the far IR/THz (as well as alternative methods).

My personal advice for the reader curious about this work would be to read first chapters 9, 10 and 11 that present our up-to-date answers to the main questions from above. However, before answering these questions one has to prove that the combination of IR-UV ion dip gas phase spectroscopy in the far infrared/THz domain with ab initio DFT-MD based molecular dynamics for infrared spectra calculations works. Furthermore one has to demonstrate that far infrared/THz spectra are sufficiently resolved in order to possible extract information, same for the dynamical spectra.

Chapter 5, 6, 7 and 8 make such demonstrations (among several other discussions). These chapters are organised around the papers we have published up to now\textsuperscript{64,65,107,108} and are dedicated respectively to the Ac-Phe-Pro-NH\textsubscript{2} system\textsuperscript{64}, a whole Ac-Phe-'AA'-NH\textsubscript{2} dipeptide series\textsuperscript{65} (where AA is replaced by one amino acid among a series of 6), a series of phenol derivatives\textsuperscript{107} and saligenin-water clusters\textsuperscript{108}.

These papers, taken separately, also partially answer the questions written above that will be fully answered in chapters 9, 10 and 11:

• Are gas phase far infrared/THz spectra systematically well resolved, with well-defined thin peaks, do we basically avoid spectral congestion?

All chapters answer to this particular question and demonstrate the strength of the far infrared/THz spectral domain in avoiding spectral congestion, at least for all systems investigated in my thesis. We will also show that the signatures are not so easy to characterize and that spectra calculations are very important into the final analysis of the peaks.

• What are the vibrational modes probed in the far infrared/THz domain? Are these modes harmonic or anharmonic?

Chapter 9 presents a far infrared vibrational mode analysis of all systems investigated in the context of this thesis (see chapter 2 for the whole list of building blocks). Our final aim is to build a general map of the motions that are responsible for the spectroscopic signatures recorded in the far infrared/THz domain (<800 cm\textsuperscript{-1}, <24 THz) as it exists in the more common 3000-4000 and 800-1800 cm\textsuperscript{-1} infrared ranges (see table 1.2). Our current table is presented at the end of chapter 9. The answer to the question
of harmonicity/anharmonicity of the identified far IR/THz vibrational modes is made in chapter 10, which also discusses what theoretical methods are adequate for spectra calculations in the far IR/THz domain.

- **What is the correct theoretical tool to calculate far-infrared/THz domain spectra and how to further improve the quality of theoretical spectra in the far infrared/THz in the future?**

In chapter 10, three theoretical methods are tested for the calculation of vibrational infrared spectra in the far infrared/THz domain, the theoretical spectra are systematically compared with the experimental ones. This is done for most of the systems introduced in chapter 2. It provides the demonstration of the quality of the DFT-MD/BLYP-D3 representation that takes into account anharmonicities of the potential energy surface and of the dipole surface as well as mode couplings. Comparisons between anharmonic DFT-MD, anharmonic VPT2 and harmonic spectra are presented and we identify here in which context (i.e. for which systems, chemical functions, vibrational motions) the anharmonicities need to be taken into account.

- **How far can we go in terms of conformational assignment using the far infrared/THz vibrational signatures? Is this range really more conformational selective, and why?**

While most of the building blocks investigated in my thesis have already been assigned in the literature in terms of their 3D conformations using other vibrational domains (despite sometimes some ambiguities in these assignments), one more ambitious building block model for β-sheets has been investigated in my work, especially chosen because the spectroscopic literature could not assign its detailed structure\(^{69–71}\). Chapter 11 thus presents the conformational assignment of the β-sheet model (Ac-Phe-OMe)\(_2\) using far IR/THz spectral signatures. This chapter shows the power of the far infrared/THz signatures for conformational assignments and in particular this chapter is a demonstration that far infrared signatures are beyond spectral congestion (here in terms of multiple conformations providing similar signatures).

I believe my thesis work also shows how essential is the synergy between experiments and theoretical calculations for far IR/THz spectroscopy in order to answer these questions.

Conclusions to this work and perspectives for the near future will be presented in chapter 12.
Chapter 2

Brief description of the systems investigated.

This chapter gathers background information about the systems studied in the context of this thesis: the Ac-Phe-AA-NH$_2$ dipeptide series (where AA stands for one Amino Acid), the Ac-Phe-OMe monomer and its dimeric form (Ac-Phe-OMe)$_2$, Z-Ala$_6$-NH$_2$ system, phenol derivatives, two nucleobases and one nucleobase analogue. The structures are presented hereafter with chemical schemes and labels that display the internal coordinates (mainly dihedral angles) or the hydrogen atoms involved in the dihedral angles used for the ICDOS decompositions presented in subsection 4.3.1 of chapter 4. All the experimental measurements have been done by Dr. Sander Jaeqx, Daniël Bakker, Dr. Faady Siouri and myself, in collaboration with the Rijs and de Vries’s groups (FELIX laboratory, Radbout University, Nijmegen, The Netherlands; Department of Chemistry & Biochemistry, University of California, Santa Barbara, USA). All the structures of the systems investigated have been found in literature except for Ac-Phe-OMe (monomer and dimer) for which a conformational search has been performed prior to the dynamics. For the structures displayed in this chapter, all have been optimised or re-optimised at the BLYP-D3/6-311+G(d,p) level with the Gaussian package$^{109}$, prior to ab initio molecular dynamics.

2.1 Peptidic systems

Proteins play an important role in biology$^{10–12}$. A protein is a peptide (i.e. a polymer made of amino acids, see figure 2.1) that plays a role in biological processes. Only 21 amino acids (that differ from another by the nature of the side chain) are being used in the synthesis of proteins in nature (see figure 2.2).

\[
\begin{align*}
\text{H}_2\text{N} & \quad \text{COOH} \\
\text{C}_\alpha & \\
\text{H}_\alpha & \quad \text{Side Chain}
\end{align*}
\]

Figure 2.1: Scheme of a generic amino-acid.
Figure 2.2: The 21 α-amino acids found in eukaryote proteins, grouped according to their residue. pH values and charges carried at physiological pH 7.4. The figure has been extracted from the Wikipedia page about amino-acids.
2.1. PEPTIDIC SYSTEMS

Considering the reductionist approach discussed in the introduction, we have been characterising the structure of small entities, namely small sequence peptides. The building blocks for peptides are amino acids. Amino acids are built up from a central carbon atom (denoted C$_\alpha$), connected to an amino group (-NH$_2$), a carboxylic acid group (-COOH), a hydrogen atom (H$_\alpha$) and a side chain, see scheme 2.1.

The side chains come with large variety. For the smallest amino acid, the residue is only one hydrogen atom (glycine). Since the central carbon atom is attached to two hydrogen atoms, this is the only achiral amino acid. All other amino acids are chiral, and occur in natural proteins almost exclusively in their left-handed form. The amino acid side chain can be acidic, e.g. glutamic acid, or basic, e.g. arginine. Three amino acids contain an aromatic moiety such as a phenyl ring (phenlyalanine), a phenol group (tyrosine) or an indole group (tryptophane). Other side chains contain only an aliphatic group, for instance alanine, valine or leucine. Another important amino acid is cysteine. Its side chain ends with an -SH group, which can form a disulfur bond (-S-S-) with another cysteine residue and is very important for protein stability and the overall structure of a peptide. Another original amino acid is proline, where the side chain is bonded to both the central carbon (C$_\alpha$) and the nitrogen atom, forming a cyclic structure.

A peptide is a polymer made of amino acids. In order to form a peptide, a condensation reaction occurs where the amino group of one amino acid reacts with the carboxylic group of another amino acid, hence forming a peptide bond, as illustrated in figure 2.3.

![Figure 2.3: Generic scheme for the peptidic bond formation (amino acids are zwitterions in aqueous solutions, thus COO$^-$ and NH$_3^+$ terminals instead of COOH and NH$_2$ in figure 2.1).](image)

To facilitate the description of peptide and protein structures, a distinction is made between four different levels of structure, illustrated in figure 2.4:

- **The primary structure** is simply the amino acid sequence of the polypeptide chain.
- **The secondary structure** describes how segments of the polypeptide chain can fold into regular patterns such as $\alpha$-helices or $\beta$-sheets (see figure 2.5). This secondary structure is the one investigated in the context of this thesis.
- **The tertiary structure** describes the overall structure of an entire protein, *i.e.* how the secondary structure elements are organised in the 3D space. None of the peptide systems investigated in my work are big enough to observe a tertiary structure.
- **The quaternary structure** describes the organisation of more than one polypeptidic chain. Examples of proteins with a quaternary structure include hemoglobin, DNA polymerase, and ion channels.
CHAPTER 2. BRIEF DESCRIPTION OF THE SYSTEMS INVESTIGATED.

Figure 2.4: Description of protein structures. Reproduced from the internet website: "https://www.boundless.com/biology/textbooks/boundless-biology-textbook/biological-macromolecules-3/proteins-56/protein-structure-304-11437/"
Figure 2.5: The \( \alpha \)-helix and \( \beta \)-pleated sheet of peptides and proteins formed because of hydrogen bonding between C=O carbonyl and N-H amino groups of the backbone. Certain amino acids have a propensity to form an \( \alpha \)-helix, while others have a propensity to form a \( \beta \)-pleated sheet. Reproduced from the internet website: "https://www.boundless.com/biology/textbooks/boundless-biology-textbook/biological-macromolecules-3/proteins-56/protein-structure-304-11437/"
In the following subsections, the primary and secondary structures of the peptidic systems investigated in my thesis are described.

The general view is that the most important structures observed in proteins are $\alpha$-helices and $\beta$-sheets structures built on "well designed/well known" intermolecular hydrogen bonds between amide NH and carbonyl C=O groups that belong to the protein backbone scaffold, see figure 1.4. A more careful analysis of the proteins structures recorded in the PDB (Protein Data Bank) shows that some subtle motifs based on specific hydrogen bonds also actively participate to the final three dimensional structures. These motifs$^{13}$ are typically called C5 found in $\beta$ strands, C7 found in $\gamma$-turns, C10 in $\beta$-turns and 3$^{10}$ helices, with Cn meaning a hydrogen bond ring/nest of size n (n stands for the number of atoms present in the ring formed by the whole hydrogen bond network). Larger hydrogen bonded rings also exist, i.e. C17, C20. It is pivotal to recognise and characterise these motifs within the whole protein scaffold, and this is where vibrational spectroscopy might be very useful.

2.1.1 Ac-Phe-AA-NH$_2$ dipeptide series

The Ac-Phe-'AA'-NH$_2$ dipeptide series has been chosen in particular because widely studied and structurally unravelled in the last decade$^{61–63}$ using the 'classical' 1000-2000 and 3000-4000 cm$^{-1}$ spectral ranges. In my work, 'AA' stands for one of the following amino acids, glycine (Gly), alanine (Ala), proline (Pro), cysteine (Cys), serine (Ser), valine (Val) and the series always contains the phenylalanine chromophore (Phe) used for the UV excitation and ionization in the experiment, see section 3.2 in chapter 3. The natural terminations of amino acids in the gas phase are the NH$_2$ and OH functions, see figure 2.1, but they are replaced by different caps in the dipeptide series investigated here. The NH$_2$ function is replaced by an "Ac" (acetyl) CH$_3$CO group, linked to the nitrogen atom of the amide N-H. The OH natural function is replaced by an NH$_2$ function.

This whole series has been studied in the context of the paper: "Mapping gas phase dipeptides motions in the far-infrared and terahertz domain"$^{65}$ presented in chapter 6. Their structures were already known from spectroscopic works in the 1000-4000 cm$^{-1}$ domain$^{61–63}$. This series is therefore a good test for our theoretical methods in the far infrared spectral domain. The structures have been extracted from literature and used after geometry optimisation at our level of representation. The experiment-theory comparison for our main tool BLYP-D3-MD is presented in this paper but also in chapter 10 where all the systems investigated have been taken into account. Once the theoretical method is validated, a mapping of the vibrational modes is possible. This analysis has been done for the dipeptide series in the same paper$^{65}$, see chapter 6, and is put into perspective in a more general mapping of the vibrational modes of all the systems I have investigated in my thesis, as presented in chapter 9.

According to our own work as well as previous studies$^{61–63}$, the Ac-Phe-AA-NH$_2$ dipeptides are folded in either a $\gamma$-(C7) or a $\beta$-(C10) turn geometry, see representations in figures 2.6 and 2.7. Figure 2.6 presents generic schemes of these dipeptides and it also displays the labels used to define the internal coordinates for the ICDOS decompositions, see subsection 4.3.1 in
chapter 4. Dihedral angles \((\phi, \psi, \omega, \chi_1 \text{ and } \chi_2)\) are directly reported in the schemes, they correspond to the Ramachandran notation\(^{110}\) defining peptidic structures.

\[
\begin{align*}
\text{Gly: } R &= \text{H} \\
\text{Ala: } R &= \text{CH}_3 \\
\text{Pro: } R &= \text{CH}_2-\text{CH}_2-\text{CH}_2-\text{N} \\
\text{Cys: } R &= \text{CH}_2-\text{SH} \\
\text{Ser: } R &= \text{CH}_2-\text{OH} \\
\text{Val: } R &= \text{CH}_-(\text{CH}_3)_2
\end{align*}
\]

Figure 2.6: Chemical schemes of Ac-Phe-AA-NH\(_2\) dipeptides for \(\gamma\)-turn C7 (left) and \(\beta\)-turn C10 (right) conformations with the labels used in the text for the atoms and dihedral angles \((\phi, \psi, \omega \text{ and } \chi)\) corresponding to the Ramachandran notation\(^{110}\). Residue 'AA' stands for Glycine (Gly), Alanine (Ala), Proline (Pro), Cysteine (Cys), Serine (Ser) and Valine (Val) in our work. This figure is extracted from our paper: "Mapping gas phase dipeptides motions in the far-infrared and terahertz domain"\(^{65}\).

The NH\(_2\) terminal function has been actually chosen to induce a folding and from motifs such as \(\gamma\) or \(\beta\)-turns. The signatures of \(\gamma\) and \(\beta\)-turns will be investigated in chapters 5, 6 and 9 but \(\omega(\text{NH})\) signatures of these NH\(_2\) functions (signatures of the out of plane motion of the hydrogen atoms) can not be predictive for protein signatures since the hydrogen bonds are formed by NH amide functions in "natural" larger peptide or protein. We believe that we do not have enough systems here to propose far infrared signature comparisons between the same motif formed by NH\(_2\) or NH groups but this would be interesting. We have no C7 formed by NH functions among the systems investigated but we have two C10 formed by NH can be found in Z-Ala\(_6\)-NH\(_2\), see subsection 2.1.3.

In this work we systematically measured REMPI spectra (presented in section 3.5 of chapter 3) to ensure their similarities with the published results. Based on that, the assignments of the different UV spectral features to conformations have been taken from these previous studies\(^{61-63}\):

- Only one conformer for the Ac-Phe-Gly-NH\(_2\), Ac-Phe-Ala-NH\(_2\) and Ac-Phe-Ser-NH\(_2\) dipeptides,
CHAPTER 2. BRIEF DESCRIPTION OF THE SYSTEMS INVESTIGATED.

Figure 2.7: Generic 3D structures of capped Ac-Phe-AA-NH$_2$ dipeptides. The 2 caps are 'Ac' for acetyl (CH$_3$-CO) and NH$_2$, respectively for the N- and C-terminals. 'AA' stands for Amino Acid, and 'AA' = Glycine (Gly), Alanine (Ala), Proline (Pro), Cysteine (Cys), Serine (Ser) or Valine (Val) in the present work.

a): γ-turn structure of Ac-Phe-Gly-NH$_2$, Ac-Phe-Ala-NH$_2$, Ac-Phe-Ser-NH$_2$, Ac-Phe-Val-NH$_2$, and Ac-Phe-Cys-NH$_2$ dipeptides. For Ac-Phe-Gly-NH$_2$, Ac-Phe-Ala-NH$_2$ and Ac-Phe-Val-NH$_2$, the amino acid residue does not interact with its immediate environment. For Ac-Phe-Ser-NH$_2$ and Ac-Phe-Cys-NH$_2$, the residue interacts with one Amide C=O group.

b): γ-turn of Ac-Phe-Pro-NH$_2$.

c): β-turn of Ac-Phe-Cys-NH$_2$ (β-turn type I).

d): β-turn of Ac-Phe-Pro-NH$_2$ (β-turn type VIa).

This figure is extracted from our paper: "Mapping gas phase dipeptides motions in the far-infrared and terahertz domain"$^{65}$. Details of the DFT-MD trajectories performed for these systems are presented in chapter 4 and results are presented in chapters 5, 6, 9 and 10.
corresponding to a γ-turn geometry where one internal hydrogen bond leads to the formation of a seven membered ring (denoted C7 interaction and leading to a γ-turn structure). Such conformation is illustrated in Figures 2.6 and 2.7-a.

- Two conformers of Ac-Phe-Val-NH₂, both with a γ-turn structure (Figures 2.6 and 2.7-a) and differing by the orientation of the CH(CH₃)₂ residue with respect to the backbone.

- For Ac-Phe-Cys-NH₂, two main conformations, one of a γ-turn type and one of a β-turn type (one internal hydrogen bond leading to the formation of a ten membered ring, denoted as C10 interaction and β-turn geometry). See Figures 2.6, 2.7-a and 2.7-c for an illustration of the 3D structures respectively for the γ-turn and β-turn geometries. For the β-turn of Ac-Phe-Cys-NH₂, the precise orientation of the S-H group was still an open issue when we started our own work, which we believe is solved in our work, see especially the supplementary material of the paper: Mapping gas phase dipeptides motions in the far-infrared and terahertz domain, presented in chapter 6. We find that the SH group interacts with one backbone C=O group and the assigned structure is presented in figure 2.7-c.

- Two structural isomers for Ac-Phe-Pro-NH₂ dipeptide, i.e. a γ-turn and a β-turn. The proline residue forms a ring and the γ-turn and β-turn structures are illustrated separately in Figures 2.7-b and 2.7-d for the sake of clarity.

2.1.2 Ac-Phe-OMe and (Ac-Phe-OMe)₂

These two systems have been chosen because the exact 3D structure and in particular the orientation of the aromatic phenylalanine ring has not been found for the dimer before we started our work. For the dimer the conformational assignment appears to be particularly challenging by using the classical ranges 1000-2000 and 3000-4000 cm⁻¹ leaving one with only the information about the global family structure, i.e. a β-Sheet structure as shown by Gerhards’s group in the last decade. A generic scheme of a β-sheet structure is presented in figure 2.8. The dimer is therefore a good model to study peptide aggregation. On the other hand for the monomer the exact conformation was already found in the same series of papers using the mid-infrared range.

![Figure 2.8: Generic scheme for an antiparallel β-sheet structure.](image_url)
Note that the NH$_2$ termination present in the dipeptide series (subsection 2.1.1) has been replaced here by an (NH)-(CO)-O-CH$_3$ cap and the COOH termination replaced by a (CO)-O-CH$_3$ cap. These caps have been chosen because the O-CH$_3$ function is not expected to induce hydrogen bonding and therefore folding, with the assumption that a linear backbone would then allow the peptide to dimerise into a $\beta$-sheet structure. Chapter 11 shows that the gas phase conformation of the monomer is not systematically conserved into the dimer.

The definitive conformational assignment of the dimer has been achieved thanks to the far infrared spectral signatures and is presented in chapter 11. This is the perfect example that shows that the far infrared signatures are more conformer selective than the classical 1000-4000 cm$^{-1}$ range and the reasons are discussed in the same chapter. A similar example is presented in the paper: "Gas-Phase Peptide Structures Unraveled by Far-IR Spectroscopy: Combining IR-UV Ion-Dip Experiments with Born-Oppenheimer Molecular Dynamics Simulations"$^{105}$ where only the far infrared spectrum can show differences between two close conformations of the Ac-Phe-Gly-NH$_2$ dipeptide.

Three different conformations of the Ac-Phe-OMe monomer are presented in figure 2.9 based on the same backbone's structure $\beta_L$ which corresponds to one possible backbone orientation with values of dihedral angles: $\phi \simeq -160^\circ$, $\psi \simeq +160^\circ$ and $\omega \simeq 180^\circ$ (in the Ramachandran notation$^{110}$), see figure 2.10. Note that a complete conformational search has been performed and is presented in chapter 11 and all the other backbone orientations have been excluded based on comparisons between theoretical and experimental infrared spectra in the 1000-2000 cm$^{-1}$ range. Three stable orientations of the aromatic ring can be found and correspond to three stable isomers: $g^+$ (the assigned conformation), a and $g^-$ corresponding to values 60$^\circ$, 180$^\circ$ and -60$^\circ$, respectively, of the dihedral angle $\chi_1$, see figures 2.9 and 2.10.

The assigned conformation of the dimer is also presented in figure 2.9, see chapter 11 for all details. The two strands in (Ac-Phe-OMe)$_2$ adopt a $\beta_L(g^+)$ and a $\beta_L(g^-)$ conformation respectively. In figure 2.10, is presented a scheme that displays the $\phi$, $\psi$, $\omega$, $\chi_1$ and $\chi_2$ dihedral angles and all the labels needed to understand the ICDOS decompositions for infrared interpretation, see subsection 4.3.1 in chapter 4 of this work.

The second reason to investigate these systems is that (Ac-Phe-OMe)$_2$ is probably the smallest peptidic system that adopts a $\beta$-sheet structure, a common structure for large size proteins and peptides. One of our goals is to provide specific signatures of protein motifs in the far infrared-THz domain. Characterising (Ac-Phe-OMe)$_2$ is a step in this direction.
2.1. PEPTIDIC SYSTEMS

Figure 2.9: Ac-Phe-OMe and (Ac-Phe-OMe)_2 peptides optimised geometries in the β_L general structural organisation. The "g+" conformation corresponds to the assigned one for the monomer (in our experimental conditions). The dimer structure corresponds to the assigned one (in our experimental conditions). Free energies have been calculated at 50K at the BLYP-D3/6-311+G(d,p) level. See scheme 2.10 for more notations. Details of the DFT-MD trajectories performed for these systems are presented in chapter 4 and results are presented in chapters 9, 10 and 11.

Figure 2.10: Chemical scheme of Ac-Phe-OMe monomer with the labels used in the text for the atoms and the definition of dihedral angles (φ, ψ, ω and χ) that correspond to the Ramachandran notation. The caps 'Ac' and 'OMe' correspond respectively to an acetyl group CH₃-CO and to a O-CH₃ group.
2.1.3 Z-(Ala)$_6$-NH$_2$

As discussed in the introduction of my thesis, one reason to investigate far infrared signatures is to find specific signatures of diverse common motifs in biology. The peptidic systems described above, i.e. the dipeptide series and Ac-Phe-OMe display a small variety of structural motifs. We thus chose to investigate the larger Z-Ala$_6$-NH$_2$ peptide because it is a large peptide that carries several amine NH functions (6) that can have a large variety of environments and therefore possibly involved in several possible hydrogen bond motifs. This is, to our knowledge, one of the biggest peptides studied in the gas phase in the far infrared range. This is really interesting in the context of the mapping of vibrational modes.

The conformational assignment of Z-Ala$_6$-NH$_2$ has been presented in the thesis by Sander Jaeqx: "Protein folding forces probed by infrared action spectroscopy" (2014, supervisor: A. M. Rijs., Nijmegen) and is therefore not presented here. Z-Ala$_6$-NH$_2$ adopts a globular conformation, so that unfortunately we might expect infrared signatures and vibrational modes different to what could be expected for organised biological materials (in helices or sheets). The same NH$_2$ C-terminal cap as in the dipeptide series is present in this system and the 'Z' cap corresponds to a benzyl group linked to the backbone via an ester group $C_6H_6-CH_2-OCO$ for the N-terminal of the peptide. The terminal ring has been added as chromophore in order to perform IR-UV ion dip spectroscopy, see section 3 for more details on the method. See figure 2.11 for a 3D representation.

As presented in Sander Jaeqx’s thesis, conformational transitions occur during the DFT-MD trajectory of Z-Ala$_6$-NH$_2$ at 50K of temperature (even already at 25K) and it seems that two structures possibly co-exist and/or interconvert in the experimental conditions. As I am only interested in the vibrational modes analysis of this system, only one conformation, presented in figures 2.11 and 2.12, is taken into account in my analyses since the vibrational modes are conformational dependent. Note that this conformation is not a minimum on the BLYP-D3/6-311+G(d,p) potential energy surface (see subsection 4.5 for our theoretical representation) and the structure displayed in figure 2.11 has been optimised at the B3LYP/6-311+G(d,p) level.

In the molecular dynamics, the NH functions of the first and second amino acids in Z-Ala$_6$-NH$_2$ are linked to the same C=O function (of the last amino acid, see figure 2.12) forming respectively a C20 and a C17 hydrogen bond interaction (corresponding to a 20 and 17 hydrogen bonded membered ring). The (NH)$_3$ function is more or less simultaneously bonded with two C=O functions at the same time (the C=O group of the first amino acid and the C=O that binds the Z-ring with the first amino acid). The (NH)$_4$ function is free of hydrogen bond. The last two NH functions (NH)$_5$ and (NH)$_6$ in Z-Ala$_6$-NH$_2$ are involved in β-turns of type I and IV respectively (in the Ramachandran notation type I: $\phi_{x+2}=-60^\circ$, $\psi_{x+2}=-30^\circ$, $\phi_{x+3}=-90^\circ$ and $\psi_{x+3}=0^\circ$, with x the amino acid that carries the CO function; type IV is a kind of miscellaneous category, which includes two of the dihedral angles more than 40$^\circ$ away from the ideal values for any other type of β-turn). In Z-Ala$_6$-NH$_2$, we have $\phi_4=-60^\circ$, $\psi_4=-21^\circ$, $\phi_5=-76^\circ$, $\psi_5=-8^\circ$, $\phi_6=74^\circ$, $\psi_6=-62^\circ$. 
2.1. PEPTIDIC SYSTEMS

Figure 2.11: 3D structure of the conformer of the Z-Ala$_6$-NH$_2$ system investigated in my work and optimised here for the picture at the B3LYP/6-311+G(d,p) level of theory. Note that this structure is not a minimum on the BLYP/6-311+G(d,p) potential energy surface. Details of the DFT-MD trajectory performed for this system are presented in chapter 4 and results are presented in chapter 9.

Figure 2.12: Chemical scheme of the conformation of Z-Ala$_6$-NH$_2$ peptide investigated in this work ($Z= C_6H_6$-CH$_2$-O-C=O). Black lines correspond to the hydrogen bonds formed in that conformer ("hydrogen bonded rings" denoted C7, C10 and C17), see figure 2.11 for the 3D globular representation.

For the ICDOS decompositions used in chapter 9, the NH wagging motions in the far infrared of the NH amide bonds are extracted using dihedral angles H-N-C-C. The number associated to each amino acid in the sequence is added to identify the function (see figure 2.12). The amino acid number 1 is the one linked to the aromatic ring and the 6$^{th}$ amino acid is the one linked to the NH$_2$ terminal function. For the NH$_2$ symmetric and asymmetric motions, the two H-N-C-C dihedral angles are combined together by adding and subtracting the two values.
2.2 Phenol derivatives

Our two papers: "Anharmonic, dynamic and functional level effects in far-infrared spectroscopy: Phenol derivatives"\textsuperscript{108} and "Fingerprints of Inter- and Intramolecular Hydrogen Bonding in Saligenin-Water Clusters Revealed by Mid- and Far-Infrared Spectroscopy"\textsuperscript{111} are presented in chapters 7 and 8 respectively. In the first paper, a phenol derivative series has been chosen because the conformational assignments are not really challenging. Therefore we can focus on vibrational modes assignments and comparison between different levels of calculations for spectroscopy which is one of my interests. The second series, saligenin and its water clusters has been chosen specifically for conformational assignment, once our first paper has shown that the methods are good enough to perform assignments without ambiguity. The large diversity of environments in which the OH function is involved allowed us to perform a correlation between the signatures and the frequencies in the far infrared domain. A last series of four phenol derivatives has been chosen for investigating the progressive enhancement of the strength of the hydrogen bond. This series is reported in chapters 9 and 10.

2.2.1 First series: phenol, catechol, saligenin, salicylic acid and nitrophenol series.

Our first paper presented in chapter 7, discusses the quality of different levels of calculations using as models phenol, catechol, saligenin, nitrophenol and salicylic acid molecules. The structures are presented in figure 2.13. The conformational assignment of these phenol derivatives is not really challenging since only one possible conformation is possible for phenol, catechol and nitrophenol. Despite several possible stable conformations, only one structure has been found in similar experimental conditions (similar to our experimental conditions) for saligenin. Saligenin's conformational assignment has been done by Kumar \textit{et al.} using microwave and IR-UV ion dip spectroscopy\textsuperscript{112}. For salicylic acid, two structures coexist in our experimental conditions. The conformational assignments have been done by Yagahi \textit{et al.} with IR-UV ion dip spectroscopy\textsuperscript{113} in the range 3000-4000 cm$^{-1}$. All structures have been used for DFT-MD after geometry optimisation.

2.2.2 Second series: saligenin and its water clusters.

Our second paper presented in chapter 8, discusses the conformational assignments of saligenin-(H$_2$O)$_n$ clusters with $n=1$-3. One single conformation is observed for saligenin-(H$_2$O)$_1$ and saligenin-(H$_2$O)$_2$, while two conformations co-exist in our experimental conditions for saligenin-(H$_2$O)$_3$. The four assigned structures are presented in figure 2.14. A conformational search has been performed by Daniël Bakker by adding water molecules to the bare saligenin or to the saligenin-(H$_2$O)$_{n-1}$ complexes, water by water. Because of the large variety of hydrogen bonds that can be formed between the OH functions in these systems, this paper also shows a correlation between the direct and indirect signatures of the modes inducing hydrogen bond deformations and the strength of the hydrogen bonds.
2.2. PHENOL DERIVATIVES

Figure 2.13: Molecular structures of the phenol derivatives, optimized at the B3LYP-D3/6-311+G(d,p) level of theory. Black spheres represent carbon, white hydrogen, red oxygen and purple nitrogen. The H-bonded OH bond-length $L_{OH}$ is listed as a measure of the strength of the hydrogen bond when present. This figure has been copied from our paper: "Anharmonic, dynamic and functional level effects in far-infrared spectroscopy: Phenol derivatives" presented in details in chapter 7. Details of the DFT-MD trajectories performed for these systems are presented in chapter 4 and results are presented in chapter 7.

Figure 2.14: Assigned structures of the measured saligenin-(H$_2$O)$_{1-3}$ species. Optimizations at the MP2/6-311+G(2d,p) level of theory. The conformational assignment is presented in our paper: "Fingerprints of Inter- and Intramolecular Hydrogen Bonding in Saligenin-Water Clusters Revealed by Mid- and Far-Infrared Spectroscopy" presented in chapter 8. This figure has been copied from this paper and the colors are used to identify each OH function in the context of this work. Details of the DFT-MD trajectories performed for these systems are presented in chapter 4 and results are presented in chapter 8.
2.2.3 Third series: phenol, catechol, saligenin and phenol-water complex.

For chapters 9 and 10 where we discuss the mapping of vibrational modes and show comparisons of methods for a large range of systems, we chose to restrict the phenol derivative series (and the specific study of the OH function) to only four systems: phenol, catechol, saligenin and the phenol-water complex. Figure 2.15 presents the 3D structures and figure 2.16 presents a generic scheme of the phenol derivatives that displays the labels of the hydrogen atoms used to extract the wagging motions for the internal coordinates decomposition.

These four systems have been chosen for the progressive enhancement in the strength of the O-H···O hydrogen bond: phenol has one free OH according to the geometry optimisation in our theoretical representation (see section 4.5), catechol has a weak OH···OH hydrogen bond (hydrogen bond length: 2.18 Å), saligenin has a strong hydrogen bond (hydrogen bond length: 1.96 Å) and the phenol-water complex has a very strong hydrogen bond (hydrogen bond length: 1.88 Å).

Note that the study of the phenol-water complex has not been presented in the two published papers reported in chapters 7 and 8. Its structure has been assigned using IR UV ion dip and stimulated Raman-UV double resonance spectroscopies by Watanabe et al.114 The assigned structure is presented in figure 2.15. Note that different initial orientations of the water molecule might be considered, but geometry optimisations with different initial water orientations all converge to the presented structure, suggesting that only this orientation of the water molecule corresponds to a minimum on the potential energy surface (BLYP-D3 level).

![Figure 2.15: 3D structures of phenol, catechol, saligenin and phenol-water species measured and optimised at the BLYP-D3/6-311+G(d,p) level.](image)

Details of the DFT-MD trajectories performed for these systems are presented in chapter 4 and results are presented in chapters 9 and 10.
2.3 Base pairs and analogues

To extend further our far infrared analyses, some nucleobase pairs have been characterised in collaboration with the group of Mattaniah de Vries (University of Santa Barbara, USA) providing us the experimental spectra. The mapping of the vibrational modes of the nucleobases can be found in chapter 9 while a method comparison is presented in chapter 10. The mode assignments are also presented in Faady Siouri’s PhD manuscript (de Vries’s group): "Excited State Dynamics of Isolated Nucleobases and Base Pairs" defended in 2017. All the structures have been found in the literature and used for DFT-MD after geometry optimisation with our level of representation.

In DNA, especially the well known double helix, see figure 2.19, strands made of sugars and phosphate groups are linked through base pairs being hydrogen bonded. In DNA (DeoxyriboNucleic Acid), there are four DNA bases, cytosine, thymine, guanine and adenine, see figures 2.17 and 2.19. The hydrogen bond patterns between the bases are of importance to characterise as they are the main elements for the double strand scaffold.

In RNA (Ribonucleic Acid), thymine is replaced by uracil. Nucleobases are presented in figure 2.17.

In DNA, adenine and thymine on the first hand, guanine and cytosine on the other hand, bind together via hydrogen bonds to form base pairs. Canonical structures are presented in
CHAPTER 2. BRIEF DESCRIPTION OF THE SYSTEMS INVESTIGATED.

figures 2.18 and 2.19.

![Chemical schemes of the two canonical base pairs, guanine-cytosine (GC) and adenine-thymine (AT).](image1)

Figure 2.18: Chemical schemes of the two canonical base pairs, guanine-cytosine (GC) and adenine-thymine (AT).

![Schematic representation of the DNA double helix.](image2)

Figure 2.19: Schematic representation of the DNA double helix (Reproduced from the internet website: "https://www.quora.com/What-makes-up-the-rungs-of-a-DNA-ladder").
In the context of this thesis, three base pairs have been investigated: guanine-guanine and guanine-cytosine and one analogue base pair, ethylated guanine-methylated cytosine. All the conformational assignments of these systems in the gas phase have been obtained in the last decade\textsuperscript{76–78}.

Although, guanine-guanine is not one of the two canonical base pairs, it can be found in telomeres (extremity of a DNA strand). Even though two isomers are present in our experimental conditions (see F. Siouri’s PhD manuscript), only one will be studied here, its structure is presented in figure 2.20. The labels of the hydrogen atoms in this figure will be used for the ICDOS internal coordinates decomposition of infrared spectra that will be applied in chapter 9. The conformational assignment of the two conformers has been done by Nir \textit{et al.}\textsuperscript{76} using IR-UV ion dip spectroscopy in the range 3000-4000 cm\textsuperscript{-1}.

![Figure 2.20: 3D structure of the guanine-guanine base pair studied in the context of this thesis. The labels correspond to the hydrogen atoms that will be used for ICDOS analyses of N-H wagging motions, presented in chapter 9. After geometry optimisation at the BLYP-D3/6-311+G(d,p) level, the three intermolecular hydrogen bond length values are, from top to bottom, 2.04, 1.86 and 1.75 Å. Details of the DFT-MD trajectory performed for this system are presented in chapter 4 and results are presented in chapters 9 and 10.](image)
At least two conformations have been found in the gas phase for the guanine-cytosine base pair but only one isomer will be studied here. The conformational assignment of this isomer is presented in the paper by Bakker et al., see figure 2.21. The labels of the hydrogen atoms in figure 2.21 illustrate the internal coordinates decomposition that will be used in chapter 9.

![Figure 2.21: 3D structure of the guanine-cytosine base pair studied in the context of this thesis. The labels correspond to the hydrogen atoms used for ICDOS analyses of N-H wagging motions, presented in chapter 9. After geometry optimisation at the BLYP-D3/6-311+G(d,p) level, the three intermolecular hydrogen bond length values are, from top to bottom, 2.04, 1.93 and 1.79 Å. Details of the DFT-MD trajectory performed for these systems are presented in chapter 4 and results are presented in chapters 9 and 10.](image)

In DNA, the guanine-cytosine base pair is organised into a Watson-Crick configuration (the canonical one, see figure 2.18). As can be seen in figures 2.20 and 2.21, neither the guanine-guanine nor guanine-cytosine base pairs adopt such specific Watson-Crick configuration in our gas phase experimental conditions and yet, no Watson-Crick structures have been observed in equivalent gas phase conditions. This is presumably due to their short electronic excited state lifetime. The 1 ps (or less) lifetime of biological DNA tautomers is believed to be a protection against photochemical damages through internal conversion back to the electronic ground state. This property prevents the G···C Watson-Crick conformation to be studied using the IR-UV ion dip method that requires a long lifetime S1 level, see subsection 3.2.

The base pair formed by 9-ethylguanine with 1-methylcytosine is interesting because it adopts a Watson-Crick conformation in our experimental conditions, that corresponds to the structure of the Guanine-Cytosine in DNA and that is not found in the gas phase for the Guanine-Cytosine base pair. The 9-ethylguanine ··· 1-methylcytosine pair in the WC (Watson-Crick) conformation is presented in figure 2.22. The paper "Photochemical selectivity in guanine-cytosine base-pair structures" presents a good agreement between the experimental infrared spectrum and the theoretical spectrum of this conformation in the range 2800-3800 cm⁻¹,
which suggests that this conformation is indeed present in the experimental conditions. The good agreement between theory and experiment in the far infrared that is presented in this thesis goes along the same line. But at my best knowledge, no full conformational assignment has been published (i.e. comparisons of the experimental spectrum with theoretical spectra of multiple conformations).

![Figure 2.22: 3D structure of the 9-ethylated guanine···1-methylated cytosine base pair in the Watson-Crick canonical configuration investigated in the context of this thesis. The labels correspond to the hydrogen atoms used for ICDOS analyses of the N-H wagging motions, presented in chapter 9. After geometry optimisation at the BLYP-D3/6-311+G(d,p) level, the three intermolecular hydrogen bond length values are, from top to bottom, 1.75, 1.89 and 1.89 Å. Details of the DFT-MD trajectory performed for this system are presented in chapter 4 and results are presented in chapters 9 and 10.](image-url)
Chapter 3

IR-UV ion dip spectroscopy.

Summary: I present in a nutshell a short overview of the principles of the experiments that we have conducted in collaboration with our partner’s group of Dr. Anouk M. Rijs at FELIX laboratory, Radbout University in The Netherlands.

The molecules are desorbed from a mixture of the investigated species and carbon black applied to a sample bar by a pulsed YAG laser operating at 1064 nm (1 mJ pulse). The cold (∼10-20 K) gas phase neutral molecules are prepared in a supersonic jet expansion of Argon by a pulsed valve with a backing pressure of 3 bar Argon. For the IR-UV ion dip experiments, the molecules are ionized via a one color Resonance Enhanced Multi-Photon Ionization also called (1+1) REMPI scheme. The $S_0 \rightarrow S_1$ transitions are selected for each conformation of each single molecule, resulting in a UV conformer selected spectroscopy. The UV pulses are produced by a frequency doubled dye laser pumped by a frequency doubled or tripled Nd:Yag laser operated at 10 Hz with typical energy pulse of 1-2 mJ. The generated ions are detected using a reflectron time of flight mass spectrometer. The UV pulses are preceded by far infrared light produced by the Free Electron Lasers FELIX located at the FELIX laboratory at the Radboud University (The Netherlands). Whenever the radiation is resonant with a vibrational transition of the system, the ground state is depleted and the UV photons are not resonant anymore with an electronic transition of the system. We therefore observe a depletion in the mass spectrometer signal. During the experiment, IR-on and IR-off signals are alternated to correct for the long term UV power drifts and the evolution of the source conditions by operating the IR laser at 5Hz and the UV laser at 10Hz. The IR intensities of the measured transitions are corrected for the wavelength dependent photon energy and absorbed number of photons. The experimental spectra presented here are obtained as an average over a minimum of three scans with a minimum of 30 averages per point.

A schematic overview of the set-up used to conduct IR-UV spectroscopy of neutral, cold, conformer selected molecules (and clusters) at the FELIX laboratory is displayed in figure 3.1 and is explained in more details in the next sections. Other gas phase spectroscopies are discussed in section 3.1 and the principles of the IR-UV ion dip method are presented in section 3.2. The production of isolated, cold, conformer selected gas phase neutral species is presented in section 3.3. Their measurements and their ionisations are presented in subsection 3.4 and 3.5 respectively. The far infrared source, the Free Electron Laser FELIX (Nijmegen, The Netherlands) that provides the infrared light down to 90 cm$^{-1}$ is presented in section 3.6.
Figure 3.1: Schematic representation of the whole set-up used for IR-UV double resonance spectroscopy (or IR-UV ion dip spectroscopy). The key components that make up the Free Electron Lasers (FEL), laser desorption sample preparation and Time-of-Flight mass spectrometer are displayed. The translation stage used to precisely position the sample bar with respect to the desorption laser, and the relative positions of the pulsed valve nozzle and the skimmer are displayed in the photos in the inset (Reproduced from "Gas phase IR spectroscopy and structure of biological molecules"\textsuperscript{58}).
3.1 Direct absorption spectroscopy, action spectroscopy and alternative methods to probe vibrational levels

In conventional absorption spectroscopy, we apply the Beer-Lambert law:

\[ I(\omega) = I_0 e^{-\sigma(\omega)ln} \]  

where \( I \) and \( I_0 \) are the intensities of the infrared light measured by the detector for the frequency \( \omega \) and emitted by the source respectively. \( I_0 \) is the path length of the infrared beamlight in the sample, \( n \) is related to the concentration of the molecules in the experimental conditions and \( \sigma(\omega) \) is the absorption cross section for the frequency \( \omega \). In molecular beam conditions, due to the small amount of molecules that will be produced in the gas phase, the difference between intensity of light before, \( I_0 \), and after absorption, \( I(\omega) \), is too small to be measurable by any available infrared detector.

One solution could be the increase of \( l \), the path length of the infrared beamlight. This solution is used for instance by Jean-Marie Flaud or Pascale Roy in France, with a path length over 100 m. But even with pressure conditions much higher than in a molecular beam and using an intense source of light such as the beamline AILES at the SOLEIL synchrotron (Saclay, France), such experimental setup can only address molecular systems of small size (typically lower than 20 atoms).

An alternative is action spectroscopy, where the consequences of IR photon(s) absorption by the molecular system are measured. Usually, a mass spectrometer is used to probe the consequences of the photon absorption, i.e. the ionisation of the system for IR-UV ion dip spectroscopy (used by groups of Rizzo, Mons, Rij, DeVries, Gerhards, ...), the fragmentation of the system for InfraRed Multi Photon Dissociation IR-MPD (used by groups of Maître, Oomens, Von Helden, Fielicke, Compagnon, ...) or the loss of a weakly attached tag for InfraRed - Photo Dissociation IR-PD (used by groups of Lisy, Johnson, Asmis, ...). For these two latter methods, the far infrared experiments might be highly challenging, in particular for the IRMPD spectroscopy since a huge amount of low energy photons would then be needed to fragment the molecular system.

For the IR-MPD and IR-PD action spectroscopies, the intensity of the bands in the spectra will be dependent of the efficiency of the fragmentation of the system and therefore (possibly) not directly comparable to the absorption spectrum in terms of relative and absolute intensities. While the IR-UV ion dip spectrum is directly comparable to the absorption spectrum in terms of relative intensities it is not true anymore for the absolute intensities. For a proper comparison between absorption and IR-UV ion dip spectra, some parameters have to be taken into account, i.e. intensities of the IR and UV sources and the efficiency of the UV photon absorption. A way to calculate the absorption cross section is applicable to action spectroscopy if we assume that the UV fluence is much smaller than the IR fluence (only way to obtain a correct \( I^{MS-\infty} \) value), as described below:
\[ \sigma(\omega) = \ln\left( \frac{I_{MS}^{+}-0 - I_{MS}^{-\infty}}{I_{MS}(\omega) - I_{MS}^{-\infty}} \right) \frac{1}{\Phi^{IR}} \] (3.2)

with \( \sigma(\omega) \) the absorption cross section for the frequency \( \omega \), \( I_{MS}(\omega) \) the mass spectrometer signal when the IR laser is fixed at the wavelength \( \omega \), \( I_{MS}^{+}-0 \) the mass spectrometer signal when the IR laser is off, \( I_{MS}^{-\infty} \) the mass spectrometer signal when the IR laser is fully saturated and \( \Phi^{IR} \) the fluence of the infrared laser. The full saturation is reached when as many molecules as possible have been depleted from the vibrational ground state. Such equation assumes that the UV excitation of the vibrationally excited molecules will be equally efficient whatever the vibrational mode excited. This point is not discussed further in the paper "Exploring the Mechanism of IR-UV Double-Resonance for Quantitative Spectroscopy of Protonated Polypeptides and Proteins"\(^{115} \) that presents equation 3.2, but we could guess that it is a strong assumption.

Other methods exist that probe vibrational levels for gas phase species, i.e. Raman\(^{46-48} \) spectroscopy or LIF (Laser Induced Fluorescence)\(^ {49} \) but these methods are challenging and therefore limited to small size systems. Note that the rules of selection are different for these methods and therefore non active modes in infrared spectroscopy could be probed here.

### 3.2 IR-UV ion dip spectroscopy principles

IR-UV ion dip spectroscopy has been developed in the 1980's\(^ {116} \). Let us consider a gas phase population of a molecule (or cluster) that could adopt multiple conformations (see subsection 3.3 for the production). A UV laser will be set at the S\( \text{0}\rightarrow\text{S1} \) transition of the selected conformation. This conformation will be ionised via a two photon scheme (see subsection 3.5), and the ion current will be measured with a mass spectrometer (see subsection 3.4). The ion current (signal coming from the mass spectrometer) is constant (or considered as constant if we ignore the non-regularity of the production source). The full process is schematically presented in figure 3.2.

In the IR-UV scheme, the infrared source of light irradiates the gas phase population before the UV irradiation. At this stage, no conformer selective (by UV) has been done. The infrared light will be scanned over the whole range where we want to plot the action spectrum and of course accessible to the lasers used. Whenever the infrared frequency will make the infrared photon resonant with a vibrational transition of the molecular system investigated, a vibrational excitation will occur due to the absorption of the infrared photon. Once the conformation is vibrationally excited, it looses its resonance with the pre-chosen UV photons, see figure 3.2-a, which stops the two photon scheme for ionisation. A depletion of the measured ion current is consequently observed, see figures 3.2-b and -c. The ion current plotted as a function of the infrared frequency is the action spectrum (in IR-UV ion dip spectroscopy), the depletions (dips) correspond to the absorption of the IR photons.

One of the strength of this method is that the spectrum can be plotted for each isomeric conformation, separately, by fixing the UV laser to the corresponding S\( \text{0}\rightarrow\text{S1} \) transition of each isomer.
Figure 3.2: Schemes a-IR-off) and a-IR-on) present respectively the spectroscopic processes happening for the IR sources on and off. For the IR off, the selected conformer B (via the S0→S1 wavelength) is ionised through a two photon scheme. When the IR is on and set at a resonant frequency, we have vibrational excitation for conformer B. This excitation causes conformation B to be off-resonance with the chosen UV photon and no electronic transition nor ionisation therefore occurs. The b-IR-off) and b-IR-on) figures are schematic representations of the set-up. When the IR is off, the conformer B is ionised and enters into the mass spectrometer. When IR is on, no ionisation occurs anymore and the amount of ions in the mass spectrometer is reduced. The c-IR-off) and c-IR-on) figures represent the signal observed on the mass spectrometer. When the infrared light is not entering into the setup, the mass spectrometer signal remains constant in time and does not change with the infrared frequency. When the infrared is on, there are depletions on the mass spectrometer signal for the resonant infrared frequencies.
The absorption of the UV photons is the most restrictive part in this set-up. Indeed, the excited S1 state has to be stable with a lifetime long enough to allow the absorption of the second photon for ionisation. The guanine-cytosine base pair investigated in my work (see section 2.3 of chapter 2) can not be investigated in its canonical Watson-Crick 3D configuration because of the S1 state fastly reconverting to the S0 state through a conical intersection which makes the IR-UV ion dip spectroscopy impossible to be performed on this system. Another requirement is a S0 electronic ground state that efficiently absorbs photons for the S0 → S1 transition. For all these reasons, an aromatic ring is used as chromophore for the UV excitations in our experiments.

For all the systems investigated in this thesis, only one frequency (or color) is used for the two photons ionisation scheme for simplicity reasons (only one UV laser used in the experimental setup). The only exception is salicylic acid, presented in section 2.2 in chapter 2, for which \( \text{EI} > 2^\ast \text{S1} \) (with \( \text{EI} \) the ionisation energy and \( \text{S1} \) the energy of the S1 excited state). A two-color ionization scheme has therefore been used (for which two UV lasers are involved in the setup).

For peptides, the UV chromophore can be either naturally present in the sequence, which is the case whenever the peptide includes the amino acids phenylalanine, tryptophan or tyrosine. It can also be added to the sequence, covalently bound to the peptide as the Z-cap in the Z-Ala\(_6\)-NH\(_2\) peptide that is investigated in our work and where the Z-cap is an aromatic ring, see figures 2.11 and 2.12 in chapter 2. One legitimate question is whether the addition of the Z-cap will influence the conformation of the peptidic system (despite it is not a concern in the context of this thesis). Note that Eric Gloaguen and Michel Mons used a non covalent bound UV-tag (toluene) attached via an NH-π bond that is shown in their paper\(^{117}\), as non-perturbative to the conformation of the studied molecule. The DNA bases and phenol derivatives investigated in this work are natural chromophores too.

Another limitation of the method is observed when the population of one conformation is dominant on the others. When a molecule absorbs an infrared photon, a possibility exists that the system now in its vibrationally excited configuration becomes resonant with the frequency of the UV photons (which are in principle resonant with another conformation). In this case, the molecule is ionised and the ion current increases (while we want only a decrease due to the infrared absorption of the chosen conformer!). Even if this process has few chances to happen, it becomes a real issue if we try to measure the spectrum of a conformation with a small population, because depletion through the normal IR-UV ion dip spectroscopy process (big probability, small population) and "anti-depletion" (small probability and big population) coming from the most populated conformations are now in competition. As example, the Z-Ala-Ala-OMe dipeptide has two populated conformations in our experimental conditions, denoted A and B, see figure 3.3, with A dominant over B (seen through the UV spectroscopy and amplitude of peaks). The infrared spectrum of each conformation is presented in figure 3.4. While the spectrum of conformation A is really well defined, the spectrum of conformation B is definitely less good. We can indeed observe negative peaks corresponding exactly to all
positive peaks of conformation A, illustrating the process described above. The study of this system is recent and therefore not presented further in this thesis.

Figure 3.3: Assigned structures for the two conformations of the dipeptide Z-Ala-Ala-OMe. Conformation A is dominant in terms of population over B in our experimental conditions.

Figure 3.4: Infrared-ultraviolet double resonance spectra of the A and B conformations of Z-Ala-Ala-OMe system. See structures in figure 3.3. The population of the A conformation is larger than the population of the B one. For each absorption peak of A, we observe negative peaks in B spectrum. See text for details.

3.3 Production of cold and isolated neutral species in the gas phase

In the context of this thesis, we want to work on isolated species free of external perturbations. To reach this aim, we will use a molecular beam that provides cold and isolated conditions. In the first step of the experiment (see figure 3.1), one produces cold and isolated conditions for the neutral species of interest. Two different tools have been used in my work to seed molecules into the gas phase, *i.e.* by a heating source or by a desorption laser. Then, the gaseous molecules are driven through a molecular beam by a 3 bar argon flow resulting in cold and isolated conditions. This part of the setup is presented in figure 3.5.
3.3.1 Heating source and desorption laser

In the context of this thesis, we chose to study neutral species only. Molecules are seeded into the gas phase using either a heating source (Ac-Phe-OMe and the phenol derivatives, see sections 2.1.2 and 2.2 of chapter 2) or a desorption laser (for the remaining systems, see chapter 2) depending on the volatility of the sample.

**Heating source**: Only phenol (see section 2.2 in chapter 2) has a sufficient vapor pressure to produce a measurable signal at room temperature. For water-phenol clusters, the argon is flowed through a water reservoir before picking up phenol molecules. For phenol derivatives, see section 2.2 of chapter 2 and Ac-Phe-OMe and (Ac-Phe-OMe)\textsubscript{2}, see section 2.1.2 of chapter 2, the vapor pressure is insufficient, so they are brought into the gas phase using a heated glass container. A temperature of 120°C has been used for the Ac-Phe-OMe powder. All the temperatures used for the phenol derivatives are reported in subsection 3.5. To avoid sample condensation in the valve, it is heated to 5-10 K higher than the heating source.

**The desorption laser** is a Nd:YAG laser emitting a 5 ns pulse energy of about 1.5 mJ at the wavelength 1064 nm. A prepared mixture of the studied molecule powder with carbon black powder is filed on a graphite bar that is heated up by the desorption laser. The sample bar is driven by a motor to constantly expose fresh part of the mixture. The power of the laser has however to be set up carefully to avoid fragmentation of the molecules that could occur at high power.

Note that the production of large neutral species in the gas phase is challenging while producing charged systems is much easier especially thanks to methods like ESI (ElectroSpray Ionisation)\textsuperscript{118}. IR-UV ion dip experiments on charged systems are performed by the group of Prof. Rizzo for example\textsuperscript{90}. 

---

Figure 3.5: Pictures of the experimental setup in the configuration using the desorption laser. The sample is visible but not the desorption laser that is located above the setup. The red arrow visible in the side view (right panel) corresponds to the path of the desorption laser beamlight. On the side view (right panel), the skimmer is visible, see subsection 3.3.2. These pictures are part of the schematic representation presented in figure 3.1.
3.3.2 Molecular beam

A schematic view of the molecular beam is presented in figure 3.6. Once the molecular systems are produced in the gas phase they are driven through the skimmer by a 3 bar Argon flow. The gas expands into the vacuum chamber at pressure $P_1$ where it is released by the pulsed valve (that is not presented in the scheme but corresponds to the junction between the $P_0$ pressure domain and the $P_1$ pressure domain). Then the central part of the expansion passes the skimmer, collimating the molecular beam.

Figure 3.6: A schematic representation of a molecular beam. A gas in the reservoir with temperature and pressure $T_0$ and $P_0$ respectively can escape into a chamber with a much lower pressure ($P_1$). In this process random motion is converted into directed motion, resulting in lowering of the temperature. The skimmer is used to select the coldest (most unidirectional) part of the beam (Reproduced from the thesis "Protein folding forces probed by infrared action spectroscopy" by S. Jaeqx, The Netherlands, 2014).

In the molecular beam, the first physical process is the reduction of the rotational and vibrational energies of the sample by collision with the argon atoms (or He atoms). Then, the gas flow is driven through the skimmer (jet expansion) and only molecules with identical axial direction will remain, reducing the number of collisions and ensuring isolated conditions. This provides low temperature systems and vibrational and rotational ground state.

Note that for the catechol molecule, presented in section 2.2 of chapter 2, a helium gas flow has been used as seeding gas. In general, helium increases the signal measured on the mass spectrometer because it decreases the chance to form a molecule-atom cluster (smaller polarisability). But on the other hand, the collisional cooling is less efficient. With catechol, the cooling by helium atoms is good enough and therefore used to increase the signal measured on the mass spectrometer.

The typical pressures in our experiments are $2.5 \times 10^{-5}$ mbar, $7.2 \times 10^{-7}$ mbar and $5.0 \times 10^{-8}$ mbar, respectively for the source chamber, the ionisation chamber and detection chamber.
3.4 Detection of the ionised systems - Reflectron time-of-flight mass spectrometer

Once the neutral molecules have been seeded into the gas phase, the second step is to ionise them and to measure the amount of ions for the final action spectroscopy. The ionisation is done by a dye laser pumped by a YAG laser following a two photons process. The ions are then driven and measured in a reflectron time-of-flight mass spectrometer.

The ion current is detected with a TOF (time of flight) mass spectrometer schematically presented in figure 3.7. In practice, the molecules become ionized inbetween two charged plates with a strong voltage gap, VA1 and VA2. The acceleration will be different for ions with different charges $q$, and masses $m$, and the needed time to go through the spectrometer will be function of the $\frac{m}{q}$ ratio.

Figure 3.7: Schematic representation of the reflectron time-of-flight spectrometer. The ions are created between the repeller (VA1) and extractor (VA2) plates, accelerated into the time-of-flight tube, and reflected into the detector (Reproduced from the thesis "Protein folding forces probed by infrared action spectroscopy", by S. Jaeqx, The Netherlands, 2014).

In practice, there are two limitations, both due to the spatial expansion of the ions in the gas phase.

- The first one is that all the ions will not be at equal distance of the two charged plates, inducing different accelerations for ions with the same ratio $\frac{m}{q}$. This problem is solved by doubling the length of the mass spectrometer (by adding the "VR1" charged plates at the end of the mass spectrometer, see figure 3.7). The charged plates at the end of the mass spectrometer will decelerate and accelerate the ions in the opposite direction function of the $\frac{m}{q} \text{ ratio} \text{ but also } \text{function of the incident velocity of the ions. It is thus possible to choose settings that will correct for the original error.}

- The second limitation arises from the time expansion of the ions packet due to the length of the UV pulse (which means that the UV pulse is longer in time than the valve opening time). All ions will therefore not arrive at the same time between the plates. By doubling
the path length, the mass spectrum resolution increases (the ions with the same \( \frac{m}{q} \) are gathered together). It is the biggest limitation in the set-up, decreasing the resolution of the mass spectrometer.

3.5 REMPI (Resonance Enhanced Multi-Photon Ionization)

As it has been discussed above and in subsection 3.2, the molecules need to be ionized to be measured by the mass spectrometer (this is the 'action' part in our action spectroscopy). For the best ionization efficiency and to be sure to have only one dominant ionized conformation over the others, only the wavelength corresponding to the \( \text{S}_0 \rightarrow \text{S}_1 \) transition of one specific conformation is used. We therefore have to measure the UV spectrum of the molecular system before anything else. The UV spectrum is actually measured through the signal of the mass spectrometer, corresponding to the ionisation occurring via a two-photon process. The UV spectrum is therefore called REMPI (Resonance Enhanced Multi Photon Ionization) spectrum.

In our experiment, the UV beam is produced by a frequency doubled dye laser pumped by a frequency double or tripled YAG laser (355 or 532 nm, originally 1064 nm). The dye coumarin 153 pumped by the frequency tripled YAG laser is used when the chromophore is the phenylalanine amino-acid or for phenol derivatives. For DNA nucleobases and analogues, a mixture of Rhodamine 610 and Rhodamine 640 is used. The frequency of the YAG laser is only doubled to pump the Rhodamine dye (532 nm).

All the REMPI spectra of all molecules in my work have been measured during my PhD, only two of them are reported here for illustration: the REMPI spectrum of the Ac-Phe-Pro-NH\(_2\) system and the ones of the Ac-Phe-OMe and (Ac-Phe-OMe)\(_2\) systems are respectively presented in figures 3.8 and 3.9. The other REMPI spectra can be found elsewhere:

- **For the dipeptides series**, presented in section 2.1.1 in chapter 2 and studied in chapters 5, 6, 9 and 10, the REMPI spectrum of the Ac-Phe-Val-NH\(_2\) dipeptide has been published in the supplementary material of the paper: "Mapping gas phase dipeptides motions in the far-infrared and terahertz domain"\(^{65}\), the REMPI spectra of the Ac-Phe-Cys-NH\(_2\) and Ac-Phe-Ser-NH\(_2\) dipeptides have been published in the paper: "A conformation-selective IR-UV study of the dipeptides Ac-Phe-Ser-NH\(_2\) and Ac-Phe-Cys-NH\(_2\): probing the SH···O and OH···O hydrogen bond interactions".\(^{62}\), the REMPI spectrum of the Ac-Phe-Pro-NH\(_2\) system has been published in the paper "Can far-IR action spectroscopy combined with BOMB simulations be conformation selective"\(^{64}\) presented in chapter 5. The \( \text{S}_0 \rightarrow \text{S}_1 \) transitions are at 37486 cm\(^{-1}\) for Ac-Phe-Gly-NH\(_2\) (\( \gamma \)-turn), 37464 cm\(^{-1}\) for Ac-Phe-Ala-NH\(_2\) (\( \gamma \)-turn), at 37435.5 cm\(^{-1}\) and 37409 cm\(^{-1}\) for the \( \gamma \) and \( \beta \)-turn conformations of Ac-Phe-Pro-NH\(_2\) respectively, at 37390 cm\(^{-1}\) for Ac-Phe-Ser-NH\(_2\) (\( \gamma \)-turn), at 37325 cm\(^{-1}\) and 37450 cm\(^{-1}\) for the \( \gamma \) and \( \beta \)-turn conformations of Ac-Phe-Cys-NH\(_2\) respectively, 37472 cm\(^{-1}\) for the A1 conformation of Ac-Phe-Val-NH\(_2\) (\( \gamma \)-turn) and 37410 cm\(^{-1}\) for the A2 conformation of Ac-Phe-Val-NH\(_2\) (\( \gamma \)-turn). The values of these transitions can also be found elsewhere in the literature\(^{61-63}\).
For Ac-Phe-OMe and (Ac-Phe-OMe)$_2$ presented in section 2.1.2 in chapter 2 and studied in chapters 9, 10 and 11, the REMPI spectra are presented in figure 3.9, and the $S_0 \rightarrow S_1$ transitions are found at 37564 cm$^{-1}$ for the monomer and at 37539 cm$^{-1}$ for the dimer (only one isomer found). Investigation not published yet.

Figure 3.8: REMPI spectrum of the Ac-Phe-Pro-NH$_2$ dipeptide, recorded in this work.

Figure 3.9: REMPI spectra of the Ac-Phe-OMe monomer (in purple) and its dimeric form (Ac-Phe-OMe)$_2$ (in green), recorded in this work.
• **For Z-Ala$_6$-NH$_2$** presented in section 2.1.3 in chapter 2 and studied in chapter 9, the REMPI spectrum is presented in Sander Jaeqx’s thesis: "Protein folding forces probed by infrared action spectroscopy". Investigation not published yet.

• **For the phenol derivatives**, presented in section 2.2 in chapter 2 and studied in chapters 7, 8, 9 and 10 we have multiple series presented:

  - For the first series of phenol derivatives introduced in the paper: "Anharmonic, dynamic and functional level effects in far-infrared spectroscopy: phenol derivatives"$^{108}$, *i.e.* phenol, catechol, salicylic acid, saligenin and nitrophenol, only the REMPI spectrum of the nitrophenol can be found in the supplementary material of the paper$^{108}$, presented in chapter 7. The other REMPI spectra are not presented. The $S_0 \rightarrow S_1$ transitions are listed in figure 3.10. The transition is located at 36348.5 cm$^{-1}$ for phenol, 36346 cm$^{-1}$ for phenol-D (identical values were found elsewhere for the phenol and its isotopologue$^{119}$), 35648 cm$^{-1}$ for catechol (as already found elsewhere$^{120}$), 30940 and 32098 cm$^{-1}$ for the two observed conformations of the salicylic acid (as already found elsewhere$^{113}$), 35503 cm$^{-1}$ for saligenin (as already found elsewhere$^{112}$), 35502.5 and 35509.5 cm$^{-1}$ for the two isotopologues of saligenin and 35940 cm$^{-1}$ for the nitrophenol.

  - For the second series (bare saligenin and saligenin-water clusters), introduced in the paper: "Fingerprints of Inter- and Intramolecular Hydrogen Bonding in Saligenin-Water Clusters Revealed by Mid- and Far-Infrared Spectroscopy"$^{111}$, all the REMPI spectra can be found in the supplementary material of the paper, see chapter 8. The values of the $S_0 \rightarrow S_1$ transitions can be found in the main text and are at 35676 cm$^{-1}$ for SLG-1w, 35650 cm$^{-1}$ for SLG-2w, 35977 cm$^{-1}$ for SLG-3w-a and 35855 cm$^{-1}$ for SLG-3w-b.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$T_{\text{source}}$ (°C)</th>
<th>$\lambda_{\text{UV}}$ (cm$^{-1}$)</th>
<th>Seed gas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenol</td>
<td>25</td>
<td>36348.5</td>
<td>Ar</td>
</tr>
<tr>
<td>Phenol-D</td>
<td>25</td>
<td>36346</td>
<td>Ar</td>
</tr>
<tr>
<td>Catechol</td>
<td>90</td>
<td>35648</td>
<td>He</td>
</tr>
<tr>
<td>Salicylic acid – rotamer 1</td>
<td>90</td>
<td>30940</td>
<td>He</td>
</tr>
<tr>
<td>Salicylic acid – rotamer 2</td>
<td>90</td>
<td>32098</td>
<td>He</td>
</tr>
<tr>
<td>Saligenin</td>
<td>80</td>
<td>35503</td>
<td>Ar</td>
</tr>
<tr>
<td>Saligenin-FD</td>
<td>80</td>
<td>35502.5</td>
<td>Ar</td>
</tr>
<tr>
<td>Saligenin-BD</td>
<td>80</td>
<td>35509.5</td>
<td>Ar</td>
</tr>
<tr>
<td>Nitrophenol</td>
<td>25</td>
<td>35940</td>
<td>He</td>
</tr>
</tbody>
</table>

Figure 3.10: Experimental parameters for our experiments on phenol derivatives.
- For DNA base pairs, presented in section 2.3 in chapter 2 and studied in chapters 9 and 10, the REMPI spectra are presented in the PhD manuscript "Excited State Dynamics of Isolated Nucleobases and Base Pairs" by Faady Mohammad Siouri (University of Santa Barbara, USA, 2017) and the frequency of the $S_0 \rightarrow S_1$ transition is $33088 \text{ cm}^{-1}$ for the studied conformer of the guanine-guanine, $33292 \text{ cm}^{-1}$ for studied the conformer of guanine-cytosine and $33071 \text{ cm}^{-1}$ for the 9-ethylguanine···1-methylcytosine complex.

Once the REMPI spectrum has been measured, one needs to assign groups of peaks to the different conformers. Multiple strategies can be chosen (UV-UV holeburning, IR-UV holeburning or an infrared spectrum can be measured for each UV transition). For the systems studied in this thesis (presented in chapter 2), all REMPI spectra were available in the literature. The published assignments have been used.

The REMPI spectrum can also be used for temperature evaluation through hot bands. These correspond to vibrational or rotational excited levels being populated and a transition $\nu=1 \rightarrow \nu=0$ can be observed, corresponding to a vibrational de-excitation during the electronic excitation process, with the associated band red-shifted with respect to the $\nu=0 \rightarrow \nu'=0$ band. Since we know the energy of the excited level, a temperature of $\sim 10-20\text{K}$ has been estimated from previous studies, and also assumed for systems investigated in this work.$^{58}$

![Figure 3.11](image.png)

Figure 3.11: Representation of a $\nu=0 \rightarrow \nu'=0$ fundamental transition and of a $\nu=1 \rightarrow \nu'=0$ hot band transition. The two surfaces correspond to electronic potential energy surfaces respectively of the $S_0$ ground state (at the bottom) and $S_1$ excited surface (at the top).
намечено использование нескольких источников в сочетании с IR-UV ионным допplerским установкой, представленной в этом разделе. Для диапазона 3000-4000 см\(^{-1}\), YAG-выкачивающаяся управляемый IR OPO лазер с KTA кристаллами является наиболее распространенным инфракрасным источником. Для диапазона 1000-2000 см\(^{-1}\), Free Electron Laser инфраустановки являются лучшим инструментом, доступным, например, в Orsay в Франции (CLIO), в Берлине в Германии (Fritz Haber Institute) или в Неймегене в Нидерландах (FELIX).

В контексте этой диссертации, мы специально интересованы в диапазоне далекого инфракрасного спектра и используем Free Electron Lasers FELIX. На данный момент в Европе три FEL установки расположены в Неймегене, Нидерланды (FELIX), в Орсе, Франция (CLIO) и в Берлине, Германия (Fritz Haber Institute), два из которых (FELIX и Fritz Haber Institute) функционируют в далеко инфракрасном/THz диапазоне.

Одной из причин очень ограниченного количества газовых исследований в далеком инфракрасном диапазоне является отсутствие интенсивных источников в этом частотном диапазоне. Некоторые эксперименты были проведены в далеком инфракрасном диапазоне с помощью источников, получаемых в результате голографии (черное тело-тип источники, см. например, исследование от Jean-Marie Flaud\(^{[121]}\)) но эти источники далеки от того, чтобы быть интенсивными достаточно, чтобы проводить эксперименты в молекулярных пучках условиях. Только недавно, Free Electron Lasers начали производить свет в далеком инфракрасном диапазоне. Все наши эксперименты были выполнены на лаборатории FELIX, в сотрудничестве с командой Dr A. M. Rijs. Три FEL (Free Electron Laser) доступны в лаборатории FELIX, (FELIX, FELICE and FLARE, см. рисунок 3.12). Здесь, только FELIX (1 и 2) был использован. Этот лазер способен производить свет до 90 см\(^{-1}\), в то время как на момент написания.

![Figure 3.12: Overview of the FELIX laser hall. This figure has been copied from the official website of the FELIX laboratory: http://www.ru.nl/felix/about-felix/about-felix/](http://www.ru.nl/felix/about-felix/about-felix/)
The principles of the free electron laser are illustrated in figure 3.13. Electrons are injected into vacuum with an electron gun and accelerated with electro-magnetic fields at speed close to the speed of light. The electrons enter into an undulator which consists of a series of positive and negative magnets. The electrons will oscillate at a specific frequency tuned by changing the distance between the rows of magnets and they therefore produce light. The emitted light is not intense enough after one single pass through the undulator. In practice the light beam is trapped between two high-reflective gold plated copper mirrors and is reflected back and forth in the cavity. The length of the cavity is adjusted for the beam light to overlap with fresh packets of electrons. By interacting with the light, the electrons will produce coherent and intense radiations at the same frequency. Part of this radiation is released through a central hole in a mirror and transported through a beam transport system to the IR-UV ion dip experimental setup. All the beam light path is kept at low pressure to avoid light absorption by water vapour naturally present in the air.

Macropulses of FELs typically have a time-duration of 6 to 10 $\mu$s. The wavelength-dependent intensity of the infrared light of FELIX varies from 20 mJ per macropulse to 80 mJ per macropulse.

Note that the UV beam is produced by a frequency doubled dye laser pumped by a frequency doubled or tripled YAG table laser, see subsection 3.5.

During the experiment, IR-on and IR-off signals are alternated to correct for the long term UV power drifts and the evolution of the source conditions by operating the IR laser at 5Hz and the UV laser at 10Hz.

Part of the FELIX beam light is driven to a calibrated home-built grating spectrometer and to a calibrated infrared powermeter. Since the experiments are performed over a very wide frequency range, the IR intensities of the measured transitions are corrected for the FELIX macropulse wavelength dependent energy and for the number of photons present in the FELIX macropulse at different wavelengths for identical pulse energy.
Chapter 4

DFT-MD theoretical method for anharmonic vibrational spectroscopy (and alternatives).

Summary: I present in a nutshell the DFT-MD (Density Functional Theory based Molecular Dynamics) method used in my thesis. In our dynamics, the nuclei are treated classically and the electrons quantum mechanically within the DFT formalism. This method is implemented in the CP2K package, used throughout all my work. We apply the Born-Oppenheimer framework.

The BLYP-D3 functional has been used for all the results presented here except whenever explicitly specified. This functional was selected based on previous works in the group on gas-phase vibrational spectroscopy over different spectral ranges. The simulations are based on the GPW (Gaussian Plane Waves) method, with a dual basis set representation constructed on a plane wave basis set with a 450 Ry energy cutoff and the aug-TZV2P (for the peptidic systems) or m-TZV2PX (for the phenol derivatives and the DNA base pairs) gaussian basis sets. Pseudopotentials of the GTH type (Goedecker-Teter-Hutter) are used. The size of the cubic box is systematically optimised for each family of system to ensure that there is no interaction between the system and its 3D replica (periodic boundary conditions are indeed systematically used in our simulations). Time-step for the dynamics is 0.4 fs. DFT-MD are performed at 50 K for the best spectral agreement between theory and experiment. Initial structures for the dynamics are re-optimised at our level of theory. Each DFT-MD infrared spectrum of peptidic systems is obtained as an average over three separate trajectories. We use the same initial geometry for all trajectories (the optimised one), different initial velocities (Boltzmann distribution), 20 ps trajectory each (after thermalisation over 1 ps) in order to obtain converged relative intensities of the absorption peaks. Peptidic systems spectra are therefore calculated over a total 60 ps trajectory except for Z-Ala$_6$-NH$_2$ for which only one trajectory of 14 ps has been used. For phenol derivatives and DNA base pairs only one trajectory of 20 ps has been used (due to a lack of time). See chapter 2 for details on the systems simulated. For the purpose of functional comparison, one trajectory of 20 ps has been run using the B3LYP-D3 hybrid GGA functional for each of the four phenol derivatives.
For vibrational spectroscopy, in MD simulations, the infrared spectrum is calculated as the Fourier transform of the time correlation function of the fluctuating dipole moment vector of the absorbing molecular system:

$$I(\omega) = \frac{2\pi\beta\omega^2}{3cV} \int_{-\infty}^{\infty} \langle \delta \vec{\mu}(t) \cdot \delta \vec{\mu}(0) \rangle e^{i\omega t} dt$$

where $\beta = 1/kT$, $T$ is the average temperature of the trajectory, $c$ is the speed of light in vacuum, $V$ is the volume of the simulation box. The angular brackets represent a statistical average of the correlation function of the molecular dipole moment vector $\vec{\mu}(t)$, where $\delta \vec{\mu}(t) = \vec{\mu}(t) - \langle \vec{\mu} \rangle$ is the dipole fluctuation, with $\langle \vec{\mu} \rangle$ the time average of $\vec{\mu}(t)$. $\frac{2\pi\beta\omega^2}{3cV}$ is a quantum correction factor related to the classical treatment of the nuclei. Our trajectories are 20 ps long, which means that a mode at 80 cm$^{-1}$ is sampled $\approx$50 times. It is therefore difficult to trust our DFT-MD spectra below 80 cm$^{-1}$.

This method takes into account modes couplings and anharmonicities of the potential energy and dipole surfaces (by opposition to the harmonic method for which the description of the potential and of the dipole is restricted to a harmonic function). Vibrational-rotational coupling is removed within the Eckart-Sayvetz conditions whenever needed. This method provides anharmonic vibrational spectra as will be discussed in this chapter.

In section 4.1, I introduce DFT as the electronic representation (basic principles of the DFT in subsection 4.1.1 as well as the BLYP-D3 and B3LYP-D3 functionals in subsection 4.1.2) systematically used in my DFT-MD simulations, its implementation in the CP2K package, i.e. the GPW, Gaussian Plane Waves scheme in subsection 4.1.3, the specific equations implemented in CP2K in subsection 4.1.4, the SCF procedure in section 4.1.5 as well as the specificities for the calculation using hybrid functionals as B3LYP-D3 in section 4.1.6.

In section 4.2, I go into more details about DFT-MD simulations (Density Functional Theory based Molecular Dynamics). Section 4.3 presents more in details the way to calculate DFT-MD spectra and vibrational modes analyses in DFT-MD.

We also need geometry optimisations, this is presented in section 4.4.

Harmonic vibrational spectra have also been calculated in my work using the Gaussian package. In these calculations, the BLYP-D3 electronic representation is applied for the best comparison with the infrared DFT-MD spectra. The gaussian basis set 6-311+G(d,p) is used. An ultrafine grid is used and the "tight" keyword is used for geometry optimisations that change the criteria of convergence from the default in the Gaussian package. This method is introduced in section 4.5.

Finally, anharmonic VPT2 infrared spectra (as implemented in the Gaussian package by the group of Barone) have also been calculated. For these, the BLYP-D3 electronic representation is also kept (for the sake of comparison with DFT-MD), the gaussian basis set 6-311+G(d,p) is used, an ultrafine grid and "tight" keyword are used for geometry optimisations. VPT2 spectroscopy is presented in section 4.6.
4.1 Electronic representation: DFT (Density Functional Theory)

4.1.1 Basic definitions

In quantum mechanics, an ensemble of particles, i.e. M nuclei and N electrons, is represented by the time independent Hamiltonian \( H \). \( H \) is an operator, sum of the kinetic and potential energies of these \((M+N)\) particles. The Hamiltonian is expressed as a function of the set of 3N cartesian coordinates for the electrons \( \{ \vec{r}_i \} = \{ \vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N \} \) and a set of 3M cartesian coordinates for the nuclei \( \{ \vec{R}_j \} = \{ \vec{R}_1, \vec{R}_2, \ldots, \vec{R}_M \} \) (as projected into the \( \vec{r} \) space):

\[
H_{Tot}(\{ \vec{r}_i \}, \{ \vec{R}_j \}) = T_N + T_e + V_{NN} + V_{ee} + V_{Ne}
\]  

(4.2)

In this expression, \( T_N \) and \( T_e \) are the kinetic energies for the nuclei and electrons respectively, \( V_{NN}, V_{ee} \) and \( V_{Ne} \) are the potential energies for the nuclei-nuclei interactions, the electron-electron interactions and nuclei-electron interactions respectively. We therefore have:

\[
H_{Tot}(\{ \vec{r}_i \}, \{ \vec{R}_j \}) = -\frac{1}{2} \sum_{A} M \frac{\hbar^2}{2m_A} \nabla_A^2 - \frac{1}{2} \sum_{i} N \frac{\hbar^2}{2m_e} \nabla_i^2 + \frac{1}{2} \sum_{A} \sum_{B \neq A} Z_A Z_B \frac{e^2}{| \vec{R}_A - \vec{R}_B |} + \frac{1}{2} \sum_{i} \sum_{j \neq i} \frac{e^2}{| \vec{r}_j - \vec{r}_i |} - \frac{1}{2} \sum_{A} \sum_{i} \frac{Z_A e}{| \vec{R}_A - \vec{r}_i |}
\]

with \( T_N = -\frac{1}{2} \sum_{A} M \frac{\hbar^2}{2m_A} \nabla_A^2 \), \( T_e = -\frac{1}{2} \sum_{i} N \frac{\hbar^2}{2m_e} \nabla_i^2 \), \( V_{NN} = \frac{1}{2} \sum_{A} \sum_{B \neq A} Z_A Z_B \frac{e^2}{R_{AB}} \), \( V_{ee} = \frac{1}{2} \sum_{i} \sum_{j \neq i} \frac{e^2}{| r_{j} - r_{i} |} \) and \( V_{Ne} = -\frac{1}{2} \sum_{A} \sum_{i} \frac{Z_A e}{| R_{A} - r_{i} |} \)

where \( M \) is the number of atoms, \( N \) the number of electrons, \( m_e \) the mass of one electron, \( m_A \) the mass of the nucleus \( A \), \( Z_A \) is the atomic number (and the charge) of nucleus \( A \), \( e \) the charge of an electron, \( | \vec{R}_A - \vec{R}_B | \) is the internuclear distance between two nuclei \( A \) and \( B \), \( | \vec{R}_A - \vec{r}_i | \) the distance between nucleus \( A \) and electron \( i \), and \( | \vec{r}_j - \vec{r}_i | \) the distance between the electrons \( i \) and \( j \).

Once the Hamiltonian has been defined, one can solve the Schrödinger equation, a mathematical equation named after Erwin Schrödinger, who derived and published the equation in 1926. We will in my work only consider the time-independent Schrödinger equation that describes the physical system and allows to calculate its energy \( E_{Tot} \) and wavefunction \( \Psi_{Tot} \):

\[
H_{Tot} \Psi_{Tot}(\{ \vec{r}_i \}, \{ \vec{R}_j \}) = E_{Tot} \Psi_{Tot}(\{ \vec{r}_i \}, \{ \vec{R}_j \})
\]

(4.4)

The time independent wavefunction \( \Psi_{Tot}(\{ \vec{r}_i \}, \{ \vec{R}_j \}) \) is a mathematical description of the quantum state of the system. The wavefunction is a function of all degrees of freedom of the system:

\[
\Psi(\{ \vec{r}_i \}, \{ \vec{R}_j \}) = \Psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N, \vec{R}_1, \vec{R}_2, \ldots, \vec{R}_M)
\]

(4.5)

This is the total wavefunction of the whole system (nuclei + electrons), and the total energy of the system is quantised and both are solutions of the Schrödinger equation (see equation 4.4).
The construction of the wavefunction will be described in subsection 4.1.3 as well as the approximations to the Hamiltonian in the form of the DFT representation in subsection 4.1.2.

The mass of an electron ($\approx 9.1 \times 10^{-31}$ kg) is approximately 1686 times smaller than the mass of a proton ($\approx 1.7 \times 10^{-27}$) which means that even for the hydrogen atom (with the lightest nucleus), there is a difference of 3 orders of magnitude between the masses of the electron and the nucleus, that therefore induces differences of speed in their respective motions. In the Born-Oppenheimer approximation we consider that the motions of the electrons will be so fast in comparison with nuclei motions that we can consider the nuclei as "fixed in space", and therefore only the electronic Schrödinger equation has to be solved (at fixed positions of the nuclei, represented by the ";" in the following equation).

In the Born-Oppenheimer framework, equation 4.5 now becomes:

$$
\Psi_e(\{r\}, \{R\}) = \Psi(\vec{r}_i, \vec{r}_j, \vec{r}_k, ..., \vec{r}_N; \vec{R}_1, \vec{R}_2, ..., \vec{R}_M) = \Psi(\vec{r}_i, \vec{r}_j, \vec{r}_k, ..., \vec{r}_N; \vec{R})
$$

which is now the electronic wavefunction of the system, explicitly dependent on the 3N cartesian coordinates of the electrons and implicitly dependant on the 3M coordinate of the nuclei.

The Schrödinger equation (equation 4.4) is now re-expressed only for the electrons as:

$$
H_e \Psi_e(\{r\}; \{R\}) = E_e \Psi_e(\{r\}; \{R\})
$$

where $H_e$ is the electronic Hamiltonian:

$$
H_e(\{r\}; \{R\}) = -\frac{1}{2} \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 + \frac{1}{2} \sum_i \sum_{j \neq i} \frac{e^2}{|\vec{r}_j - \vec{r}_i|} - \frac{1}{2} \sum_A \sum_i \frac{Z_A e}{|\vec{R}_A - \vec{r}_i|}
$$

and $E_e$ is the electronic energy. In all my work, I will consider nuclei as classical particles, therefore I will forget the Schrödinger equation for the nuclei and never solve it. In the following, $H_e$ is replaced by $H$, and $E_e$ by $E$.

In practice the analytical solution to the electronic Schrödinger equation can be only obtained for the hydrogen atom and any hydrogenoid atom as well as the $H_2^+$ molecule. In practice, one therefore has to apply approximations in order to obtain approximate values of the wavefunction and energy. There exists a large variety of ways that correspond to different levels of approximations and different computational costs. One can cite the Hartree-Fock or post Hartree-Fock methods (Møller Plesset perturbation theory\textsuperscript{138}, coupled cluster\textsuperscript{139}, ...) based on the wavefunction only, the DFT method that will be described in details here, in its modern form based on both the wavefunction and on the electronic density, or semi-empirical methods (AM1\textsuperscript{140}, PM3\textsuperscript{141}, PM6\textsuperscript{142}, PM7\textsuperscript{143}, ...). Note that the cheapest way available to calculate the energy of the system is to use a force field (AMBER\textsuperscript{53}, CHARMM\textsuperscript{54}, ...) because it does not take into account the electrons anymore (not a quantum method). None of these methods will be described in this thesis. Only DFT will be discussed.
All the quantum methods above (DFT, HF, ...) are based on the use of one single Slater determinant:

\[
\Psi(\{\vec{r}\}; \{\vec{R}\}) = \Psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N; \{\vec{R}\}) = \frac{1}{\sqrt{N}} \begin{bmatrix}
\psi_i(\vec{r}_1) & \psi_j(\vec{r}_1) & \ldots & \psi_N(\vec{r}_1) \\
\psi_i(\vec{r}_j) & \psi_j(\vec{r}_j) & \ldots & \psi_N(\vec{r}_j) \\
\vdots & \vdots & \ddots & \vdots \\
\psi_i(\vec{r}_N) & \psi_j(\vec{r}_N) & \ldots & \psi_N(\vec{r}_N)
\end{bmatrix}
\]  

(4.9)

with \(\psi_i(\vec{r}_j)\) the spinorbital (mono electronic wavefunction) number \(i\) populated by the electron labelled \(j\). I have removed any notation related to spins for convenience of writing. The spinorbitals are expressed on a basis set, which will be described later in this chapter.

For molecular ground states which are quasi-degenerate with low-lying excited states or in bond breaking situations, HF or post Hartree-Fock methods using a single Slater determinant might not be suited and multi-configurational methods are needed to correctly represent the electronic wavefunction (e.g. MCSCF\textsuperscript{144}, MR-CI\textsuperscript{145}). These methods use a linear combination of Slater determinants. The DFT method uses one single Slater determinant. Because of the large amount of functionals developed in the literature, with some of them fitted on high level multi-configurational post Hartree Fock methods, the mono determinant character can be somehow empirically corrected.

The DFT approach is extensively used nowadays because of the good balance between computational cost and accuracy. The main reason of the cheap computational cost is that we replace the 3N cartesian degrees of freedom (with N the number of electrons in \(\Psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N; \{\vec{R}\})\)) typically needed in a Hartree-Fock or post-Hartree-Fock method, by 3 cartesian degrees of freedom that correspond to the 3D representation in space necessary to represent the electronic density in the DFT approach. While the wavefunction is described using mathematical functions (as gaussian or plane wave functions), the density is described on a 3D grid with \(\rho(\vec{r})\) the density value at each point mesh \(\vec{r}\) of the grid.

### 4.1.2 Brief history of the DFT methodology and presentation of the BLYP-D3 density functional, and beyond.

In the following, a brief history of the DFT method is presented as well as the BLYP-D3 functional which is the main representation used in the context of this thesis.

**Hohenberg-Kohn equations.**

In the seminal work of Hohenberg and Kohn\textsuperscript{146} that introduced the Density Functional Theory in 1964, the electronic energy of the system is calculated with the following equation:

\[
E[\rho] = T[\rho] + E_{\text{ext}}[\rho] + E_H[\rho] + E_{\text{HK}}^{\text{XC}}[\rho]
\]  

(4.10)

with \(\rho(\vec{r})\) the electronic density at \(\vec{r}\) in space. To get the total energy of the system, one has to add up \(V_{NN}\) (see equations 4.2 and 4.3) to \(E[\rho]\). In equation 4.10, \(T[\rho]\), \(E_{\text{ext}}[\rho]\) and
are respectively the kinetic energy of the electrons, the external energy due to the nuclei interacting with the electrons (this is replacing $V_{Ne}$ in equation 4.2) and the Hartree energy (electron-electron direct Coulomb repulsion). $E_{HC}^{K} [\rho]$ is the exchange-correlation energy functional.

In details:

$$E[\rho] = T[\rho] + \int \rho(\vec{r}) V_{ext}(\vec{r}) d\vec{r} + \frac{1}{2} \int \int \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + E_{HC}^{K} [\rho]$$ (4.11)

The problem at this point is that we do not know how to calculate the kinetic energy functional $T[\rho]$ (the exact expression is unknown for interacting electrons, only known for non interacting electrons) neither the exchange-correlation part $E_{HC}^{K} [\rho]$ (note that the exchange part is exact only in the Hartree-Fock expression).

Kohn-Sham equations.

A solution was given by the Kohn-Sham equations in 1965, introducing at the same time the modern DFT equations still used nowadays. The new equations are based both on the spinorbitals $\psi_i(\vec{r})$ (mono-electronic wavefunctions) and electronic density $\rho(\vec{r})$. We indeed know how to calculate the exact kinetic energy for non interacting electrons based on the mono-electronic wavefunctions:

$$T_{\text{Non--interacting}} = - \frac{1}{2} \sum_i \langle \psi_i(\vec{r}) | \nabla^2 | \psi_i(\vec{r}) \rangle$$ (4.12)

To take into account the interactions between the electrons, a correction has however to be applied. The form of this correction is basically unknown, and in modern DFT, this correction will be part of the exchange-correlation functional where all the unknown quantities are gathered.

For non interacting electrons, the electronic density $\rho(\vec{r})$ and the mono-electronic wavefunctions $\psi_i$ are linked together via the following equation:

$$\rho(\vec{r}) = \sum_i |\psi_i(\vec{r})|^2$$ (4.13)

Using this density (of non interacting electrons) and the associated kinetic energy from eq. 4.12, we now have the following expression for the electronic energy of our N electrons system:

$$E[\rho] = T + E_{ext}[\rho] + E_{H}[\rho] + E_{HC}^{K} [\rho]$$ (4.14)

$$= T_{\text{Non--interacting}} + E_{ext}[\rho] + E_{H}[\rho] + (E_{HC}^{K} [\rho] + T_{\text{correction}})$$ (4.15)

$$= - \frac{1}{2} \sum_i \langle \psi_i(\vec{r}) | \nabla^2 | \psi_i(\vec{r}) \rangle + \int \rho(\vec{r}) V_{ext}(\vec{r}) d\vec{r}$$

$$+ \frac{1}{2} \int \int \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + (E_{HC}^{K} [\rho] + T_{\text{correction}})$$ (4.16)
In equation 4.16, \( E_{HK}^{XC}[\rho] \) remains an unknown expression and \( T_{\text{correction}} \) is added up to this unknown. The total is still called \( E_{XC}[\rho] \) functional for exchange and correlation. This is where DFT becomes empirical: several expressions exist in the literature that give rise to different DFT functionals. Research in DFT consists in particular in the development of new forms and parameters for the best exchange correlation functional in terms of accuracy and computational price. Hereafter, I describe briefly some approximations proposed over time for the exchange correlation functional. All other terms in equation 4.16 are "easy" to calculate.

**Local Density Approximation (LDA)**

The LDA approach gathers one of the first and simplest approximations for the XC (exchange-correlation) functional. For convenience the exchange-correlation functional was decomposed into an exchange \( E_X[\rho] \) and a correlation \( E_C[\rho] \) term:

\[
E_{XC}[\rho] = E_X[\rho] + E_C[\rho]
\]  

The approximation for the XC in LDA relies only upon the density. The approach is based on the homogeneous electron gas equations. This is constructed by placing \( N \) interacting electrons into a volume \( V \) with a positive background charge keeping the system neutral. In particular, for an homogeneous density in space, the exchange energy density is proportional to \( \rho^{1/3} \). The exchange energy density of a homogeneous electron gas is known analytically. The LDA for exchange employs this expression under the approximation that the exchange energy in a system where the density is not homogeneous, is obtained by applying the homogenous electron gas results pointwise, yielding the expression 4.18. The exchange functional in LDA comes from Dirac's work for the homogeneous electron gas and is expressed as:

\[
E_{X}^{\text{LDA}}[\rho] = -\frac{3}{4} \left( \frac{3}{\pi} \right)^{1/3} \sum_{\sigma} \int \rho_\sigma(\vec{r})^{4/3} d\vec{r}
\]  

where \( \sigma \) denotes either "up" or "down" electron spin.

Several correlation functionals were proposed in the local density approximation, i.e. the VWN (Volo-Wilks-Nusair - 1980) or the PZ (Perdrew-Zunger - 1981) functionals. The VWN correlation functional is:

\[
E_{C}^{\text{VWN}}[\rho] = A \left\{ \ln \frac{\rho(\vec{r})}{\rho(\vec{r}) + b\sqrt{\rho(\vec{r})} + c} + \frac{2b}{\sqrt{4c-b^2}} \tan^{-1} \frac{\sqrt{4c-b^2}}{2\sqrt{\rho(\vec{r})} + b} - \frac{bx_0}{(x_0)^2 + bx_0 + c} \right\}
\]

with known values for \( A, x_0, b \) and \( c \).

To overcome some known weaknesses of the LDA functional (mainly underestimation of bond lengths), corrections have been developed. At this point, the DFT method is 'local' in the sense that we probe the electronic interactions between each \( \vec{r}_i \) and \( \vec{r}_j \) point mesh of the
density grid only (for which the electronic densities are defined) but we have no information for the remaining locations. One source of improvement is to go in the direction of the non locality and most of the developments in the DFT go into this direction. With these approaches density inhomogeneities, or in other words density spatial variations, can be taken into account without changing the grid that maps the electronic density.

- One approach is to express the density using a Taylor development (for the exchange-correlation part only): this gives rise to Generalized-Gradient Approximation (GGA) functionals when including first derivatives of the electron density, and gives rise to meta-GGA functionals when including second derivatives of the electron density.

- The other approach is to include (in the exchange-correlation part only) terms known from other methods (local methods). Hence hybrid functionals include part of the exact Hartree-Fock exchange and double-hybrid functionals include parts of the exact Hartree-Fock exchange and of the MP2\textsuperscript{138} correlation.

**BLYP: a Generalized-Gradient Approximation (GGA) functional**

**BLYP** is one example of a GGA functional which is based on the B88 (1988) exchange functional from Becke\textsuperscript{123} and on the LYP (1988) correlation functional from Lee, Yang and Parr\textsuperscript{124}.

\[ E_{XC}^{BLYP}(\rho) = E_{XC}^{B88}(\rho) + E_{XC}^{LYP}(\rho) \]  \hspace{1cm} (4.20)

The LDA generally underestimates the exchange energy by more or less 10%. To improve this result, the B88 exchange functional is written:

\[ E_{X}^{B88}(\rho) = E_{X}^{LDA}(\rho) - \beta \sum_{\sigma} \int \rho_{\sigma}(r)^{3} \frac{x_{\sigma}^{2}}{1 + 6x_{\sigma} \sinh^{-1}(x_{\sigma})} dr \]  \hspace{1cm} (4.21)

with

\[ x_{\sigma} = \frac{|\nabla \rho_{\sigma}|}{\rho_{\sigma}^{3/2}} \]  \hspace{1cm} (4.22)

where \( \beta = 0.0042 \). This constant has been found through the minimisation of the relative deviation of the exchange energy with respect to the exact Hartree Fock exchange energy for a set of atoms.

The LYP correlation functional is written:

\[ E_{C}^{LYP}(\rho) = -0.04918 \int \frac{1}{1 + 0.349 \rho^{-3/5}} \left( \rho + 0.132 \rho^{-2/5} \left[ \frac{3}{10} (3\pi^{2})^{2/5} \rho^{2/5} - 2 \frac{1}{8} \frac{\left| \nabla \rho \right|^{2}}{\rho} - \frac{1}{8} \nabla^{2} \rho + \left( \frac{1}{9} \frac{\left| \nabla \rho \right|^{2}}{\rho} - \frac{1}{8} \nabla^{2} \rho + \frac{1}{18} \nabla^{2} \rho \right) e^{-0.2533 \rho^{rac{1}{3}}} \right) dr \]  \hspace{1cm} (4.23)
Numerous other GGA functionals can be found in the literature. One can cite for instance the other popular PBE\textsuperscript{150} (Perdrew-Burke-Ernzerhof - 1996) XC functional.

**Beyond the GGA XC functionals**

**Meta-GGA** (**Generalized Gradient Approximation**). In the meta-GGA XC functionals, $\nabla^2 \rho$ terms are now included. One can cite the M06-L (**Minnesota06 - Local**) functional\textsuperscript{151} for instance. Such functional is highly parametrised. I am not going to describe it any further.

**Hybrid functionals.** A part of the (exact) Hartree-Fock exchange is now included into the exchange functional. B3LYP\textsuperscript{152} and PBE0\textsuperscript{153,154} which are the hybrid counterparts of the BLYP\textsuperscript{123,124} and PBE\textsuperscript{155} GGA functionals include such terms.

In this thesis, the B3LYP functional will be used, its expression is:

$$E_{xc}^{B3LYP} = 0.08E_{xc}^{LDA} + 0.72E_{xc}^{B88} + 0.20E_{xc}^{HF} + 0.19E_{xc}^{VWN} + 0.81E_{xc}^{LYP} \quad (4.24)$$

In this expression, $E_{xc}^{LDA}$ and $E_{xc}^{VWN}$ are coming from the LDA approximation as defined in equations 4.18 and 4.19, $E_{xc}^{B88}$ is coming from the B88 exchange functional defined in equation 4.21 and $E_{xc}^{LYP}$ is coming from the LYP correlation functional defined in equation 4.23. The coefficients are fitted to a set of atomisation energies, ionisation potentials, proton affinities, and total atomic energies.\textsuperscript{152} $E_{xc}^{HF}$ is the Hartree-Fock exchange written:

$$E_{xc}^{HF} = -\frac{1}{2} \sum_{i}^{N} \sum_{i \neq j}^{N} \sum_{\mu \lambda}^{m} c_{\mu i} c_{\lambda j} \int \frac{\phi_{\mu}(\vec{r}) \phi_{\lambda}(\vec{r}) \phi_{\nu}(\vec{r}) \phi_{\sigma}(\vec{r})}{|\vec{r}_i - \vec{r}_j|} d\vec{r}_i d\vec{r}_j \quad (4.25)$$

with $\mu$, $\nu$, $\lambda$ and $\sigma$ gaussian atomic functions, $c_{\mu i}$ is a coefficient that determines the participation of the gaussian function $\phi_{\mu}(\vec{r})$ into the mono-electronic wavefunction $\psi_i$.

To reduce the computational price of a hybrid functional in DFT-MD simulations, specificities are presented in subsection 4.1.6.

**Long range hybrid functionals.** Long range hybrid functionals are hybrid functionals in which the percentage of Hartree Fock exchange $E_{xc}^{HF}$ in the total of exchange-correlation functional is varying depending on the distance between the two interacting electrons. One can cite the CAM-B3LYP\textsuperscript{156} functional which is the long range counterpart of the hybrid B3LYP functional. In the CAM-B3LYP, the composition of the exchange functional is a function of the $r_{ij}$ distance, with at small range distance: $E_{xc}^{CAM-B3LYP} = 0.81 E_{xc}^{B88} + 0.19 E_{xc}^{HF}$ and at long range: $E_{xc}^{CAM-B3LYP} = 0.35 E_{xc}^{B88} + 0.65 E_{xc}^{HF}$.

**Further and most recent developments.** DFT is a field with strong ongoing developments. Among the latest developments one can cite double-hybrid functionals that also include part of the MP2 correlation. Let us cite the B2-PLYP double hybrid functional for which the exchange-correlation functional is now written:
\[ E_{xc}^{B2-PLYP} = 0.47E_{xc}^{B88} + 0.53E_{xc}^{HF} + 0.73E_{c}^{LYP} + 0.27E_{c}^{PT2} \]  

with:

\[ E_{c}^{PT2} = \frac{1}{4} \sum_{i \neq j} \sum_{a}^{N} \sum_{b \neq a}^{N} \sum_{m}^{N} \sum_{n}^{N} \frac{(i|a|b)^{2}}{\epsilon_{a} + \epsilon_{b} - \epsilon_{i} - \epsilon_{j}} \]  

with \( a \) and \( b \), unoccupied (virtual) Kohn-Sham molecular orbitals, \( \epsilon_{a} \) and \( \epsilon_{b} \) their energies and \( \epsilon_{i} \) and \( \epsilon_{j} \) the energies of the Kohn-Sham molecular orbitals occupied by the electrons \( i \) and \( j \).

**Dispersion corrections**

GGA functionals like BLYP are known to struggle to possibly reproduce the weak van der Waals interactions and several solutions have been developed to overcome the underestimation of dispersion interactions. One can cite non local van der Waals functionals (vdW-DF)\(^{157} \), dispersion-corrected atom-centered potentials (DCAPs)\(^{158} \), but we will focus on the solution chosen for this work, i.e. the D3 correction\(^{125} \) from Grimme and collaborators. This is a computationally not too expensive method, thus its easy use in DFT-MD simulations. The D3 term is an additional term to the Kohn-Sham energy. The Grimme corrections are popular due to the good results provided at a reasonable cost in terms of computational resources. One has to be careful: this correction does not correct the electronic density, it corrects only the associated electronic energy, and it is somehow its main drawback.

The BLYP-D3 energy is calculated as:

\[ E_{BLYP-D3}[\rho] = E_{BLYP}[\rho] - E_{Disp-D3} \]  

The way to calculate \( E_{BLYP}[\rho] \) has been described above (see equations 4.16, 4.21 and 4.23) and the term \( E_{Disp-D3} \) is calculated as:

\[ E_{Disp-D3} = E^{(2)} + E^{(3)} \]  

where \( E^{(2)} \) and \( E^{(3)} \) respectively refer to a two-body and three-body term.

The two-body \( E^{(2)} \) term is written as:

\[ E^{(2)} = \sum_{A}^{M} \sum_{B > A}^{M} \sum_{n=6,8}^{N} s_{n} \frac{C_{A}^{AB} f_{d,n}(R_{AB})}{(R_{AB})^{n}} \]  

with \( C_{n}^{AB} \) the nth-order dispersion coefficient for the atom pair \( AB \) (see ref\(^{125} \) for the equations to calculate \( C_{n}^{AB} \) and \( C_{8}^{AB} \) coefficients), \( M \) is the number of atoms, \( f_{d,n}(R_{AB}) \) is a damping function (see equation 4.31), \( R_{AB} \) is the AB internuclear distance, and \( s_{n} \) is a scaling factor equal to unity for \( n=6 \) (\( s_{6}=1 \)) while it is fitted to match the energies from a set of systems for each exchange-correlation functional for \( n=8 \) (for BLYP: \( s_{8}^{BLYP}=1,682 \), for B3LYP: \( s_{8}^{B3LYP}=1,703 \)). The damping function \( f_{d,n}(R_{AB}) \) corrects the \( \frac{1}{r^{n}} \) form that is not valid at short distances\(^{159} \):
4.1 ELECTRONIC REPRESENTATION: DFT (DENSITY FUNCTIONAL THEORY)

\[
f_{d,n}(r_{AB}) = \frac{1}{1 + 6 \left( \frac{R_{AB}}{R_{0}^{AB}} \right)^{-\alpha_n}} \tag{4.31}
\]

\( t_n \) is a scaling factor of the cutoff radii \( R_{0}^{AB} \) (see ref\textsuperscript{125} for the equations to calculate the cutoff radii \( R_{0}^{AB} \), \( t_6 \) is determined by a standard least-square error fitting procedure (\( t_6 = 1.094 \) for BLYP and \( t_6 = 1.261 \) for B3LYP) and \( t_8 = 1 \) for all exchange-correlation functionals. \( \alpha_n \) is the steepness parameter adjusted manually by Grimme. \( \alpha_6 = 14 \) and \( \alpha_8 = 16 \) for all the functionals.

The three-body term \( E^{(3)} \) is written as:

\[
E^{(3)} = \sum_{ABC} f_{d,(3)}(\bar{R}_{abc})E^{ABC} \tag{4.32}
\]

The damping function \( f_{d,(3)} \) is the same as for the two-body correction (equation 4.31) using the scaling factor \( t_n = \frac{4}{3} \) and the steepness parameter \( \alpha = 16 \). \( E^{ABC} \) is equal to:

\[
E^{ABC} = \frac{\mathcal{C}^{ABC}(3\cos \theta_a \cos \theta_b \cos \theta_c + 1)}{(R_{ab}R_{bc}R_{ca})^3} \tag{4.33}
\]

where \( \theta_a, \theta_b \) and \( \theta_c \) are the internal angles of the triangles formed by the 3 atoms A, B and C and \( R_{ab}, R_{bc} \) and \( R_{ca} \) are the internuclear distances. \( \mathcal{C}^{ABC} \) can be approximated as:

\[
\mathcal{C}^{ABC} \approx -\sqrt{\mathcal{C}^{AB}_6 \mathcal{C}^{BC}_6 \mathcal{C}^{CA}_6} \tag{4.34}
\]

with \( \mathcal{C}^{AB}_6 \) dispersion coefficients calculated as in the work from Grimme et al.\textsuperscript{125}

4.1.3 How to build the wavefunction/electronic density in the GPW Gaussian Plane Wave scheme

If we restart from the time independent Schrödinger equation in eq 4.7, we now have defined the Hamiltonian \( H \), the BLYP-D3 functional, but we still have to discuss the way to calculate the wavefunction and the electronic density.

The CP2K package used in my work for DFT-MD trajectories is built on two different basis sets: a gaussian basis set localised around each nucleus and a plane waves basis set delocalised in space in the whole box of simulation. For mathematical reasons that ease up and decrease the computational cost, two spaces are used in the CP2K code that we are using for DFT-MD simulations: the real space (\( \vec{r} \), cartesian coordinates) applied to use the gaussian function and its reciprocal space (\( \vec{G} \)) applied to use the plane waves functions. The two spaces are Fourier transforms of each others.

Gaussian basis set.

We first have to build the mono-electronic wavefunctions \( \psi_i \) on a gaussian basis set. The Kohn-Sham spinorbitals \( \psi_i \) (Slater determinant, equation 4.9) are expanded as a linear combination of the gaussian atomic functions \( \phi_{\mu}(\vec{r}) \) of the gaussian basis set:
\[ \psi_i(\vec{r}) = \sum_{\mu=1}^{m} c_{i\mu} \phi_\mu(\vec{r}) \]  

(4.35)

with \( m \) the number of gaussian atomic functions (it depends on the gaussian basis set chosen) and \( c_{i\mu} \) a coefficient that determines the participation of the gaussian function \( \phi_\mu(\vec{r}) \) into the (Kohn-Sham spinorbital) mono-electronic wavefunction \( \psi_i \). Note that each gaussian atomic function is a linear combination of gaussian functions that are called primitives. This combination is particularly important for the most localised electrons since it is difficult to reproduce narrow orbitals with gaussian-type functions.

To reach the "best" wavefunction or electronic density (within the limitation of the BLYP-D3 functional), an infinite basis set would be needed, which would unfortunately lead to an infinite time of calculations. In practice we have to make some compromise on the basis set size and two different gaussian basis sets have been used in my work. The aug-TZV2P basis set (for peptidic systems, see section 2.1 in chapter 2) and the m-TZV2PX\(^{129}\) basis set (for the DNA base pairs and the phenol derivatives, see sections 2.3 and 2.2 of chapter 2). The aug-TZV2P basis set is named following the Dunning basis style\(^{160}\). "TZV" corresponds to the three basis functions assigned to the valence electrons (equivalent to the "p" type of atomic orbitals). "2P" refers to the two polarisation functions that have a bigger angular momentum than the valence functions (equivalent to the "d" type of atomic orbitals). "aug" refers to diffuse functions needed to describe electrons distant from the nuclei. The second basis set, m-TZV2PX, is part of the basis sets developed by VandeVondele and Hutter\(^{129}\) for the CP2K package and optimised to work in combination with the GTH pseudopotentials\(^{130}\). The "m" in the name is for "molopt". Diffuse primitives are systematically included. This reoptimisation provides better accuracy for an equal amount of basis functions in comparison with non "molopt" basis sets. "m" refers to the re-optimisation, and "TZV2P" has the same meaning than in the previous functional (3p2d). Finally "X" refers to the inclusion of Gaussian functions of the "f" type.

For the B3LYP-D3-MD calculations, an ADMM (Auxiliary Density Matrix Method)\(^{161}\) has been used. Density matrices are introduced in subsection 4.1.5 and the auxiliary density matrix method is briefly described in section 4.1.6. Anyway, to build this auxiliary matrix, a second gaussian basis set is needed and the "cpFIT3" basis set\(^{161}\) has been selected after several tests in my work.

The electronic density is obtained from the mono-electronic wavefunctions via the equation:

\[ \rho(\vec{r}) = \sum_i^N |\psi_i(\vec{r})|^2 \]  

(4.36)

By mixing equation 4.35 and 4.36, we have:

\[ \rho(\vec{r}) = \sum_i^N \sum_{\mu}^m \sum_{\nu}^m c_{i\mu} \phi_\mu(\vec{r}) c_{i\nu} \phi_\nu(\vec{r}) = \sum_{\mu}^m \sum_{\nu}^m P^{\nu\mu} \phi_\mu(\vec{r}) \phi_\nu(\vec{r}) \]  

(4.37)
with \( m \) is the amount of gaussian atomic functions (it depends on the gaussian basis set chosen) and \( c_\mu \) and \( c_\nu \) are the coefficients of participation of the gaussian atomic functions \( \phi_\mu \) and \( \phi_\nu \) in the electronic density. Here we write the gaussian functions as real functions. In this equation, we have written \( P^{\nu \mu} = \sum_i^N c_{i\mu} c_{i\nu} \), the density matrix elements that will be used in the next section.

**Plane Waves basis set.**

In CP2K, there is also a PW (Plane Waves) basis set used on top of the gaussian one. This is the dual basis set approach or the GPW scheme (Gaussian Plane Waves). The electronic density described in equation 4.37 in the real space is now projected into the reciprocal space using a Fourier transform (Fast Fourier Transform). Note that multiple grids (with more or less sparse plane waves) are used in the reciprocal space. The narrow and large gaussian atomic functions will not be projected on the same grid, in order to speed up the calculations and in order to describe each gaussian atomic function with the same precision. By default, CP2K uses four different grids with a relative energy cut-off of 40 Ry between each successive grid. The density projected on the plane waves is now:

\[
\rho(r) = \sum_G \tilde{\rho}_G f_G(r)
\]  

(4.38)

\( \tilde{\rho}_G \) is the expansion coefficient that ensures the following condition: \( \rho(r) = \rho(r) \), \( f_G(r) \) is a plane wave written as:

\[
f_G(r) = e^{i\vec{G} \cdot \vec{r}} \sqrt{V}
\]

(4.39)

\( V \) is the volume of the simulation box and \( \vec{G} \) a vector of the reciprocal space.

As for the Gaussian basis set, an infinite plane waves basis set would be needed to converge to the "best" wavefunction in the limit of our electronic representation. In practice we have to limit its size for obvious computational reasons:

\[
\frac{1}{2} ||\vec{G}||^2 < E_{\text{Cut-off}}
\]

(4.40)

After convergence tests (not discussed here), we chose a value of \( E_{\text{Cut-off}} = 450 \text{ Ry} \) in all our presented works.

**Periodic boundary conditions (PBC).**

As most of MD codes, the CP2K package uses periodic boundary conditions in the three directions of space. This mimics a continuous medium composed of an infinite number of the unit cell. One of the cells is the simulation box, and the other cells are the replica or images. An illustration of a simulation box containing one dialanine peptide and its replica (here displayed
along y and z directions only) is presented in figure 4.1. The main box is highlighted with the blue lines. One can see immediately that each box interacts with its replica.

Figure 4.1: Illustration of PBC. Here, the gas phase dialanine dipeptide and its replica. Replica are in all three directions of space, for the picture only y and z directions replica are visible. The blue square is the limit of the central simulation box.

The size of the simulation box is thus crucial. A too big simulation box would be of course the best in order to reduce as much as possible interactions between the replica, but it will be computationally expensive. One has to choose a box reasonable in size for reasonable computational costs and no artefact of interaction. In order to choose the best compromise in box size, in practice we take the largest molecule of each family of system studied and we do geometry optimisations for different dimensions of the simulation box (always cubic boxes). The size chosen for the simulations corresponds to the size for which the electronic energy reaches a plateau (convergence). The size chosen is 20 Å³ for the dipeptide series and for the Ac-Phe-OMe monomer, 25 Å³ for the DNA base pairs and 24 Å³ for Z-Ala₆-NH₂, the (Ac-Phe-OMe)₂ dimer and the phenol derivatives.

Pseudopotentials.

It is already difficult to reproduce the behaviour of the core electrons with a gaussian basis set, this is even more challenging with a plane waves basis set because it requires extremely large basis sets (i.e. for plane waves it requires very large $E_{\text{CUT-OFF}}$), this is too computationally expensive. Therefore we use pseudopotentials to describe core electrons. By eliminating the core electrons, the number of electrons that has to be explicitly described is reduced. We use the norm conservative, separable, dual space pseudo-potentials from Goedecker, Teter and Hutter¹³⁰ which are specifically compatible with the dual PW/Gaussian basis sets representation in the CP2K package. During the parametrisation of these pseudopotentials, all scalar relativistic corrections have been taken into account. See work from Goedecker et al.¹³⁰ These
pseudopotentials are separated into two equivalent parts: a local and a non-local part. The local and non-local parts will be alternatively used to calculate the different parts of the energy in equation 4.16, function of their computational efficiency (cost). The local potential \( V_{\text{loc}}(r) \) (local in the sense that it only depends on the radial distance \( r \)) is given by:

\[
V_{\text{loc}}(r) = -\frac{Z}{r-R_{\text{ion}}} \text{erf} \left( \frac{r-R_{\text{ion}}}{\sqrt{2} r_{\text{loc}}} \right) + e^{-\frac{1}{2} \left( \frac{r-r_{\text{loc}}}{r_{\text{loc}}} \right)^2} \left( C_1 + C_2 \left( \frac{r-R_{\text{ion}}}{r_{\text{loc}}} \right)^2 + C_3 \left( \frac{r-R_{\text{ion}}}{r_{\text{loc}}} \right)^4 + C_4 \left( \frac{r-R_{\text{ion}}}{r_{\text{loc}}} \right)^6 \right)
\]

(4.41)

with \( \text{erf} \) the error function that determines the range of the ionic charge fixed by the coefficient \( r_{\text{loc}} \), \( Z \) is the ionic charge calculated as the charge of the nucleus minus the charge of the electrons described by the pseudopotential, \( r-R_{\text{ion}} \) is the distance electron/nucleus and \( C_n \) are optimised coefficients. For the non-local potential (i.e. depending on \( \vec{r} \) and \( \vec{r}' \) in space):

\[
V_{\text{non-loc}}(\vec{r}, \vec{r}') = \sum_{l=0}^{l} \sum_{m=-l}^{l} N_l Y_{l,m}(\vec{r}) \left( r^{l+2j-2} e^{-\frac{1}{2} \left( \frac{r}{r_{\text{loc}}} \right)^2} \right) h_{ij} N_j Y_{l,m}(\vec{r}') \left( r'^{l+2j-2} e^{-\frac{1}{2} \left( \frac{r}{r_{\text{loc}}} \right)^2} \right)
\]

(4.42)

There is a summation over \( l \) and \( m \) respectively the azimuthal and magnetic quantum numbers. \( N_l \) are normalisation constants, \( Y_{l,m}(\vec{r}) \) are the spherical harmonics with only angular dependencies (l,m). \( e^{l+2j-2} e^{-\frac{1}{2} \left( \frac{r}{r_{\text{loc}}} \right)^2} \) are gaussian radial projectors. \( h_{ij} \) and \( r_l \) are optimised parameters.

Both local and non-local pseudopotentials can be transformed in the reciprocal space \( \vec{G} \), with the following expressions:

\[
V_{\text{loc}}(\vec{G}) = -4\pi \frac{Z}{V} e^{-\frac{1}{2} (G_{\text{loc}})^2} + \sqrt{\frac{(2\pi)^3}{V}} e^{-\frac{1}{2} (r_{\text{loc}})^2} \left( C_1 + C_2 \left( 3 - (r_{\text{loc}})^2 \right) \right) + \sum_{l=0}^{l} \sum_{m=-l}^{l} N_l Y_{l,m}(\vec{G}) p_{ij} N_j Y_{l,m}(G') p_j
\]

(4.43)

where \( V \) is the volume of the simulation box and \( \vec{G} \) is vector of the reciprocal space, and:

\[
V_{\text{non-loc}}(\vec{G}, \vec{G}') = \sum_{l=0}^{l} \sum_{m=-l}^{l} \sum_{i,j=0}^{l} N_l Y_{l,m}(\vec{G}) p_{ij} N_j Y_{l,m}(\vec{G}') p_j
\]

(4.44)

with \( p_{ij} \) and \( p_j \) the radial projectors written in the reciprocal space.

The local part of the pseudopotential can be furthermore separated into two components, a short range contribution and a long range contribution for convenience of calculations.

\[
V_{\text{loc}}(r) = V_{\text{loc}}^{LR}(r) + V_{\text{loc}}^{SR}(r)
\]

(4.45)
with (written in the real space):

\[
V_{LR}(r) = -Z_{\text{ion}} \frac{r - R_{\text{ion}}}{r - R_{\text{ion}}} \text{erf} \left( \frac{r - R_{\text{ion}}}{\sqrt{2} r_{\text{loc}}} \right) \tag{4.46}
\]

\[
V_{\text{SR}}^{\text{loc}}(r) = e^{-\frac{1}{2} \left( \frac{r - R_{\text{ion}}}{r_{\text{loc}}} \right)^2} \left( C_1 + C_2 \left( \frac{r - R_{\text{ion}}}{r_{\text{loc}}} \right)^2 + C_3 \left( \frac{r - R_{\text{ion}}}{r_{\text{loc}}} \right)^4 + C_4 \left( \frac{r - R_{\text{ion}}}{r_{\text{loc}}} \right)^6 \right) \tag{4.47}
\]

The \( r_{\text{loc}} \) parameter has been optimised for the short range to perfectly cancel the long range in the overlapping part. The list of \( r_{\text{loc}} \) values for the two first rows of the periodic table can be found in literature.\(^{130}\)

For the B3LYP-D3 molecular dynamics simulations, we have created our own pseudopotentials for the B3LYP density functional using functionalities implemented in the CP2K package\(^{122}\). In practice, pseudopotentials are optimised with respect to a calculation taking into account all electrons.

We will see in the next subsection how these expressions are integrated into the density functional.

### 4.1.4 Kohn-Sham DFT in the CP2K package

In the CP2K package, the different parts of the energy (equation 4.16) are calculated on the different basis sets presented above, as a function of what is the cheapest in terms of computational cost. The exact equations can be found in the paper "Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach"\(^{122}\) that describes the implementation of the method in CP2K. I summarise a few equations below.

If we go back to the general expression of the electronic energy of the system in DFT, we have already written previously:

\[
E[\rho] = T[\rho] + E_{\text{ext}}[\rho] + E_H[\rho] + E_{\text{XC}}[\rho] \tag{4.48}
\]

to which the nuclei-nuclei \( V_{NN} \) interaction can be added-up, for mathematical tricks that will become clear just afterwards.

The kinetic energy is calculated using the mono-electronic wavefunctions, in the real space, following the Kohn-Sham scheme, see equation 4.12:

\[
T[\rho] = -\frac{1}{2} \sum_i N \langle \psi_i(\vec{r}) | \nabla^2 | \psi_i(\vec{r}) \rangle = -\frac{1}{2} \sum_{\mu \nu} p_{\mu \nu} \langle \phi_\mu(\vec{r}) | \nabla^2 | \phi_\nu(\vec{r}) \rangle \tag{4.49}
\]

using the gaussian expansion 4.35 for the \( \psi_i \) in the real space, and the density matrix elements...
\( P^{\mu\nu} \) defined previously.

From now on, the nuclei-electron interactions and the Hartree electron-electron interactions are going to be gathered, and their expressions will be rewritten to work with the pseudopotentials. Added to these two terms, is now included the nuclei-nuclei energy interaction, giving the following expression:

\[
\sum_{I} \sum_{J \neq I} \int \rho(\vec{r}) V_{\text{ext}}(\vec{r}) d\vec{r} + \frac{1}{2} \int \int \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + \frac{1}{2} \sum_{I} \sum_{J \neq I} Z_{I} Z_{J} |\vec{R}_{I} - \vec{R}_{J}| \tag{4.50}
\]

where the first term is \( V_{\text{NN}}[\rho] \), the second is \( E_{\text{ext}}[\rho] \), and the third is \( E_{H}[\rho] \) in eq. 4.48.

From equation 4.50, the interactions are now separated into short range interactions that will be treated in the real space (i.e. \( \vec{r} \)) and long range interactions that will be treated in the reciprocal space (i.e. \( \vec{G} \)). The gaussian basis set is used in the real space and the plane waves basis set is used in the reciprocal space. In particular, the non-local part of \( E_{\text{tot}} \) tends rapidly to zero outside the core region of the nuclei. Combining the local part of \( E_{\text{ext}} \), the \( E_{H} \) Hartree energy for the (valence) electrons and the nuclei-nuclei interaction provides the electrostatic energy, as written in Equation 4.51, where \( Z_{I} \) refers to the charge of the ionic core (nuclei) \( I \):

\[
E^{ES} = \sum_{\mu\nu} P^{\mu\nu} \langle \phi_{\mu}(\vec{r}) | V_{\text{loc}}(\vec{r}) | \phi_{\nu}(\vec{r}) \rangle + \frac{1}{2} \int \int \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + \frac{1}{2} \sum_{I} \sum_{J \neq I} Z_{I} Z_{J} |\vec{R}_{I} - \vec{R}_{J}| \tag{4.51}
\]

For a clear separation of the ionic cores/nuclei, a gaussian charge distribution is introduced for each nucleus \( I \):

\[
\rho_{c}^{I}(\vec{r}) = -\frac{2}{(R_{I}^{c})^{3}} \pi^{-3/2} \exp \left[ -\frac{1}{2} \left( \frac{|\vec{r} - \vec{R}_{I}|}{R_{I}^{c}} \right)^{2} \right] \tag{4.52}
\]

Choosing \( R_{I}^{c} = \sqrt{2} r_{\text{loc}} \) in this expression enables an exact cancellation of the long-range term arising from the local pseudopotential. Thus, the interaction between the nuclei and the electronic density of the electrons is reduced to \( \int V_{\text{loc}}^{SR}(\vec{r}) \rho(\vec{r}) d\vec{r} = \int \int \frac{\rho_{c}(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' \), with \( \rho_{c}(\vec{r}) = \sum_{I} \rho_{c}^{I}(\vec{r}) \). We thus have:

\[
E^{ES} = \frac{1}{2} \int \int \frac{\rho_{c}(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + \frac{1}{2} \sum_{I} \sum_{J \neq I} Z_{I} Z_{J} |\vec{R}_{I} - \vec{R}_{J}| \tag{4.53}
\]

If a total charge distribution \( \rho_{\text{tot}}(\vec{r}) = \rho_{c}(\vec{r}) + \rho(\vec{r}) \) is introduced, the electrostatic energy becomes:

\[
E^{ES} = \frac{1}{2} \int \int \frac{\rho_{\text{tot}}(\vec{r}) \rho_{\text{tot}}(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' - \frac{1}{2} \int \int \frac{\rho_{c}(\vec{r}) \rho_{c}(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + \frac{1}{2} \sum_{I} \sum_{J \neq I} Z_{I} Z_{J} |\vec{R}_{I} - \vec{R}_{J}| \tag{4.54}
\]

also written:

\[
E^{ES} = \frac{1}{2} \int \int \frac{\rho_{\text{tot}}(\vec{r}) \rho_{\text{tot}}(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + E_{\text{overlap}} - E_{\text{self}} \tag{4.55}
\]
It can be shown that $E_{\text{overlap}}$ is the nuclei-nuclei short-range pair energy:

$$E_{\text{overlap}} = \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} \times \text{erfc} \left( \frac{|\vec{R}_I - \vec{R}_J|}{\sqrt{(R_{ic})^2 + (R_{jc})^2}} \right)$$

(4.56)

(\text{erf} + \text{erfc} = 1) and that the self-energy reads:

$$E_{\text{self}} = -\sum_I \frac{1}{\sqrt{2\pi}} Z_I^2 \frac{1}{R_{ic}^I}$$

(4.57)

Both $E_{\text{overlap}}$ and $E_{\text{self}}$ are computed in the real space.

The new Hartree energy function of the total charge density of the system is non-local, i.e. depends on $\vec{r}$ and $\vec{r'}$, and has to be handled in the Fourier space with plane-waves since $\rho_{\text{tot}}(\vec{r})$ is easily written in Fourier space as:

$$\rho_{\text{tot}}(\vec{r}) = \frac{1}{\Omega} \sum_\vec{G} \tilde{\rho}_{\text{tot}}(\vec{G}) \exp(i\vec{G} \cdot \vec{r})$$

(4.58)

with $\vec{G}$ representing the reciprocal lattice vectors and $\Omega$ the volume of the unit cell. More precisely, the total density is first calculated on a grid ($\vec{r}$ is discretized); then a Fourier transform is applied to get $\tilde{\rho}_{\text{tot}}(\vec{G})$. In the Fourier space, the Hartree energy is computed via

$$\frac{2\pi}{N} \sum_\vec{G} \tilde{\rho}_{\text{tot}}(\vec{G}) \tilde{\rho}_{\text{tot}}(\vec{G}) \frac{\vec{G}^2}{G^2}.$$ 

The exchange-correlation energy is discretized on a grid in the real space; density gradients are approximated numerically on the grid and the exchange-correlation energy as well as its derivatives are evaluated on each grid point. This allows the calculation of the exchange-correlation potential, and thus the final energy contributions after multiplication by Gaussian basis functions.

4.1.5 SCF procedure: optimisation of the wavefunction

We will now see how to solve the time independent electronic Schrödinger equation $H\Psi = E\Psi$ in order to get $\Psi$ and $E$. This is done by starting with an initial guess of $\Psi$ that will be refined (optimised) in a self consistent procedure (SCF). This is based on the variational principle.

The secular equation

If we restart from the Schrödinger equation (equation 4.7), developed using equation 4.35 for the wavefunction and written here on the gaussian basis set only, we have:

$$\sum_v \sum_\mu c_\mu c_v \phi_\mu H \phi_v = E \sum_v \sum_\mu c_\mu c_v \phi_\mu \phi_v$$

(4.59)

This can be rewritten in matrix language as:

$$\vec{F} \vec{C} = S \vec{C} \vec{E}$$

(4.60)
which is the secular equation, with $F$ the Kohn-Sham matrix:

$$
\tilde{F} = \begin{bmatrix}
F_{\nu\nu} & \ldots & \ldots & F_{\nu\mu} \\
\vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
F_{\mu\nu} & \ldots & \ldots & F_{\mu\mu}
\end{bmatrix}
$$

(4.61)

where $F_{\mu\nu} = \phi_\mu H \phi_\nu$ are elements of the $\tilde{F}$ matrix,

$\tilde{C}$ is the density matrix:

$$
\tilde{C} = \begin{bmatrix}
C_{\nu\nu} & \ldots & \ldots & C_{\nu\mu} \\
\vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
C_{\mu\nu} & \ldots & \ldots & C_{\mu\mu}
\end{bmatrix}
$$

(4.62)

where $C_{\mu\nu} = c_\mu c_\nu$ are elements of the $\tilde{C}$ matrix,

$\tilde{S}$ is the overlap matrix:

$$
\tilde{S} = \begin{bmatrix}
S_{\nu\nu} & \ldots & \ldots & S_{\nu\mu} \\
\vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
S_{\mu\nu} & \ldots & \ldots & S_{\mu\mu}
\end{bmatrix}
$$

(4.63)

where $S_{\mu\nu} = \phi_\mu \phi_\nu$ are elements of the $\tilde{S}$ matrix,

and $\tilde{E}$ is a diagonal matrix whose elements are the energies of the one electron molecular orbitals:

$$
\tilde{E} = \begin{bmatrix}
\epsilon_1 \\
\epsilon_2 \\
\epsilon_3 \\
\vdots \\
\epsilon_m
\end{bmatrix}
$$

(4.64)

with $m$ the number of atomic functions in the chosen basis set, and $\epsilon_i$ is the energy of the molecular orbital number $i$.

Once the eigenvalue problem (equation 4.60) is solved using an initial guess for $\Psi$, we obtain $m$ (as many as the number of atomic functions) molecular orbitals, with a resulting $\frac{N}{2}$ occupied orbitals (lowest energy ones) and $m - \frac{N}{2}$ non-occupied orbitals (with $N$ the total number of electrons). Equation 4.60 is solved self consistently starting from the guess and getting to the solution. But this first wavefunction is usually not the one of lowest energy for a given set of positions of the nuclei. Algorithms are therefore applied in order to find new guesses in order to obtain the lowest energy possible. In practice forces on the nuclei ($\frac{\partial E}{\partial c}$ with $c$ an element of the density matrix) are calculated. If the convergence criteria in terms of energy and forces (defined at the start of the calculation) are not satisfied, a new $\tilde{C}$ density matrix is created and the secular equation is solved again, until convergence is reached.
Once the calculation has converged, $\bar{C}$ and $\bar{E}$ are thus known, and therefore $\Psi$ and $E$, the wavefunction and energy of the system respectively. In CP2K, the OT transformation is used instead of the traditional diagonalisation but the spirit of SCF remains similar.

### 4.1.6 Specific characteristics for hybrid functionals in CP2K

For the purpose of functionals comparison investigated in this work, see chapter 10, one trajectory of 20 ps using the B3LYP-D3 functional for each of the four phenol derivatives (presented in section 2.2 of chapter 2) has been calculated.

The B3LYP functional energy expression is presented in equation 4.24, and it uses a part of the (exact) Hartree-Fock exchange $E_{HF}^x$:

$$E_{HF}^x = -\frac{1}{2} \sum_i^{N} \sum_{i \neq j}^{N} \sum_{\mu \lambda \nu \sigma}^{m} c_{i \mu} c_{i \lambda} c_{j \nu} c_{j \sigma} \int \frac{\phi_{\mu}(\vec{r}_i) \phi_{\nu}(\vec{r}_i) \phi_{\lambda}(\vec{r}_j) \phi_{\sigma}(\vec{r}_j)}{|\vec{r}_i - \vec{r}_j|} \ d\vec{r}_i \ d\vec{r}_j \quad (4.65)$$

with $\mu$, $\nu$, $\lambda$, and $\sigma$ gaussian atomic functions and $C$ elements of the $\bar{C}$ density matrix presented in equation 4.62.

As for the other methods, the computational cost of the Hartree-Fock exchange calculation varies with the size of the system and/or the basis set employed. A brute force implementation of the Hartree-Fock exchange term leads to a cost growth proportional to $m^4$, with $m$ the number of basis functions. The cost of calculation thus quickly becomes inaccessible even for systems of rather small size.

To reduce this computational cost, different screenings\textsuperscript{[163]} are implemented into the CP2K package\textsuperscript{[122]}. A first screening is applied to the electron repulsion integrals in equation 4.65, that avoids calculating interactions whose contributions are smaller than a given threshold $\varepsilon$ (defined in Hartree). If a sufficient amount of integrals are ignored, the cost of HFX (Hartree-Fock eXchange) calculations can be reduced to $m^2$. In our calculation, we have set the $\varepsilon$ value to $1 \times 10^{-8} \text{ Ha}$ after convergence tests. A second screening can be applied on the density matrix at the first SCF iteration. This second solution can reduce the calculation cost to $m^1$ but has not been chosen in our work. Another source of reduction of computational cost arises from a truncated Coulomb operator for the Hartree-Fock exchange\textsuperscript{[164]}. In practice we have:

$$g_{TC}(r_{12}) = \begin{cases} \frac{1}{r_{12}}, & r_{12} \leq R_{TC} \\ 0, & r_{12} > R_{TC} \end{cases} \quad (4.66)$$

with $R_{TC}$ a cut-off value fixed at 12 Å in our work after convergence tests. The missing long-range exchange above $R_{TC}$ is corrected by a term applied on the long-range exchange part of the GGA density functional.

Even after applying these screening procedures, the calculation of Hartree-Fock exchange (HFX) remains computationally demanding and even with a linear scaling scheme (for the
number of atoms) as implemented in CP2K, the cost increases with a fourth power of the number of (primitive) basis functions. To reduce the calculation cost related to the size of the basis set, an ADMM (Auxiliary Density Matrix Method)\(^1\) has been recently implemented in CP2K to calculate the HFX energy only. The gaussian atomic functions of this ADMM basis are narrow which increases the speed of calculations but induces approximations. To ensure that the quality of the calculation is influenced as little as possible by the quality of the auxiliary density matrix, a correction term is added to the exchange and correlation functional. The gaussian basis set used for the auxiliary density matrix is the “cpFIT3” basis set\(^1\), it has been selected after several tests in our work.

For the B3LYP-D3 molecular dynamics simulations, we have created our own pseudopotentials for the B3LYP density functional, using functionalities implemented in the CP2K package\(^2\). In practice, pseudopotentials are optimised based on comparisons with all electrons calculations (without pseudopotentials).

### 4.2 DFT-MD (Density Functional Theory based Molecular Dynamics)

In practice, once the electronic density and the associated electronic energy have been obtained through the SCF procedure described in subsection 4.1.5, the forces acting on the nuclei can be calculated, so that the nuclei can be moved in space following the resolution of the equations of motions (velocity Verlet algorithm in our work). With these new nuclei positions, a new electronic wavefunction and density as well as a new energy are calculated, which is the basis for the next step of the molecular dynamics, and so on and so forth.

#### 4.2.1 Velocity Verlet algorithm

In dynamics, the Newton equations of motion are solved for each nucleus:

\[
\sum_{B \neq A}^{M} \vec{F}_A = m_A \vec{a}_A
\]  

with \(\vec{a}_A\) the acceleration of the nucleus A: \(\vec{a}_A = \frac{d}{dt} \vec{v}_A = \frac{d^2}{dt^2} \vec{r}_A\) (second derivative of the position with respect to time, first derivative of velocity) and \(\vec{F}_A\) the total force acting on nucleus A defined as:

\[
\sum_{B \neq A}^{M} \vec{F}_A = -\frac{\partial E}{\partial \vec{r}_B}
\]  

with \(\vec{r}_B\) the cartesian coordinates x, y and z of the nucleus B, and E is the total energy (electronic+nuclei) of the system (see section 4.1).

In molecular dynamics, one needs to propagate the motion of each atom in terms of positions and velocities. To that end, the velocity Verlet algorithm is used:
\[ \vec{r}_A(t + \delta t) = \vec{r}_A(t) + \vec{v}_A(t) \delta t + \frac{1}{2 m_A} \sum_{B \neq A}^{M} \vec{F}_A(t) \delta t^2 \]  

(4.69)

with \( \vec{r}_A(t) \) the position of the A nucleus at time \( t \), \( \vec{r}_A(t + \delta t) \) its position at time \( t = t + \delta t \), with \( \delta t \) the incremental time step, \( \vec{a}_A(t) \) the acceleration of the same nucleus at time \( t \) and \( \vec{v}_A(t) \), the velocity of the same nucleus at time \( t \):

\[ \vec{v}_A(t) = \vec{r}_A(t - \delta t) + \frac{1}{2 m_A} \left( \sum_{B \neq A}^{M} \vec{F}_A(t - \delta t) + \sum_{B \neq A}^{M} \vec{F}_A(t) \right) \delta t \]  

(4.70)

The choice of the integration time \( \delta t \) for the Verlet algorithm is crucial and has to be chosen carefully to ensure the velocity Verlet algorithm to be reversible, and to ensure that \( E \) is numerically constant as it should be in the NVE ensemble that we simulate. In our simulations, the integration time is 0.4 fs.

### 4.2.2 NVE ensemble

Our simulations are done in the NVE microcanonical ensemble, where the number of particles \( M + N \), the volume \( V \) and the total energy \( E \) (kinetic+potential) of the system, all remain constant over time. In NVE simulations, the temperature \( T \) is not a constant of motion but it fluctuates in time. It is calculated at each time step of the dynamics through the velocities of the atoms using the following equation:

\[ \sum_{A}^{M} \frac{1}{2 m_A} \vec{v}_A^2 = \frac{3}{2} M k_B T \]  

(4.71)

with \( M \) the number of atoms, \( m_A \) the mass of atom \( A \), \( \vec{v}_A \) the velocity of atom \( A \) and \( k_B \) the Boltzmann constant \( (k_B \simeq 1.38*10^{23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1}) \). From the knowledge of the instantaneous temperature at each time step, one can calculate the average temperature \( \langle T \rangle \) and its fluctuation.

In our work, the simulations aim at a temperature of 50 K for the best agreement with the experimental spectra. We usually have a fluctuation of \( \simeq 10 \text{K} \) from the average.

Before the NVE ensemble simulation, few picoseconds of simulation are done using a rescaling procedure of the velocities of the atoms in order to equilibrate the system towards the targeted temperature. The rescaling occurs whenever the temperature is out of a tolerance gap fixed at 20K (temperature is thus allowed to fluctuate between 30 and 70K) around the 50K targeted temperature, as highlighted in figure 4.2.

This equilibration period could have been done using a NVT ensemble simulation but in practice we usually find NVE with velocities rescaling more efficient (even in the gas phase). The simulations are 20 ps long (after the equilibration period described above). In the gas
4.3. Calculation of an infrared spectrum from molecular dynamics

During the trajectory, the dipole moment of the whole system at any time \( t \) is calculated using:

\[
\bar{\mu}(t) = \int \langle \psi | \hat{r} | \psi \rangle d\hat{r}
\]  

(4.72)

with \( \hat{r} \) the position operator and \( \psi \) the wavefunction of the system. This provides the knowledge of the evolution with time of the dipole moment of the system.

In molecular dynamics simulations, an infrared spectrum is calculated through the Fourier transform of the time dependent correlation function of the dipole moment:

\[
I(\omega) = \frac{2\pi \beta \omega^2}{3cV} \int_{-\infty}^{\infty} dt \langle \delta\bar{\mu}(t) \cdot \delta\bar{\mu}(0) \rangle e^{-i\omega t}
\]  

(4.73)

where \( \beta = 1/kT \), \( T \) is the average temperature of the trajectory, \( c \) is the speed of light in vacuum, \( V \) is the volume of the simulation box. The angular brackets represent a statistical average of the correlation function of the molecular dipole moment vector \( \bar{\mu}(t) \), where \( \delta\bar{\mu}(t) = \bar{\mu}(t) - < \bar{\mu} > \) is the dipole fluctuation, with \( < \bar{\mu} > \) the time average of \( \bar{\mu}(t) \). The calculation in equation 4.73 is done in the absence of an applied external field. For the prefactor this equation, we
take into account an empirical quantum correction factor (multiplying the classical line shape) of the form $\beta \bar{h} \omega / (1 - \exp(-\beta \bar{h} \omega))$, which was shown by the group in Evry and others to give accurate results on calculated IR intensities\textsuperscript{165–167}. For more detailed discussions on quantum corrections, see for instance refs.\textsuperscript{168–171}.

Equation 4.73 gives the whole infrared spectrum of one given molecular system in a single calculation, \textit{i.e.} the band positions, the band intensities and the band-shapes, through the Fourier transform of a time correlation function. Our trajectories are 20 ps long, which means that a mode oscillating at the frequency of 80 cm\textsuperscript{−1} is sampled $\approx$50 times.

From equation 4.73, we see that the calculation of an infrared spectrum with molecular dynamics simulations is related only to the time-dependent dipole moment of the molecular system. Since the gas phase molecule evolves \textit{on the fly} on the PES (\textit{potential energy surface}), the infrared spectrum takes into account anharmonicities of the PES and of the dipole moment surface (by opposition to the harmonic method for which the description of the potential and of the dipole is restricted to a harmonic function). Also modes couplings are taken into account in this expression by construction.

The main advantages of the molecular dynamics (MD) approach in equation 4.73 for the calculation of infrared spectra (also called "dynamical spectra") can be listed as follows. This has been detailed in two reviews by the Evry group\textsuperscript{172,173}. I am discussing below the main points.

- There are no approximations made in equation 4.73 apart from the hypothesis of linear response theory, \textit{i.e.} a small perturbation from the applied external electric field on the absorbing molecular system. Such condition is always fulfilled in vibrational spectroscopy of interest here. There are no harmonic approximations made, be they on the potential energy surface or on the dipole moment, contrary to the static harmonic calculations (described more in details in section 4.5). These approximations are not needed in equation 4.73. As a consequence, vibrational anharmonicities are naturally taken into account in equation 4.73: one thus needs the knowledge of the time evolution of the dipole moment of the system in order to calculate an anharmonic IR spectrum. This is naturally done with molecular dynamics simulations. In fact, the finite temperature dynamics takes place on all accessible parts of the potential energy surface, be they harmonic or anharmonic. The quality of the potential energy surface is entirely contained in the "ab-initio" force field used in the dynamics, calculated at the DFT/BLYP-D3 level in all works presented here (see section 4.1). The group in Evry has in particular shown that good to excellent agreements can be obtained for the absolute (and relative) positions of the different active bands obtained in dynamical IR spectra with this DFT-MD method, in the gas phase\textsuperscript{172,174–178}, in the liquid phase\textsuperscript{165,179–182}, and at solid-liquid and liquid-air interfaces\textsuperscript{183–185}.

- The calculation of IR spectra with MD is related \textit{only} to the time-dependent dipole moment of the molecular system, neither requiring any harmonic expansion of the transition dipole moment nor the knowledge of normal modes, in contrast to harmonic calculations. Therefore, if the dipole moments and their fluctuations are accurately calculated along the trajectory, the
resulting IR spectrum should be reliable. The vibrations therefore do not directly rely on the
curvature of the potential energy surface at the minima on the PES (i.e. normal modes and
derivatives using these normal modes) but rather on the time evolution of the electric dipole
moment of the molecular system, which is governed by the conformational dynamics at the
finite temperature of the simulation. As a consequence, dynamical anharmonic spectra from
equation 4.73 and harmonic spectra rely on strictly different properties.

- Equation 4.73 gives the whole infrared spectrum of a molecular system in one single
calculation, i.e. the band positions, the band intensities and the band shapes. There are no
approximations applied, in particular the shape and broadening of the vibrational bands result
from the underlying dynamics and mode-couplings of the system at a given temperature. Also,
there are no corrections applied a posteriori, particularly no scaling factor applied for band
positions.

- Dynamics simulations are performed at finite temperature. At a given temperature, and
depending on the energy barriers on the potential energy surface (PES), conformational dy-
namics between different isomeric forms of the absorbing gas phase molecular entity can be
sampled by MD. All conformations populated over time are thus taken into account in the final
calculation of the infrared spectrum from equation 4.73. Beyond the conformational dynamics
between conformers and isomers of the molecule of interest, any population dynamics (around
the minima structures), typically H-bond dynamics, local geometry rearrangements, also give
rise to a natural broadening of the calculated IR active bands, which is essential for the compar-
ison to the experimental spectra. In the present work, simulations are performed at very low
temperature (typically 50K) to be compared with the molecular beam experiment conditions
(experiments described in chapter 3). At this low temperature, we observe no conformational
dynamics between isomeric forms (at the exception of the Z-Ala₆-NH₂ peptide, see chapter 2).
The other advantage of working at finite temperature is that one can somehow probe some
effects around the Zero Point Energy (ZPE) of the vibrational modes (possibly already at an
anharmonic zone on the potential energy surface) even though the classical nuclei representa-
tion used here does not include the quantification of the vibrational levels. One could argue
that in molecular beam conditions, only the v=0 vibrational ground level of the vibrational
modes are populated but even at the ZPE, the anharmonicities of the surface can be significant.

A trajectory can include end-over-end rotation of the molecule that can then couples to the
vibrations, especially for small systems as the phenol derivatives. We removed these rotations
using the Eckart-Sayvetz conditions¹³²,¹³³. As can be seen in figure 4.3, it is important to apply
this correction.

4.3.1 Vibrational modes analyses - VDOS / ICDOS

It is one thing to be able to calculate an infrared spectrum (equation 4.73), it is something
else to unravel the motions behind the peaks. In molecular dynamics simulations, assignment
of the vibrational bands into molecular motions can be achieved through several ways. The
most popular tool is probably VDOS (velocity - Vibrational Density Of States)¹⁸⁶. The VDOS are
obtained through the Fourier Transform of the atomic velocity autocorrelation function and the
Figure 4.3: Theoretical DFT-MD IR spectrum of the phenol molecule. In black, the spectrum resulting from the trajectory including rotation. In the black spectrum, peaks are systematically doubled, because of rotation-vibration couplings. In red, the spectrum resulting from the trajectory of which rotation was removed.

VDOS spectrum, denoted here $I_{\text{V DOS}}(\omega)$, is calculated as:

$$I_{\text{V DOS}}(\omega) = \sum_{i=1}^{M} \int_{-\infty}^{\infty} \langle \vec{v}_i(t) \cdot \vec{v}_i(0) \rangle e^{i\omega t} dt$$

(4.74)

where $i$ runs over all the $M$ atoms of the investigated system. $\vec{v}_i(t)$ is the velocity vector of atom $i$ at time $t$. The $\Sigma$ is over the $M$ atoms of the system. The angular brackets represent a statistical average of the correlation function. The VDOS spectrum provides all vibrational modes of the molecular system whether active in infrared or Raman (or not active at all). The expression in equation 4.74 can also be restrained by replacing $M$ in the summation by a selection of given atoms instead of all the atoms. In that case, the $I_{\text{V DOS}}(\omega)$ spectrum gives the signatures of the selected atoms only.

In the context of this thesis, we chose another strategy for modes assignment by using the Fourier transform of time correlation functions of internal coordinates (bond-lengths, angles, dihedral angles):

$$I_{\text{ICDOS}}(\omega) = \int_{-\infty}^{\infty} \langle IC_i(t) \cdot IC_j(0) \rangle e^{i\omega t} dt$$

(4.75)

where $IC(t)$ is the time evolution of the selected internal coordinate (bond-length, angle and dihedral angle). We only calculate autocorrelation functions in this work ($i = j$). The cross correlations ($i \neq j$) would give us some supplementary knowledge on the couplings between internal coordinates, but we did not calculate and did not analyse them.

In our works, we have found that the ICDOS toolkit is more useful to analyse the vibrational motions in the far infrared domain than the VDOS. These latter indeed do not include any information about the nature of the motions, i.e. stretches, bends, torsions. Only the ICDOS does
include this information through the definition of the internal coordinates. Therefore VDOS
will not be used in this thesis. ICDOS (Internal Coordinates - vibrational Density Of States)
spectra are presented in our paper "Can far IR action spectroscopy combined with BOMB simu-
lations be conformation selective"64, presented in chapter 5. Section 9.1 of chapter 9 presents
the advantages and limitations of the ICDOS analytical tool in details.

Note that neither VDOS nor ICDOS can be compared directly with infrared spectra in terms
of intensity of the bands because the dipole moment is not taken into account in the expression
(necessary for IR selection rules). Dr. Galimberti and Prof. Gaigeot in the group have devised a
simplification method for infrared spectra calculations based on VDOS and including the dipole
selection rules.187

4.4 Geometry optimisation

All the conformations discussed and displayed in the figures of this thesis are optimised struc-
tures. These structures are used as starting points for the DFT-molecular dynamics simulations.
They are also needed to perform harmonic and VPT2 calculations presented in sections 4.5
and 4.6 respectively.

We have described the process to optimise the electronic wavefunction above in subsec-
tion 4.1.5. We can also "optimise geometries", i.e. find minima on the potential energy surface
(PES) of the molecular system of interest. We have used the Gaussian package for geometry
optimisations and the underlying algorithms that find the paths towards minima on the PES.
In practice, the Gaussian package used here for geometry optimisations uses the Berny algo-
rithm188 based on a conjugate gradient algorithm which means that the second derivatives
do not have to be computed at every step but instead the algorithm corrects the Hessian (or
second derivative matrix) from the previous step. For my calculations, the BLYP-D3 electronic
representation is applied, with the gaussian basis set 6-311+G(d,p) is used, an ultrafine grid
(that changes the mesh of the grid for the description of the electronic density) and the "tight"
keyword (modifying the criteria of convergence for the geometry optimisation from the default
adopted in Gaussian).

The algorithms do not differentiate minima and extrema on the PES. To know this, one has
to calculate and diagonalise a Hessian matrix. This matrix gathers all the second derivatives of
the energy along the cartesian coordinates (3M matrix if M=number of atoms):

\[
H = \begin{bmatrix}
\frac{\partial^2 E}{\partial \zeta_i \partial \zeta_j} & \frac{\partial^2 E}{\partial \zeta_i \partial \zeta_k} & \cdots & \frac{\partial^2 E}{\partial \zeta_i \partial \zeta_M} \\
\frac{\partial^2 E}{\partial \zeta_j \partial \zeta_i} & \frac{\partial^2 E}{\partial \zeta_j \partial \zeta_k} & \cdots & \frac{\partial^2 E}{\partial \zeta_j \partial \zeta_M} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 E}{\partial \zeta_M \partial \zeta_i} & \frac{\partial^2 E}{\partial \zeta_M \partial \zeta_j} & \cdots & \frac{\partial^2 E}{\partial \zeta_M \partial \zeta_M}
\end{bmatrix}
\] (4.76)

After diagonalisation of the matrix, if all eigenvalues are positive, the system is located on
a minimum on the potential energy surface, if one eigenvalue is negative, the system is located
on a transition state.
4.5 Harmonic approximation for infrared spectroscopy - Vibrational Analysis in the Gaussian package

Beside the DFT-MD spectra introduced in section 4.3, harmonic vibrational spectra have also been calculated in my work using the Gaussian\textsuperscript{109} package. In these calculations, the BLYP-D3 electronic representation is applied for the best comparison with the DFT-MD spectra. The gaussian basis set 6-311+G(d,p) is used, an ultrafine grid and the "tight" keyword. Note that in the Gaussian package, only a gaussian basis set is used instead of the mixed representation gaussian/plane waves basis sets used in the CP2K package. No pseudopotentials are used, it is an all electron representation. Despite these differences between the Gaussian calculations and the ones in CP2K, we believe that the harmonic/DFT-MD comparisons presented and discussed in chapter 10 are relevant.

Two approximations are made in harmonic spectra calculations, which are used in the equations presented below. First approximation is that the spectrum is calculated for a molecular conformation which corresponds to a minimum on the potential energy surface, i.e. for an optimised geometry (see the previous section 4.4) which has no temperature (0 K) and therefore has an energy which is at a minimum on the potential energy surface (PES). At this point on the PES, the energy is purely harmonic. This is the first (mechanical) harmonic approximation. The second approximation is an electrical approximation related to the expression of the dipole moment (this is used for calculating the IR intensity of each normal mode). The dipole moment is usually expressed as a Taylor expansion with respect to derivatives of the normal modes (normal modes will appear through the diagonalisation of a Hessian matrix, see the matrix in 4.78 below). When this expansion is stopped at the first derivatives, we call it a harmonic expansion. This is what is done in harmonic spectra calculations.

To obtain the harmonic frequencies and harmonic normal modes, we start from the the mass-weighted second derivatives calculated during geometry optimisation and presented in section 4.4 above:

\[
\frac{\partial^2 E}{\partial \zeta_A \partial \zeta_B} \frac{1}{\sqrt{m_i} \sqrt{m_j}}
\]  

(4.77)

with \(\zeta\) representing the x, y and z cartesian coordinates of atom A and \(m_A\) and \(m_B\) the masses of the atoms A and B, respectively.

The \(3N\) (\(N\) is the number of atoms) mass-weighted Hessian matrix is:

\[
H = \begin{bmatrix}
\frac{\partial^2 E}{\partial \zeta_A \partial \zeta_A} \frac{1}{\sqrt{m_A} \sqrt{m_A}} & \frac{\partial^2 E}{\partial \zeta_A \partial \zeta_B} \frac{1}{\sqrt{m_A} \sqrt{m_B}} & \cdots & \frac{\partial^2 E}{\partial \zeta_A \partial \zeta_M} \frac{1}{\sqrt{m_A} \sqrt{m_M}} \\
\frac{\partial^2 E}{\partial \zeta_A \partial \zeta_B} \frac{1}{\sqrt{m_B} \sqrt{m_A}} & \frac{\partial^2 E}{\partial \zeta_B \partial \zeta_B} \frac{1}{\sqrt{m_B} \sqrt{m_B}} & \cdots & \frac{\partial^2 E}{\partial \zeta_B \partial \zeta_M} \frac{1}{\sqrt{m_B} \sqrt{m_M}} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{\partial^2 E}{\partial \zeta_M \partial \zeta_A} \frac{1}{\sqrt{m_M} \sqrt{m_A}} & \frac{\partial^2 E}{\partial \zeta_M \partial \zeta_B} \frac{1}{\sqrt{m_M} \sqrt{m_B}} & \cdots & \frac{\partial^2 E}{\partial \zeta_M \partial \zeta_M} \frac{1}{\sqrt{m_M} \sqrt{m_M}}
\end{bmatrix}
\]  

(4.78)

The Hessian matrix is then diagonalised, giving a set of \(3N\) eigenvectors and \(3N\) eigenvalues. The eigenvectors Q are the normal modes of the molecule.
One could apply the same procedure working in internal coordinates (bond-length, angles, dihedral angles) and a 3N-6 Hessian could be created and diagonalised:

\[
H = \begin{bmatrix}
\frac{\partial^2 E}{\partial q_i \partial q_i} & \frac{\partial^2 E}{\partial q_i \partial q_j} & \ldots & \frac{\partial^2 E}{\partial q_i \partial q_k} \\
\frac{\partial^2 E}{\partial q_j \partial q_i} & \frac{\partial^2 E}{\partial q_j \partial q_j} & \ldots & \frac{\partial^2 E}{\partial q_j \partial q_k} \\
\ldots & \ldots & \ldots & \ldots \\
\frac{\partial^2 E}{\partial q_k \partial q_i} & \frac{\partial^2 E}{\partial q_k \partial q_j} & \ldots & \frac{\partial^2 E}{\partial q_k \partial q_k}
\end{bmatrix}
\] (4.79)

with \(q_i\) and \(q_j\), internal coordinates. After diagonalisation of this matrix, 3N-6 normal modes \(Q\) will be produced, directly expressed in internal coordinates.

The infrared intensity of each normal mode \(Q_i\) (in cartesian or internal coordinates) is calculated as:

\[
I_{\text{harm}} = \left( \frac{\partial \bar{\mu}}{\partial Q_i} \right)^2
\] (4.80)

with \(\bar{\mu}\), the dipole moment of the molecule.

In practice the intensities are usually calculated as small variations of \(\bar{\mu}\) upon small variations of the normal mode:

\[
I_{\text{harm}} = \frac{\delta \bar{\mu}}{\delta Q_i}
\] (4.81)

### 4.6 VPT2 anharmonic spectroscopy in the Gaussian package

VPT2 anharmonic spectra (as implemented in the Gaussian package by the group of Barone\textsuperscript{134–136}) have been calculated in this work and are presented in chapter 10. The gaussian basis set 6-311+G(d,p) is used, an ultrafine grid and the "tight" keyword.

The VPT2 method is applied as a correction to the harmonic approximation (section 4.5). Third order as well as a subset of the fourth order derivatives of the potential energy with respect to the dimensionless internal coordinates \(\{q\}\) are first computed:

\[
\frac{\partial^3 E}{\partial q_i \partial q_j \partial q_k} \text{ and } \frac{\partial^4 E}{\partial q_i \partial q_j \partial q_k \partial q_l}
\] (4.82)

Then fundamentals, overtones and combination bands frequencies are computed using the following equations:

**Fundamentals:** \(v_{i,0 \rightarrow 1} = \omega_i + 2\chi_{ii} + \frac{1}{2} \sum_{j=1}^{N} \chi_{ij} \) (4.83)

**Overtones:** \(v_{i,0 \rightarrow 2} = 2\omega_i + 6\chi_{ii} + \frac{1}{2} \sum_{j=1}^{N} \chi_{ij} \) (4.84)
Combinations: \( \nu_{i,0 \rightarrow j,0 \rightarrow 1} = \omega_i + \omega_j + 2\chi_{ii} + 2\chi_{jj} + 2\chi_{ij} + \frac{1}{2} \sum_{k=1}^{N} [\chi_{ik} + \chi_{jk}] \)  

(4.85)

where the anharmonic \( \chi \) matrix has the form,

\[
16\chi_{ii} = \frac{\partial^4 E}{\partial q_i^4} - \frac{5 (\frac{\partial^3 E}{\partial q_i^3})^2}{3\omega_i} - \sum_{j \neq i}^{N} \frac{(8\omega_i^2 - 3\omega_j^2)(\frac{\partial^3 E}{\partial q_i \partial q_j^3})^2}{\omega_j (4\omega_i^2 - \omega_j^2)}
\]

\[
4\chi_{ij} = \frac{\partial^4 E}{\partial q_i^2 \partial q_j^2} - \frac{2\omega_i (\frac{\partial^3 E}{\partial q_i \partial q_j^3})^2}{4\omega_i^2 - \omega_j^2} - \frac{2\omega_j (\frac{\partial^3 E}{\partial q_i \partial q_j^3})^2}{4\omega_j^2 - \omega_i^2} - \frac{(\frac{\partial^3 E}{\partial q_i \partial q_j^3})(\frac{\partial^3 E}{\partial q_i \partial q_j^3})}{\omega_i} - \frac{(\frac{\partial^3 E}{\partial q_i \partial q_j^3})(\frac{\partial^3 E}{\partial q_i \partial q_j^3})}{\omega_j}
\]

\[
+ \sum_{k \neq i,j}^{N} \left[ \frac{2\omega_k (\omega_i^2 + \omega_j^2 - \omega_k^2)(\frac{\partial^3 E}{\partial q_i \partial q_j \partial q_k})^2}{\omega_k (4\omega_i^2 + \omega_j^2 - 2(\omega_i^2 + \omega_j^2 + \omega_k^2))} - \frac{(\frac{\partial^3 E}{\partial q_i \partial q_j \partial q_k})(\frac{\partial^3 E}{\partial q_i \partial q_j \partial q_k})}{\omega_k} \right] + \sum_{\xi=x,y,z} B_{e^\xi}^2 (C_{RV}^\xi)^2
\]

(4.86)

(4.87)

with \( \omega_i \) the harmonic wavenumber \((\text{cm}^{-1})\) associated with mode \( Q_i \), \( B_{e^\xi} \), the diagonal inertia tensor and \( C_{RV} \) the rovibronic (Coriolis) constant coupling. These three values are computed during the harmonic calculations.

The main problem with this VPT2 approach is that it depends on the accuracy of the initial harmonic approximation with in particular the same vibrational modes.

This approach, developed and implemented by the group of Barone, must not be mistaken with the VSCF\(^{189,190}\) or the VSCF-PT2\(^{191}\) methods. The VSCF-PT2 method is a perturbational correction to the VSCF method as developed by Gerber and Bowman’s groups that are much more advanced and expensive approaches. For the VSCF method, a fourth order polynomial function is usually used instead of harmonic oscillators and the vibrational Hamiltonian takes into account interactions between modes (which makes the use of a SCF procedure mandatory).
Chapter 5

Far infrared spectroscopy of the Ac-Phe-Pro-NH$_2$ dipeptide.

The work presented in this chapter is the second work combining IR-UV ion dip spectroscopy with DFT-MD calculations in the far infrared/THz spectral domain after the seminal paper from Rijs and Gaigeot’s groups: "Gas-Phase Peptide Structures Unraveled by Far-IR Spectroscopy: Combining IR-UV Ion-Dip Experiments with Born-Oppenheimer Molecular Dynamics Simulations"$^{105}$, where the emphasis was to demonstrate that very tiny structural differences could be unraveled using far infrared signatures. See chapter 3 and 4 for a detailed description of the experiments and simulations respectively. Hence, Rijs & Gaigeot’s groups showed that using far infrared signatures one can distinguish between two similar, almost iso-energetic structures that only differ by their peptidic backbone orientations, i.e. C7eq or C7ax that could not be separated using classical infrared ranges. Spectral differences between the two structures could only be seen in the 0-400 cm$^{-1}$ frequency range.

The present paper: "Can far-IR action spectroscopy combined with BOMD simulations be conformation selective?"$^{64}$, that we report hereafter, analyses in details the dipeptide Ac-Phe-Pro-NH$_2$ for which two very different conformations are observed in the gas phase (at least in our experimental conditions), i.e. a γ-turn conformation (C7, the hydrogen bond forms a 7-membered ring) and a β-turn conformation (C10, the hydrogen bond forms a 10-membered ring) both presented in figure 5.1. We study the vibrational signatures possibly impacted by the structural differences between these two turns. The far infrared experimental spectra of this peptide have been measured using IR-UV ion dip spectroscopy and the calculated spectra are by the BLYP-D3-MD method.

The gas phase conformations were already assigned in the paper: "Competition between local conformational preferences and secondary structures in gas-phase model tripeptides as revealed by laser spectroscopy and theoretical chemistry"$^{192}$ from Chin et al., using the 3000-4000 cm$^{-1}$ spectral domain. While the γ-turn conformation was unambiguously assigned, the orientation of the aromatic ring of the phenylalanine amino acid for the β-turn conformation was not assigned in this paper. Therefore we have calculated the BLYP-D3-MD spectra for the three stable orientations of the ring (a, g$^+$, g$^-$) and the match to the experimental spectrum led
Figure 5.1: 3D structures of capped Ac-Phe-Pro-NH$_2$ dipeptides. The 2 caps are 'Ac' for acetyl (CH$_3$-CO) and NH$_2$, respectively for the N- and C-terminals. a): $\gamma$-turn structure. b): $\beta$-turn structure.

Beside conformational assignment, one new contribution in the theoretical analyses introduced in our paper is the definition and the use of the ICDOS tool (Internal Coordinates - vibrational Density Of States). VDOS (velocity -Vibrational Density Of States) are more traditionally used in combination with molecular dynamics simulations$^{52,186}$ in order to extract vibrational motions. While the VDOS spectra are obtained through the Fourier transform of the autocorrelation function of velocities, ICDOS are based on the autocorrelation of internal coordinates. See section 4.3 in chapter 4 for a detailed description of the ICDOS method.

We have found that the ICDOS were more useful to analyse the vibrational motions in the far infrared domain, in particular in the context of this paper. Indeed, VDOS only give information about which atom(s) participate to a normal mode while ICDOS give real insights on the nature of the modes (participations of stretch, bend, torsion, wagging) since internal coordinates are components of the modes.

However the ICDOS decomposition is not fully satisfactory either. In the far IR/THz domain, one wants to decipher local modes from delocalised modes. The distinction between local modes (one main internal coordinate dominates the composition of the mode) and delocalised modes (multiple internal coordinates involved in the mode) is not introduced here but in our subsequent paper: "Mapping gas phase dipeptide motions in the far-infrared and terahertz domain."$^{65}$, see also chapter 6 and 9 of this thesis. ICDOS provides information of the modes in which any chosen internal coordinate is involved, which is especially useful when
one internal coordinate is the main component in one single normal mode. This easily provides the definition of local modes. Local modes for the dipeptides investigated here are found in the 400-800 cm\(^{-1}\) spectral domain. When several internal coordinates have small contributions into one given mode, we then have the definition of one delocalised mode. Delocalised modes are found in the range 0-800 cm\(^{-1}\) for the dipeptides investigated here but the range 400-800 cm\(^{-1}\) is dominated by wagging motions more intense than the delocalised modes. For these delocalised modes, the ICDOS decomposition is however not enough to get a precise insight, providing only a list of internal coordinates involved in the mode. For both local and delocalised modes ICDOS unfortunately lacks a quantitative analysis (in terms of % of each motion involved in each mode). A more detailed discussion on ICDOS spectra for infrared analyses (and their advantages/disadvantages) can be found in chapter 9.

In the attached paper, vibrational modes analyses and conformational assignments are combined to see which modes are affected by the differences in the \(\gamma\) vs \(\beta\)-turn conformations. In a nutshell, we would like to know if some vibrational peaks are conformation selective, i.e. if some peaks specifically allow to distinguish a \(\gamma\)-turn from a \(\beta\)-turn conformation without doing any spectral calculations.

We find that the range 520-800 cm\(^{-1}\) does not appear to be conformation selective since bands predominantly arise from out of plane wagging motions of hydrogen atoms that belong to the phenyl ring or to the proline residue that display similar environments in both \(\gamma\) and \(\beta\)-turn conformations. In chapters 6 and 9, we show that these \(\omega(CH)\) are only weakly conformational selective, for a broad range of molecules investigated in my work.

The 400-550 cm\(^{-1}\) spectral domain is the one that appears to be conformational selective for Ac-Phe-Pro-NH\(_2\). In this range, one can find signatures that mainly arise from the out of plane motions of the hydrogen atoms of the two NH amide functions or from the NH\(_2\) terminal (wagging motions). That \(\omega(NH_2)\) modes are conformation selective is not surprising since the NH\(_2\) function is the one involved into the \(\gamma\)-turn or \(\beta\)-turn interactions. It therefore should be conformation selective indeed. Due to the small sampling in our work (at this stage in time), \(i.e.\) 1 \(\gamma\)-turn conformation, 1 \(\beta\)-turn conformation, and 2 more \(\gamma\)-turn structures if we add the two systems Ac-Phe-Gly-NH\(_2\) and Ac-Phe-Ala-NH\(_2\) from the previous paper\(^{105}\), it is at this point impossible to determine whether the signatures of the Ac-Phe-Pro-NH\(_2\) \(\beta\)-turn that differ from the \(\gamma\)-turn conformation signatures are specific of a \(\beta\)-turn conformation. But, it is already clear that there is no \(\gamma\)-turn specific signatures because of the already large variety of hydrogen bond lengths in the three \(\gamma\)-turn systems already investigated. This is confirmed in our following paper: "Mapping gas phase dipeptide motions in the far-infrared and terahertz domain"\(^{65}\), see chapter 6, that takes into account many more dipeptide systems.

The last discussion in our paper concerns the 0-400 cm\(^{-1}\) range. The striking thing is that the \(\beta\) and \(\gamma\)-turn spectra of Ac-Phe-Pro-NH\(_2\) display a common number of peaks deviating in frequency by only 2-14 cm\(^{-1}\). However, the vibrational assignments of these common peaks are different between the two conformers. For instance, the 306/308 cm\(^{-1}\) peaks, respectively for the \(\gamma\) and \(\beta\)-turns, do not arise from the same motion (different combination of internal
coordinates however related to the backbone in both cases). Nevertheless, as shown in chapters 6 and 10 of this thesis, BLYP-D3-MD spectra for peptidic systems are precise enough to distinguish between these two conformations using this range despite the spectral similarities.

This spectral similarity between $\beta$ and $\gamma$-turn conformations in the 0-400 cm$^{-1}$ range is even more surprising because in the paper "Gas-Phase Peptide Structures Unraveled by Far-IR Spectroscopy: Combining IR-UV Ion-Dip Experiments with Born-Oppenheimer Molecular Dynamics Simulations" that discusses the case of the Ac-Phe-Gly-NH$_2$, it is shown that structural differences (that are much smaller than the differences for the two conformers of Ac-Phe-Pro-NH$_2$) induce large modifications in the delocalised modes in terms of composition and frequency.
Can far-IR action spectroscopy combined with BOMD simulations be conformation selective?†

Jérôme Mahê,‡a Sander Jaeqx,‡b Anouk M. Rijs*bd and Marie-Pierre Gaigeot*ac

The combination of conformation selective far-IR/UV double resonance spectroscopy with Born–Oppenheimer molecular dynamics (BOMD) simulations is presented here for the structural characterization of the Ac-Phe-Pro-NH₂ peptide in the far-infrared spectral domain, i.e. for radiation below 800 cm⁻¹.

Two conformers have been shown to be present in the experiment, namely a conformer with a γ-turn fold (C7 interaction) and a β-turn fold (C10 interaction). The combined experimental and theoretical work presented here aims to provide spectral features typical of each conformer in this far-IR domain. The simulated BOMD far-IR spectra agree well with the experimental spectra and allow direct assignment of the observed bands. These assignments show that the 400–550 cm⁻¹ spectral domain is conformer selective, allowing us to distinguish the H-bond signature of the γ-turn from the β-turn.

1 Introduction

Understanding the forces that govern the complex protein folding process is one of the holy grails in modern biophysical science.1 One way to obtain insights into this process is to study the folding propensities of small isolated peptides by observing the emergence of secondary structures through intramolecular hydrogen bonding.2 Conformation-selective mid-IR spectroscopy in combination with harmonic density functional theory (DFT) calculations has proven to be a powerful tool to this end, and is nowadays applied on a routine-basis.3–6 The far-IR region (radiation < 800 cm⁻¹) has often been ignored, due to the possible deficiencies of theoretical tools needed to interpret the experimental spectra in this domain.7–9 This far-IR regime not only complements the mid-IR one, but also yields information that is not accessible in the mid-IR region. The mid-IR region mostly probes localized structural information, due to the localized character of the vibrations found in this region, such as the structurally diagnostic amide A (NH stretch vibration), amide I (C=O stretch vibration) and amide II (NH in-plane-bending vibration) modes of peptides. On the other hand, the far-IR region is characterized by large-scale delocalized vibrations. These vibrations are expected to contain detailed structural information on the overall structure and are therefore directly diagnostic to various backbone conformations.10–12 Since these vibrations often extend over a large part of the peptide backbone, such delocalized modes are also expected to be important for the dynamical and flexible nature of proteins.13 Additionally, intrinsic hydrogen bond vibrations can be directly probed in the far-IR region,14,15 as previously shown for condensed phase systems using low-frequency FTIR spectroscopy.16,17 In contrast, the mid-IR region can indirectly probe hydrogen bonds through frequency shifts of the amide vibrations. One more advantage of far-IR probing is the possibility to probe larger and more complex molecules. For such molecular systems, the mid-IR region is often spectrally congested due to many overlapping amide bands and thereby limiting the conformational assignment to families rather than to one specific conformation.1,18,19

In these cases the far-IR spectra often still show resolved absorption bands. This is a consequence of the bandwidth of the free electron laser used in these experiments, which is proportional to the output photon energy.20

Synergy between experiments and theoretical calculations is essential to obtain structural information from these low-frequency motions. Static harmonic DFT calculations are known to be insufficient to achieve this task. In a previously published paper,21 we have shown that Born–Oppenheimer molecular dynamics (BOMD) is able to reproduce the far-IR absorption spectra of gas phase peptides, and can therefore be employed to obtain structural information from the far-IR absorption region of peptides. This combination is able to distinguish between subtle differences in peptide conformations, superior to static DFT calculations in combination with mid-IR spectroscopy. For example, far-IR spectroscopy coupled with BOMD simulations could differentiate the axial and equatorial
forms of the γ-turn interaction in Ac-Phe-Gly-NH₂, which was not possible with mid-IR spectroscopy and static (harmonic) DFT calculations. As will be described in the present paper, BOMD simulations take the anharmonic character of the delocalized vibrations directly into account, providing reliable spectroscopic predictions, vibrational assignments and structural interpretations. The combination between experiment and theory provides direct insight into the nature of the low frequency motions.

Apart from our developed far-IR set-up for the investigation of neutral gas phase peptides using the FELIX free electron laser, such spectroscopy has been developed in the Havenith group with investigations specifically focused on the probing of the solvation of biomolecules, see for instance ref. 24–26. In the groups of Plusquellie and Markelz, terahertz studies are performed on condensed phase biological systems, from simple amino acids to complete proteins. Far-IR spectra of tagged gas phase ionic clusters have also been obtained by the Asmis group in Berlin, see their review. Protonated clusters and ionic clusters of atmospheric interest have mainly been investigated with this technique. Lastly, far-IR studies are employed to study the structural properties of metal clusters.

As yet, the far-IR part of the absorption spectrum of gas phase peptides is in many aspects an uncharted territory. To identify the functional vibrations and distinguish them from other bands in the spectrum demands a large understanding of the low frequency modes. We have started such mapping in ref. 21 by combining far-IR experiments and BOMD simulations, and we continue this approach here with Ac-Phe-Pro-NH₂ to assess the conformation selectivity capabilities of our approach. Previous experiments on Ac-Phe-Pro-NH₂ performed in the 3 micron region (probing the NH stretch vibrations) by Mons et al. have shown that the backbone of this peptide can fold either via a β-turn (the backbone adopts a C10 interaction, with an unusual cis conformation for the Phe-Pro peptide bond) or a γ-turn (the backbone exhibits a C5 and C7 interaction). These two conformations co-exist in the molecular beam expansion experiments. Choosing this specific peptide allows us to directly observe the difference between backbone folding, here between the C10 and C7 interactions, in far-IR patterns. In that respect, going from Ac-Phe-Gly-NH₂ and Ac-Phe-Ala-NH₂ (previous work, axial and equatorial C7 γ-turns), to Ac-Phe-Pro-NH₂ (present work, C7/C10 H-bond folding), provides the opportunity to probe different structural motifs whose signatures are identified through far-IR and BOMD vibrational spectroscopy.

2 Methods

2.1 Experimental details

Ac-Phe-Pro-NH₂ (95% purity) was purchased from Genecust (Dudelange, Luxembourg) and used without further purification. Here, the experimental set-up is briefly described. A complete description of the set-up is described elsewhere. The sample was mixed with graphite powder and applied on a solid graphite bar. A pulsed YAG laser operating at 1064 nm (Polaris Pulsed Nd:YAG Laser, NewWave Research) with a pulse energy of about 1.5 mJ was used to desorb the sample molecules from the graphite substrate as intact neutral molecules. The gas-phase molecules are entrained in a supersonic molecular beam of argon, produced by a pulsed valve (Jordan) and a backing pressure of 3 bar of argon. In the molecular beam, the peptide molecules are cooled towards their rotational and vibrational ground state. The molecular beam travels through a skimmer of about 10 cm downstream to enter a differentially pumped chamber equipped with a reflection time-of-flight mass spectrometer. Here, the molecules interact with a UV beam produced by a pulsed Nd:YAG laser (either Innolas GmbH Spittal 1200 or Quanta-Ray Lab Series) coupled to a frequency doubled dye laser (Radiant Dyes NarrowScan, laser dye: coumarin 153). The UV laser was operated at 10 Hz with typical pulse energies of 1–2 mJ. The molecules are 2-photon ionized via a one color (1 + 1) REMPI scheme. Conformation selection is achieved by fixing the UV laser at a specific S₁ ← S₀ transition. Since different conformations have a slightly different electronically excited state energy, they will therefore appear as different peaks in the UV excitation spectrum. The generated ions are accelerated into the time-of-flight tube, and reflected into the detector.

For the IR-UV double resonance spectra, the IR and UV beams were spatially overlapped, but the IR pulse precedes the UV pulse by ~200 ns. The IR radiation is produced by the Free Electron Laser for Infrared eXperiments (FELIX) located in the FELIX laboratory at the Radboud University. The frequency of the UV probe pulse is fixed on a transition producing a constant ion signal. For the γ-turn conformation the UV laser was set at 37435.5 cm⁻¹, while at 37 409 cm⁻¹ for the β-turn conformer (see Fig. S1 of the ESI†). Whenever the IR hole-burn laser excites a transition that shares the same ground state as the probe laser (and thus is the same conformation), a dip in the ion signal is observed since the ground state is depleted by the IR laser. An IR absorption spectrum of a single conformer can thus be constructed by taking the logarithm of the ion signal without the IR pulse divided by the ion signal with IR pulse. To correct for long-term UV power drifts and changing source conditions, alternating IR-on and IR-off signals are measured by operating the IR laser at 5 Hz and the UV laser at 10 Hz. Since the experiments are performed over a very wide frequency range, the IR laser intensity needs to be corrected for the photon energy. A photon flux resulting in a 1 mJ pulse energy at 1000 cm⁻¹ is equal to a photon flux resulting in a 0.1 mJ pulse energy at 100 cm⁻¹, assuming that the laser pulse profile is identical at the two photon energies. Therefore, the observed absorption intensities are multiplied with the photon energy (in wavenumbers) and renormalized to correct for this effect.

2.2 Theoretical details

Our theoretical methodology consists of DFT-based molecular dynamics simulations, performed within the Born–Oppenheimer (BOMD) framework using the CP2K package. The methods and algorithms employed in the CP2K package are described in detail in ref. 33. In our dynamics, the nuclei are treated classically.
and the electrons quantum mechanically within the DFT formalism. Dynamics consist of solving Newton’s equations of motion at a finite temperature, with the forces that act on the nuclei derived from the Kohn–Sham energy. In BOMD, the Schrödinger equation for the electronic configuration of the system is solved at each time step of the dynamics. Mixed plane waves and gaussian basis sets are used in CP2K. Only the valence electrons are taken into account and pseudo-potentials of the Goedecker–Tetter–Hutter (GTH) form are used.\textsuperscript{35–37} We use the Becke, Lee, Yang and Parr (BLYP) gradient-corrected functional\textsuperscript{38,39} for the exchange and correlation terms. Dispersion interactions have been included with the Grimme D3 corrections.\textsuperscript{40} Calculations are restricted to the Γ point of the Brillouin zone. We employed plane-wave basis sets with a kinetic energy cut-off of 450 Ry and gaussian basis sets of aug-TZV2P type. The cubic box size is of 20 Å length. The kinetic energy cut-off, basis set size and cubic box size have been selected subsequent to energy convergence tests.

The first 3 ps of the trajectory was used for the thermalisation of the system with temperature control through velocity rescaling. Hereafter, pure NVE trajectories were accumulated over 20 ps for the IR spectra calculations and trajectory analyses. Periodic boundary conditions were applied (neutral molecule). The time step in the simulations is 0.4 fs. The temperature of the trajectories was 48 ± 4 K and 53 ± 5 K, respectively, for the γ-turn and β-turn conformers.

Within statistical mechanics and linear response theory,\textsuperscript{41,42} an infrared spectrum can be calculated by the Fourier transform of the time correlation function of the fluctuating dipole moment vector of the absorbing molecular system:

\[
I(\omega) = \frac{2\pi\beta\alpha_0^2}{3eV} \int_{-\infty}^{\infty} dt \langle \delta \mathbf{M}(t) \cdot \delta \mathbf{M}(0) \rangle \exp(i\omega t) \tag{1}
\]

where \( \beta = 1/\hbar T \), \( T \) is the temperature, \( \epsilon \) is the speed of light in a vacuum, \( V \) is the volume. The angular brackets represent a statistical average of the time correlation function of the dipole vector, where \( \delta \mathbf{M}(t) = \mathbf{M}(t) - \langle \mathbf{M} \rangle \) with \( \langle \mathbf{M} \rangle \) the time average of \( \mathbf{M}(t) \). The calculation in eqn (1) is done in the absence of an applied external field. For the prefactor in eqn (1), we have taken into account an empirical quantum correction factor (multiplying the classical line shape) of the form \( \beta\hbar\omega / (1 - \exp(-\beta\hbar\omega)) \), which was shown by us and others to give accurate results on calculated IR intensities.\textsuperscript{43–45} For more detailed discussions on quantum corrections, see for instance ref. 46–49.

The main advantages of the molecular dynamics (MD) approach in eqn (1) for the calculation of infrared spectra (also called “dynamical spectra” in the remainder of the text) are discussed in detail in our review\textsuperscript{50} and are briefly listed as follows:

- There are no approximations made in eqn (1) apart from the hypothesis of linear response theory, i.e. a small perturbation from the applied external electric field on the absorbing molecular system. Such conditions are always fulfilled in vibrational spectroscopy of interest here. There are no harmonic approximations made, be they on the potential energy surface or on the dipole moment, in contrast to the usual static calculations used in the literature. These approximations are not needed in eqn (1).

- As a consequence, vibrational anharmonicities are naturally taken into account in eqn (1): one thus only needs the knowledge of the time evolution of the dipole moment of the system in order to calculate an anharmonic IR spectrum. This is naturally achieved with molecular dynamics simulations. In fact, the finite temperature dynamics takes place on all accessible parts of the potential energy surface, be they harmonic or anharmonic. The quality of the potential energy surface is entirely contained in the “\textit{ab initio}” force field used in the dynamics, calculated at the DFT/BLYP + dispersion level in the work presented here. The good to excellent agreement of the absolute (and relative) positions of the different active bands obtained in our theoretical works (see for instance dynamical spectra in the gas phase,\textsuperscript{51–55} in the liquid phase,\textsuperscript{43,56–59} and at solid–liquid and liquid–air interfaces\textsuperscript{60–62}) is a demonstration that this level of theory is correct.

- Crucial to the present discussion, the calculation of IR spectra with MD is related \textit{only} to the time-dependent dipole moment of the molecular system, neither requiring any harmonic expansion of the transition dipole moment nor the knowledge of normal modes, in contrast to harmonic calculations. Therefore, if the dipole moments and their fluctuations are accurately calculated along the trajectory, the resulting IR spectrum should be reliable. The vibrations therefore do not directly rely on the curvature of the potential energy surface at the minima on the PES (i.e. normal modes and derivatives using these normal modes used in static DFT calculations) but rather on the time evolution of the electric dipole moment of the molecular system, which is governed by the conformational dynamics at the finite temperature of the simulation. As a consequence, dynamical anharmonic spectra from eqn (1) and harmonic spectra rely on strictly different properties, and presumably require different levels of accuracy for the evaluation of these properties.

- Eqn (1) gives the whole infrared spectrum of a molecular system in one single calculation, \textit{i.e.} the band positions, the band intensities and the band shapes, through the Fourier transform of a time correlation function. There are no approximations applied, in particular the shape and broadening of the vibrational bands result from the underlying dynamics and mode-couplings of the system at a given temperature.

No scaling factors of any kind are applied to the vibrations extracted from the dynamics. The sampling of vibrational anharmonicities, \textit{i.e.} potential energy surface, dipole anharmonicities, mode couplings, anharmonic modes, is included in our simulations, by construction, and the application of a scaling factor to the band positions is therefore unnecessary. As reviewed in previous papers, excellent agreements between dynamical spectra and IR-MPD, IR-PD and IR-UV ion dip experiments have been achieved. Any remaining discrepancies between dynamical and experiment spectra should mainly be due to the choice of the DFT/BLYP + dispersion functional as DFT-based dynamics are only as good as the functional itself allows.

The length of the trajectory is related to the vibrational domain to be sampled. One has to keep in mind that the
time-length has to be commensurate with the investigated vibrational motions. Hence, trajectories at around 5 ps are just enough in order to sample stretching motions in the high frequency domain of 3000–4000 cm⁻¹, provided that several trajectories starting from different initial conformations (structure and/or velocities) are accumulated and averaged for the final IR dynamical spectrum. In the mid-IR domain, trajectories of at least 10 ps each are needed in order to sample the slower stretching and bending motions of the 1000–2000 cm⁻¹ domain. In the far-IR below 1000 cm⁻¹ of interest to the present work, longer trajectories are needed in order to properly sample the much slower motions typical of that domain, i.e. torsional motions and possibly opening/closure of structures, typical of peptide chains. In Fig. 1, we have reported the dynamical IR spectrum of the γ-turn conformer calculated at each 5 ps of trajectory over 20 ps, in the critical 100–400 cm⁻¹ lower frequency part. One can see that a 20 ps trajectory already allows an excellent convergence of the dynamical IR spectrum (a 100 cm⁻¹ motion is sampled 60 times in that time period).

An accurate calculation of anharmonic infrared spectra is one goal to achieve, the assignment of the active bands into individual atomic displacements or vibrational modes is another one. This issue is essential to the understanding of the underlying molecular, structural and dynamical properties. In molecular dynamics simulations, the interpretation of the infrared active bands into individual atomic displacements traditionally relies on Fourier transforms of time correlation functions based on velocities or on positions.⁵⁰ Because of the intrinsic nature of delocalised and coupled modes in the far-IR spectral range, we have adopted here a strategy of assignments in terms of Fourier transforms of intramolecular coordinate (IC) time correlation functions, named ICDOS in the rest of the paper, following our previous work on far-IR spectra.²¹

\[
I_{\text{ICDOS}}(\omega) = \int_{-\infty}^{\infty} \langle \text{IC}(i) \cdot \text{IC}(0) \rangle \exp(i\omega t) dt.
\]

(2)

Fig. 2 provides a scheme of the PhePro molecule together with the labelling of atoms and definitions of dihedral angles

discussed below. As the spectral domain <1000 cm⁻¹ is related to large amplitude motions, we have chosen intramolecular coordinates such as dihedral angles (from the backbone, see \(\phi, \psi\) and \(\omega\) in Fig. 2 and Table S2 [ESI†], from the side chain that carries the phenyl ring, see \(\chi_1\) and \(\chi_2\) in Fig. 2), out-of-plane (wagging) motions of H atoms that belong to the phenyl ring or to the proline residue (labelled dihedral-ring-H \(\gamma_{\gamma}^{\text{ring}}\) and dihedral-pro-H \(\gamma_{\gamma}^{\text{pro}}\) in Fig. 2 and 5). We have also calculated the vibrational signature of two coordinates directly related to the C10/C7 hydrogen bond motion, the H-bond stretching defined as the NH \( \cdot \cdot \cdot \) O=C distance, and the dihedral angle H\(_{\text{bond}}\)=N–C–C where H\(_{\text{bond}}\) is the NH \( \cdot \cdot \cdot \) atom hydrogen bonded to the C=C group. We have also calculated the vibrational signature of the dihedral angle H\(_{\text{free}}\)=N–C–C where H\(_{\text{free}}\) is the free hydrogen atom of the NH \( \cdot \cdot \cdot \) group. The signature of the wagging motion of the backbone-N=H amide group (labelled dihedral–H \( \gamma_{\gamma}^{\text{ring}}\) in Fig. 5) has also been calculated.

3 Results

3.1 Conformational assignment

The REMPI spectrum of Ac-Phe-Pro-NH\(_2\) is shown in the ESI† (Fig. S1). The spectrum closely resembles the one previously measured by Mons et al.²⁵ We recorded the IR spectra of the major peaks observed in the spectrum and found two different IR spectra. This suggests that two conformations of Ac-Phe-Pro-NH\(_2\) are present in our molecular beam expansion experiment. Fig. 3 shows the IR spectra of the γ-turn (red line) and β-turn (blue line) recorded from 1850 down to 100 cm⁻¹ with the UV frequency fixed at 37435.5 and 37 409 cm⁻¹, respectively. Each IR spectrum shows intense bands throughout the complete IR region, and both spectra show many well resolved features down to the far-IR region where we observed narrow peaks with a FWHM of about 3 cm⁻¹, limited by the bandwidth of the free electron laser.

The mid-IR region is commonly used to identify peptide structure(s). Here, the amide I region of the γ-turn conformer is
composed of three clear peaks between 1780–1610 cm\(^{-1}\), originating from the three backbone C–O stretch modes. In contrast, only two peaks are observed for the β-turn peptide. Here, the three C–O stretching modes lie too close to each other to be resolved. The three C–O stretch frequencies are calculated to differ by around 22 cm\(^{-1}\). Considering a FWHM of 1–2% for the FELIX IR source, it is not surprising that these absorption bands are not fully resolved. The medium intense band observed at about 1580 cm\(^{-1}\) (γ-turn) and 1600 cm\(^{-1}\) (β-turn) results from the NH\(_2\) scissor vibration.

The experimental spectrum shown in red in Fig. 3 is readily assigned to the gamma-turn (β-γ\(_L\)) conformation of PhePro with the phenyl group in the “a orientation”. Here, the “β” does refer to a C\(_5\) interaction which is responsible for the formation of β-sheets in protein structures. This is the lowest energy structure found in our conformational search and it has the best match in the 1000–1850 cm\(^{-1}\) region, see Fig. S2 (ESI†). Mons et al. have also assigned the “a phenyl orientation” to the structure based on the Franck-Condon patterns in the REMPI spectrum.\(^{29}\)

For the spectrum shown in blue in Fig. 3, the assignment is not that straightforward. This spectrum was previously assigned to a type VIa β-turn conformation with a cis conformation of the Phe-Pro peptide bond by Mons et al.\(^{29}\) The “VIa” is used to classify different types of β-turn conformations. This is also our conclusion from the present work. However, the orientation of the phenyl ring was not discussed in detail in that paper. We therefore performed geometry optimizations of the three different a, g\(_+\) and g\(_-\) possible orientations of the phenyl group in the β-turn geometry, and we calculated the associated harmonic IR spectra of the mid-IR region (1000–1800 cm\(^{-1}\), with the B3LYP functional) and for the 3 μm region (3000–400 cm\(^{-1}\), with the B97-D functional), including mode dependent scaling factors.\(^{63}\) We also performed BOMD simulations (with the BLYP+D3 functional) in order to get the anharmonic far-IR spectra of these three phenyl orientations in the β-turn geometry of PhePro, see Fig. S3 (ESI†). The goal is to assign one of these orientations to the experimental spectra. Note that the energies of the a, g\(_+\) and g\(_-\) orientation of the phenyl group in the β-turn optimized structures are 1.22 kcal mol\(^{-1}\), 3.85 kcal mol\(^{-1}\) and 3.79 kcal mol\(^{-1}\), respectively, using the B97-D/6-311G+(d,p) level of theory (the 0 kcal mol\(^{-1}\) is assigned to the γ-turn structure).

These calculations clearly show that the phenyl group is in its “a orientation”, see Fig. S3, S4 and Table S1 (ESI†). For discarding the “g\(_+\)” orientation”, the most convincing evidence is found in the 3 μm range. As Table S1 (ESI†) shows, in this orientation the NH backbone group of Phe interacts with the phenyl ring, causing a strong red shift for the NH stretch vibration. This shift is not observed in the experimental spectrum. The “g\(_-\)” orientation” can be excluded due to the far too intense
absorption peaks in the far-IR at 408 cm\(^{-1}\) predicted by the BOMD simulations (Fig. S3, ESI\(^{\dag}\)). The “a orientation” also reproduces the peak patterns in the 1000–1400 cm\(^{-1}\) range (Fig. S4, ESI\(^{\dag}\)). The assigned \(\gamma\) and \(\beta\)-turn structures with the “a-orientation” of the phenyl group have been further investigated in our BOMD simulations.

### 3.2 BOMD spectra

Fig. 4 presents the experimental and theoretical IR spectra of the \(\gamma\)-turn (top) and \(\beta\)-turn (bottom) conformers of PhePro in the 100–400 cm\(^{-1}\) region and 400–800 cm\(^{-1}\) region respectively. Note that the absorption intensity is lower below 400 cm\(^{-1}\) and the scales in Fig. 4 are adjusted accordingly.

We remind that the dynamical theoretical spectra have not been adjusted in any way (neither band positions, nor band-widths and shapes). The first observation is that theory and experiment bear remarkable agreement. The theoretical spectra display a number of peaks, positions, band-shapes and intensities that are indeed in good to excellent agreement with the experiments in this anharmonic far-IR frequency range. A few theoretical bands are however too broad in comparison to the experiments and there are also a few bands that either lack intensity or on the contrary carry too much intensity in the theoretical spectra. We will come back to these issues in the discussion. The agreement between experiment and theory is especially remarkable for the \(\gamma\)-turn conformer. The theoretical spectrum of the \(\beta\)-turn shows deficiencies with respect to the experiment, mainly in the higher energy region, that are not observed in the theoretical spectrum of the \(\gamma\)-turn.

One can also observe from the experiments and calculations that the \(\beta\)-turn and \(\gamma\)-turn conformers have different signatures in the 100–800 cm\(^{-1}\) domain, thus this region certainly allows us to distinguish both conformers from their IR spectra alone in the far-IR range.

Fig. 5 reports band assignments of the dynamical IR spectra following the method described in Section 2. These assignments are discussed hereafter.

From the experimental and theoretical spectra, one can observe that the 550–800 cm\(^{-1}\) IR spectral region does not appear to be very much conformer selective, as the number of peaks and their positions are very similar between the \(\beta\)- and \(\gamma\)-turn conformers. Making a one-to-one comparison of the position of the peaks between the two conformers in this region, the maximum band-shift that can be observed is roughly 16 cm\(^{-1}\), with an average difference in the peak positions between the two conformers of \(\sim 7\) cm\(^{-1}\).

---

**Fig. 5** Decomposition of the vibrational bands of the \(\gamma\)-turn conformer (top) and the \(\beta\)-turn conformer (bottom) of PhePro in terms of delocalised motions, using ICDOS, the chosen IC are dihedral motions, C7 H-bond motion and NH\(_2\) free H atom dihedral motion. Dihedrals \(\phi\), \(\psi\), \(\omega\) and \(\chi\) are defined in Fig. 2. See also eqn (2) for the definition of ICDOS.
That this domain does not appear conformer selective can be very well understood by the band assignments, as we find that the 550–800 cm\(^{-1}\) bands predominantly arise from out-of-plane wagging motions of H atoms, for both \(\beta\)- and \(\gamma\)-turn conformers of PhePro. These waggings are either due to H atoms that belong to the phenyl ring or H atoms belonging to the proline ring residue. As can be seen in the 3D structures in Fig. 3, the two conformers display very similar environments around the Pro and Phe rings, so that one would indeed expect very similar signatures of the H waggings on these two rings, whether they belong to the \(\beta\)- or \(\gamma\)-turn conformers.

The 700–800 cm\(^{-1}\) domain is hence composed of a mixing of out-of-plane wagging motions of H atoms that belong to the phenyl ring and of H atoms that belong to the Pro ring. The band located at 700 cm\(^{-1}\) is on the contrary predominantly arising from the out-of-plane waggings of H atoms that belong to the phenyl ring. One can see that the signatures arising at 705, 738 and 765 cm\(^{-1}\) for the \(\gamma\) turn (702, 717 and 748 cm\(^{-1}\) for the \(\beta\) turn) from H atoms that belong to the phenyl ring were also present in our previous work about PheAla and PheGly systems.\(^{21}\) These features seem to be specific to the aromatic ring. The peaks located at 648 cm\(^{-1}\) for the \(\gamma\)-turn and 641 and 653 cm\(^{-1}\) for the \(\beta\)-turn are due to out-of-plane wagging motions of H atoms of the proline residue.

The C10 (\(\beta\)-turn) and C7 (\(\gamma\)-turn) hydrogen bond signature seen through the torsional motion of the H-bonded hydrogen atom of NH\(_2\) (signature of H\(_{\text{bond}}\)-N=C=C in Fig. 5) also appears in this region. It explains the 633 cm\(^{-1}\) peak for the \(\gamma\) turn while this H-bond signature overlaps with the out-of-plane wagging motions of H atoms of the proline residue for the \(\beta\)-turn conformer. The two thin peaks at 563 and 578 cm\(^{-1}\) for the \(\gamma\)-turn conformer and the broader peaks at 569 and 591 cm\(^{-1}\) for the \(\beta\)-turn are also due to the out-of-plane wagging motions of H atoms of the proline residue. The participation of the C10 H-bond is also appearing in these bands for the \(\beta\)-turn conformer. One peak of the doublet lacks intensity in the dynamical spectrum of the \(\gamma\)-turn conformer, while both peak intensities of the \(\beta\)-turn conformer are underestimated in the dynamical spectrum.

The 400–550 cm\(^{-1}\) spectral domain appears conformer selective, providing distinct signatures for the \(\beta\)- and \(\gamma\)-turns, as will be discussed below, although the general spectral features in this domain are similar between the two conformers. Both conformers hence give rise to a triplet between \(~490–540\) cm\(^{-1}\) and to a doublet between \(~400–470\) cm\(^{-1}\). Note that only the higher frequency band of the triplet (518 cm\(^{-1}\) for the \(\gamma\)-turn, 537 cm\(^{-1}\) for the \(\beta\)-turn) shows a substantial 19 cm\(^{-1}\) upshift when comparing the spectra of the two conformers. On the contrary, the doublet peaks are substantially shifted in position between the two conformers, with the two signatures being upshifted by +38 and +29 cm\(^{-1}\) for the \(\beta\) conformer with respect to the \(\gamma\) conformer. It is also interesting to note that all peaks (triplet and doublet) in this domain are substantially more intense in the spectrum of the \(\gamma\)-turn conformer.

For the \(\gamma\)-turn, only a single intense peak is predicted in place of the experimental triplet, although much lower intensity peaks can be seen in the tail of the intense peak, while the triplet is correctly predicted for the \(\beta\)-turn. The doublet of the \(\beta\)-turn from the dynamical spectrum is blue-shifted from experiment. Interestingly, the subtle details of the \(\gamma\)-turn doublet are well predicted in the dynamical spectrum, up to the 475 cm\(^{-1}\) tail peak. We find that the triplet and doublet features are due to N–H torsional/out-of-plane motions, whether N–H belongs to the NH\(_2\) function (free and H-bonded N–H signatures in this domain) or to the amide backbone N–H. The triplet in the \(\gamma\)-turn arises from a combination of all N–H signatures, merging into one single peak in the BOMD spectrum, in contrast to the experiment. The triplet in the \(\beta\)-turn is on the contrary solely due to the free N–H group of the NH\(_2\) terminus of the peptide. For both \(\beta\)- and \(\gamma\)-turns, the doublet clearly reflects the signature of the free N–H of the NH\(_2\) function, while this signature is also overlapping with the NH···O H-bond signature for the \(\beta\)-turn conformer.

The doublet is thus directly (\(\beta\)-turn)/indirectly (\(\gamma\)-turn) related to the C10/C7 hydrogen bond. It is clear from Fig. 6 where the evolution with time of the O···H bond of the C10 \(\beta\)-turn (blue line) and of the C7 \(\gamma\)-turn (red line) conformers is reported, that the average H-bond length and the fluctuations around the average differ between the two conformers (same temperature for the two trajectories). While this H-bond is 2.01 ± 0.08 Å on average for the \(\gamma\)-turn conformer, it is 1.95 ± 0.09 Å, on average for the \(\beta\)-turn conformer (see Table S2, ESI†). This shows how the C10 ring of the \(\beta\)-turn is more tightly H-bonded than the C7 ring of the \(\gamma\)-turn. This difference directly reflects the strength of the hydrogen bond, with the C7 \(\gamma\)-turn H-bond being weaker. With this strength difference in mind, one is therefore not surprised that the signature of the C7 H-bond in the \(\gamma\)-turn conformer appears at lower frequencies, and that the C10 has a more direct signature in the doublet assignment than the C7. Experiments and dynamical spectra show that the C10 H-bond signature is up-shifted from the C7 signature, roughly by ~30–40 cm\(^{-1}\), providing a distinct signature of the \(\gamma\) versus \(\beta\)-turn for the PhePro peptide.

One more comment about the weaker C7 H-bond (\(\gamma\)-turn) with respect to the C10 one (\(\beta\)-turn). It is well-known that GGA
functionals such as BLYP underestimate H-bond strengths, which consequently might allow too large amplitude motions of the hydrogen atoms involved in the hydrogen bond. This subsequently might lead to too large associated vibrational bands, which is indeed observed in the 450–480 cm\(^{-1}\) doublet in the dynamical spectrum of the \(\beta\)-conformer. These two bands also have lower intensities in the experimental spectra, showing that our representation presumably overestimates the H-bond motions in the \(\beta\)-turn C10 interaction. This is also observed for the C7 interaction of the \(\gamma\)-turn, but to a lower extent.

To understand the spectral domain below 400 cm\(^{-1}\), we have used the ICDOS vibrational signatures (Internal Coordinate Density Of States) of backbone torsional motions, of the hydrogen bond and of the out-of-plane motion of H atoms that belong to the NH\(_2\) functional group. This is shown in Fig. 5. First of all, the out-of-plane motion of the hydrogen atoms that belong to NH\(_2\) function do not seem to be relevant to explain the low-frequency vibrational features below 350 cm\(^{-1}\) since they do not show clear activity in this region. Only the peak at 378 cm\(^{-1}\) for the \(\gamma\)-turn is explained by the out-of-plane motion of the free hydrogen atom of the NH\(_2\) function. Second, as the C7/C10 hydrogen bond leads to the folding of the peptide, we expect that both backbone motions and hydrogen bond motions are strongly coupled. This is indeed observed for the two conformers, as the hydrogen bond stretching signature and dihedral backbone motion signatures share common features in the ICDOS spectra.

The spectra of the \(\beta\)- and \(\gamma\)-turns do not appear to be strongly conformer selective in the 240–400 cm\(^{-1}\) range. There are 7 peaks in common between the two spectra, deviating in position by only 2–14 cm\(^{-1}\). Despite these similarities in the IR signatures, the assignments of the bands are surprisingly rather different. Although the intrinsic nature of the motions is similar, namely backbone motions, the coupling between these motions is different due to the two different backbone 3D folded structures in the \(\gamma\) - and \(\beta\)-turns. For instance, the 306/308 cm\(^{-1}\) peak, respectively, for the \(\gamma\)- and \(\beta\)-turns, does not arise from the same motion. The former is related to the \(\Phi_{\text{pro}}\) torsional motion, while the latter comes from \(\Phi_{\text{phe}}\) torsion. Also, the two peaks located at 378 cm\(^{-1}\) for the \(\gamma\)-turn and 386 cm\(^{-1}\) for the \(\beta\)-turn are explained by different motions: the out-of-plane motion of the NH\(_2\) free hydrogen atom for the \(\gamma\)-turn, and coupled \(\omega_{\text{pro}}\) and \(\psi_{\text{phe}}\) motions for the \(\beta\)-turn. Of special interest are the two peaks (experiment and dynamical spectra) recorded in the 250–300 cm\(^{-1}\) domain that carry \(\gamma\)- versus \(\beta\)-turn selectivity. They, respectively, differ by +17 and +14 cm\(^{-1}\) (from lower to higher frequency), going from \(\gamma\)- to \(\beta\)-conformer spectra: these two peaks are due to the amide peptide backbone motions namely \(\Phi_{\text{pro}}\) (lower frequency) and \(\Phi_{\text{phe}}\) (higher frequency). These backbone motions are direct probes of the C7/C10 folding of PhePro, thus providing conformer specific spectral signatures.

Although the H-bond strength is clearly reflected by the peak position of the \(\gamma\) and \(\beta\)-turn conformers, the band pattern of the C7/C10 H-bond is very similar below 200 cm\(^{-1}\). The H-bond signatures dominate the spectral assignments together with couplings to backbone torsions which are directly involved in the H-bond motions. There are four dominant bands related to this H-bond motion in the far-IR part of the spectra, respectively, located at 115 cm\(^{-1}\), 140 cm\(^{-1}\), 163 cm\(^{-1}\) and 174 cm\(^{-1}\) in the BOMD spectrum of the \(\gamma\)-turn and 111 cm\(^{-1}\), 143 cm\(^{-1}\), 148 cm\(^{-1}\) and 175 cm\(^{-1}\) for the \(\beta\)-turn.

4 Conclusions and outlook

In this work, far-IR spectroscopy is shown to be a relevant tool for the characterization of peptide structures, since the observed far-IR features are a direct result from peptide backbone motions and hydrogen bond vibrations, and thereby directly reflect the secondary structure of peptides. The relationship between the delocalized backbone motions and the functional flexibility/rigidity of peptides and proteins can thus be probed with this approach.

The experimental spectra reported here for two conformers of the Ac-Phe-Pro-NH\(_2\) peptide show very well-defined and well-resolved peaks from 800 down to 120 cm\(^{-1}\). To fully exploit the far-IR region and to retrieve the structural information hidden in this region, reliable calculations are key. The presented combination of conformation selective far-IR/UV double resonance spectroscopy with Born–Oppenheimer molecular dynamics simulations brings this synergy.

The present work is a follow-up on our previous initial demonstration\(^{21}\) that such synergy was indeed able to differentiate the axial and equatorial forms of the \(\gamma\)-turn interaction in Ac-Phe-Gly-NH\(_2\), which was not possible with mid-IR spectroscopy and harmonic DFT calculations.\(^{22,23}\) Here, we have investigated two different turns in the conformation of Ac-Phe-Pro-NH\(_2\), i.e. \(\gamma\) (C7 H-bond interaction) and \(\beta\)-turns (C10 H-bond interaction), and have highlighted their specific far-IR signatures.

Clearly, the BOMD dynamical spectrum of the \(\gamma\)-turn conformation of Ac-Phe-Pro-NH\(_2\) provides a better agreement with the experiment than the dynamical spectrum of the \(\beta\)-turn. This is especially true below 350 cm\(^{-1}\) where the H-bond signatures are present. One reason might be the use of the GGA/BLYP functional (although augmented here by D3 van der Waals interactions), known to underestimate the strength of H-bonds, and thus allowing too large amplitude motions of the hydrogen atoms involved in H-bonds. One has also to keep in mind that nuclei quantum effects, especially of relevance to hydrogen atom motions, have not been taken into account in the present simulations. Such effects might help reduce some of the band-breadths, observed for H-bonded motions of the \(\beta\)-turn conformer.

One main purpose of the combined experiment/BOMD spectra simulations was to provide conformer selective IR signatures of the \(\gamma\) versus \(\beta\)-turns in the far-IR region. We have shown that the 400–550 cm\(^{-1}\) domain indeed provides such a distinction between the two conformers. This domain is predominantly due to N–H motions, indirectly probing the NH···O H-bond motion. We have shown that there is a 29–38 cm\(^{-1}\) downshift in the positions of the associated bands for the weaker C7.
interaction in the γ-turn conformer with respect to the C10 interaction in the β-turn.

The 800–550 cm\(^{-1}\) spectral domain was shown not to be conformer selective as bands in both conformers have very similar positions and same assignments from out-of-plane H atom motions of the phenyl ring and pro residue, not sensitive to the γ/β turns. Also the backbone torsional domain in the 250–400 cm\(^{-1}\) does not provide much conformer selectivity. Only the 250–300 cm\(^{-1}\) peaks carry γ versus β-turn selectivity, as they, respectively, differ by +17 and +14 cm\(^{-1}\) (from lower to higher frequency) from the γ to the β-conformer. These two signatures are due to the amide peptide backbone motions, namely \(\Phi_{\text{pro}}\) (lower frequency) and \(\Phi_{\text{phe}}\) (higher frequency), which are direct probes of the C7/C10 folding of PhePro. The supplementary H-bond signatures below 350 cm\(^{-1}\) have been shown to be of limited use for conformer selectivity.

Comparing the results of the γ-turn conformer of PhePro presented here with our previously published results on PheGly and PheAla γ-turns,\(^{21}\) one worth comment concerns the H-bond length and its signature within the three systems. This H-bond length is substantially shorter in PhePro (2.01 Å for the γ-turn and 1.95 Å for the β-turn) than in PheGly (2.14 Å) and PheAla (2.12 Å). As a result, this leads to higher frequencies associated with the H-bond stretching vibration for PhePro (140–150 cm\(^{-1}\) domain) than for PheGly and PheAla (~130 cm\(^{-1}\) domain). Probably due to the stronger H-bond strength, the H-bond stretching vibration then couples more strongly to backbone torsional vibrations, as can be seen from the hydrogen bond stretching activity at higher frequencies (e.g. 380 cm\(^{-1}\)) in the γ-turn of PhePro. In PheGly and PheAla, those bands were not observed.

Acknowledgements

Computational resources were provided by support from HPC resources from GENCI-France (Grant 2012-2014 [072484]). This work is part of the research programme of the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). Part of the FELIX laboratory is financed through the NWO BIG-programme. The calculations were sponsored by NWO Physical Sciences (EW) for the use of the supercomputer facilities at SurfSara. We also want to acknowledge Prof Michel Mons for providing the initial structure for the β-turn and for very fruitful discussions. We also acknowledge the very helpful comments from the reviewers, which significantly improved the quality of this paper.

References

Can far-IR action spectroscopy combined with BOMD simulations be conformation selective?

Jerome Mahé, a Sander Jeaqx, b Anouk M. Rijs, a,b Marie-Pierre Gaigeot ac*

a Laboratoire Analyse et Modélisation pour la Biologie et l’Environnement, LAMBE UMR CNRS 8587, Université d’Evry val d’Essonne, Blvd F. Mitterrand, Bat Maupertuis, 91025 Evry, France

b Radboud University Nijmegen, Institute for Molecules and Materials, FELIX Facility, Toernooiveld 7, 6525 ED Nijmegen, the Netherlands

c Institut Universitaire de France IUF, 103 Blvd St Michel, 75005 Paris, France

* corresponding authors: mgaigeot@univ-evry.fr and a.rijs@science.ru.nl

Supporting Information

Table of content

Figure SI-1: REMPI spectrum Ac-Phe-Pro-NH₂……………………………………………………S2

Figure SI-2: Theoretical (DFT / B3LYP) and experimental spectra for the γ-turn conformation with different phenyl orientations in the 1000 – 1850 cm⁻¹ region……………S3

Figure SI-3: Born-Oppenheimer molecular dynamics spectra for the β-turn conformation with different phenyl orientations……………………………………………………..S4

Figure SI-4: Theoretical (DFT / B3LYP) and experimental spectra for the β-turn conformation with different phenyl orientations in the 1000 – 1850 cm⁻¹ region……………S5

Table SI-1: Theoretical (DFT / B97-D) and experimental vibrational frequencies for the β-turn conformation with different phenyl orientations in the 3 μm region………………..S6

Table SI-2: Values of dihedral angles of the observed β-turn and γ-turn conformations….S7
Figure SI-1: REMPI spectrum Ac-Phe-Pro-NH$_2$. 
Figure SI-2: Experimental (black) and harmonic DFT / B3LYP / 6-311+G(d,p) spectra for the γ-turn conformation with various phenyl ring orientations. The energies are calculated using DFT / B97-D / 6-311+G(d,p).
Figure SI-3: Dynamical BOMD spectra (black) and experimental spectra (red) in the far-IR region for the β-turn conformation with various phenyl ring orientations.
Figure SI-4: Experimental (black) and harmonic DFT / B3LYP / 6-311+G(d,p) spectra for the β-turn conformation with various phenyl ring orientations. The energies are calculated using DFT / B97-D / 6-311+G(d,p).
Table SI-1: Experimental vibrational frequencies and scaled DFT / B97-D / 6-311+G(d,p) vibrational frequencies for the β-turn conformation with various phenyl ring orientations. The frequencies are scaled according to $f_{scaled} = a \times f_{DFT} + b$. The set of parameters $(a,b)$ are $(0.92135, 188 \text{ cm}^{-1})$ for the NH stretch, $(0.63115, 1210 \text{ cm}^{-1})$ for the symmetric NH$_2$ stretch and $(0.60872, 1324 \text{ cm}^{-1})$ for the NH$_2$ antisymmetric stretch.

<table>
<thead>
<tr>
<th></th>
<th>NH stretch Phe</th>
<th>NH$_2$ symmetric</th>
<th>NH$_2$ antisymmetric</th>
</tr>
</thead>
<tbody>
<tr>
<td>g-</td>
<td>3470</td>
<td>3340</td>
<td>3502</td>
</tr>
<tr>
<td>g+</td>
<td>3424</td>
<td>3352</td>
<td>3504</td>
</tr>
<tr>
<td>a</td>
<td>3451</td>
<td>3340</td>
<td>3504</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>3460</td>
<td>3345</td>
</tr>
</tbody>
</table>

*Unable to record due to spectral absorption of LiNbO$_3$ crystal:
Table SI-2: Average values and deviations of dihedral angles (º) and of the C10/C7 H-Bond length (Å) for the γ and β conformers, extracted from the 50 K dynamics performed here. Values also given as reference for the geometry optimised structures with the level of calculation: all electron Gaussian B3LYP/6-311+G(p,d), and geometry optimisation with the set-up of the dynamics. The angles are defined as follows: \( \phi_{\text{Phe}} = (C_{e},C_{d},N_{c},C_{b}) \), \( \psi_{\text{Phe}} = (N_{f},C_{e},C_{d},N_{c}) \), \( \omega_{\text{Phe}} = (O,C_{b},N_{c},H_{\text{Phe}}) \), \( \chi_{1} = (N_{c},C_{d},C_{j},C_{k}) \), \( \chi_{2} = (C_{d},C_{j},C_{k},C_{l}) \), \( \phi_{\text{Pro}} = (C_{e},N_{f},C_{m},C_{h}) \), \( \psi_{\text{Pro}} = (N_{f},C_{m},C_{h},N_{i}) \), \( \omega_{\text{Pro}} = (O,C_{e},N_{f},C_{m}) \).

<table>
<thead>
<tr>
<th>γ-turn - a phenyl orientation</th>
<th>( \phi_{\text{Phe}} )</th>
<th>( \psi_{\text{Phe}} )</th>
<th>( \omega_{\text{Phe}} )</th>
<th>( \chi_{1} )</th>
<th>( \chi_{2} )</th>
<th>( \phi_{\text{Pro}} )</th>
<th>( \psi_{\text{Pro}} )</th>
<th>( \omega_{\text{Pro}} )</th>
<th>H-Bond (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opt/Gaussian-B3LYP</td>
<td>-151.39</td>
<td>139.26</td>
<td>178.56</td>
<td>-167.97</td>
<td>83.45</td>
<td>-82.87</td>
<td>70.12</td>
<td>174.58</td>
<td>1.99</td>
</tr>
<tr>
<td>MD/CP2K-BLYP-D3</td>
<td>-155.89</td>
<td>143.69</td>
<td>176.41</td>
<td>-172.20</td>
<td>87.93</td>
<td>-79.58</td>
<td>75.80</td>
<td>174.37</td>
<td>2.01</td>
</tr>
<tr>
<td>E = 0.00 kcal / mol</td>
<td>±9.07</td>
<td>±4.47</td>
<td>±6.06</td>
<td>±4.24</td>
<td>±4.44</td>
<td>±3.64</td>
<td>±5.05</td>
<td>±3.07</td>
<td>±0.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β-turn - a phenyl orientation</th>
<th>( \phi_{\text{Phe}} )</th>
<th>( \psi_{\text{Phe}} )</th>
<th>( \omega_{\text{Phe}} )</th>
<th>( \chi_{1} )</th>
<th>( \chi_{2} )</th>
<th>( \phi_{\text{Pro}} )</th>
<th>( \psi_{\text{Pro}} )</th>
<th>( \omega_{\text{Pro}} )</th>
<th>H-Bond (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opt/Gaussian-B3LYP</td>
<td>-63.61</td>
<td>145.09</td>
<td>-163.31</td>
<td>49.53</td>
<td>81.17</td>
<td>-90.89</td>
<td>14.71</td>
<td>-3.82</td>
<td>1.93</td>
</tr>
<tr>
<td>MD/CP2K-BLYP-D3</td>
<td>-63.47</td>
<td>147.46</td>
<td>-161.36</td>
<td>53.31</td>
<td>81.21</td>
<td>-87.01</td>
<td>11.98</td>
<td>-2.37</td>
<td>1.95</td>
</tr>
<tr>
<td>E = 1.22 kcal / mol</td>
<td>±8.79</td>
<td>±3.95</td>
<td>±5.48</td>
<td>±4.38</td>
<td>±5.30</td>
<td>±6.43</td>
<td>±9.56</td>
<td>7.79</td>
<td>±0.09</td>
</tr>
</tbody>
</table>
Chapter 6

Mapping gas phase dipeptide motions in the far-infrared domain.

The work presented in this chapter is the third work combining IR-UV ion dip spectroscopy with DFT-MD calculations focusing on peptidic systems in the far infrared after the two papers of our groups: "Gas-Phase Peptide Structures Unraveled by Far-IR Spectroscopy: Combining IR-UV Ion-Dip Experiments with Born-Oppenheimer Molecular Dynamics Simulations"\textsuperscript{105} where the emphasis was to demonstrate that very tiny structural differences could be unraveled using far infrared signatures and "Can far-IR action spectroscopy combined with BOMD simulations be conformation selective?"\textsuperscript{64}, where a detailed analysis of the conformational assignment and of the vibrational modes of the Ac-Phe-Pro-NH$_2$ dipeptide system were presented. This second paper is presented in chapter 5 of this manuscript.

Combining the results of these two papers, we observed on the one hand some features in common between Ac-Phe-Pro-NH$_2$, Ac-Phe-Gly-NH$_2$ and Ac-Phe-Ala-NH$_2$ spectra, for example in the 700-800 cm$^{-1}$ range. On the other hand, notable differences for the NH$_2$ motions signatures were observed, even though the function is involved in a similar $\gamma$-turn structure in these peptides (C7, the hydrogen bond forms a 7-membered ring, see figure 6.1). This is probably due to different hydrogen bond strengths that can be explained only by indirect influence of the chemical nature of the lateral chain in these peptides.

Based on these observations, we chose to study a large sample of dipeptides Ac-Phe-’AA’-NH$_2$, only changing one amino acid at the time, to observe the influence of the different lateral chains (also influencing the NH$_2$⋯O=C hydrogen bond strength) on the far infrared signatures. Here ’AA’ = Gly, Ala, Pro, Cys, Ser, Val, and Phe stands for phenylalanine. We did record the far infrared experimental spectra using IR-UV ion dip spectroscopy in the far IR/THz spectral domain (100-800 cm$^{-1}$, 3-24 THz) and we calculated the theoretical spectra using the BLYP-D3-MD method. Note that both experimental and theoretical methods are introduced in the paper and discussed in details respectively in chapters 3 and 4 of this thesis. Our combined experimental and theoretical characterization of Ac-Phe-’AA’-NH$_2$ peptides in the far infrared domain is reproduced in this chapter by our paper: "Mapping gas phase dipeptide motions in the far-infrared domain"\textsuperscript{65} that can be found hereafter.
One reason to choose these particular dipeptides is that they have already been structurally well characterised using the 'classical' 1000-2000 and 3000-4000 cm$^{-1}$ spectral ranges within the last decade by Mons and Rijs's groups$^{61-63}$ (see structures in figure 6.1). Only one $\gamma$-turn conformation can be found for Ac-Phe-Gly-NH$_2$, Ac-Phe-Ala-NH$_2$ and Ac-Phe-Ser-NH$_2$. Two distinct $\gamma$-turn conformations co-exist in our experimental conditions for Ac-Phe-Val-NH$_2$. For Ac-Phe-Pro-NH$_2$ and Ac-Phe-Cys-NH$_2$, $\gamma$ and $\beta$-turn conformations co-exist in our experimental conditions. Among these $\gamma$-turn systems, we observe a large diversity of NH$_2$····O=C hydrogen bond lengths with a maximum distance value of 2.21$\pm$0.11 Å for one conformer of Ac-Phe-Val-NH$_2$ and a minimum distance of 1.95$\pm$0.09 Å for Ac-Phe-Pro-NH$_2$.

Figure 6.1: Generic 3D structures of capped Ac-Phe-AA-NH$_2$ dipeptides. The 2 caps are 'Ac' for acetyl (CH$_3$-CO) and NH$_2$, respectively for the N- and C-terminals. 'AA' stands for Amino Acid, and 'AA'= Glycine (Gly), Alanine (Ala), Proline (Pro), Cysteine (Cys), Serine (Ser) or Valine (Val) in the present work.

a): $\gamma$-turn structure of Ac-Phe-Gly-NH$_2$, Ac-Phe-Ala-NH$_2$, Ac-Phe-Ser-NH$_2$, Ac-Phe-Val-NH$_2$, and Ac-Phe-Cys-NH$_2$ dipeptides. For Ac-Phe-Gly-NH$_2$, Ac-Phe-Ala-NH$_2$ and Ac-Phe-Val-NH$_2$, the amino acid residue does not interact with its immediate environment. For Ac-Phe-Ser-NH$_2$ and Ac-Phe-Cys-NH$_2$, the residue interacts with one Amide C=O group.

b): $\gamma$-turn of Ac-Phe-Pro-NH$_2$.

c): $\beta$-turn of Ac-Phe-Cys-NH$_2$ ($\beta$-turn type I).

d): $\beta$-turn of Ac-Phe-Pro-NH$_2$ ($\beta$-turn type VIa).

This figure is extracted from the attached paper: "Mapping gas phase dipeptides motions in the far-infrared and terahertz domain"$^{65}$. 

---

**Figure Legend:**

- **a)**: $\gamma$-turn structure of Ac-Phe-Gly-NH$_2$, Ac-Phe-Ala-NH$_2$, Ac-Phe-Ser-NH$_2$, Ac-Phe-Val-NH$_2$, and Ac-Phe-Cys-NH$_2$ dipeptides. For Ac-Phe-Gly-NH$_2$, Ac-Phe-Ala-NH$_2$ and Ac-Phe-Val-NH$_2$, the amino acid residue does not interact with its immediate environment. For Ac-Phe-Ser-NH$_2$ and Ac-Phe-Cys-NH$_2$, the residue interacts with one Amide C=O group.

- **b)**: $\gamma$-turn of Ac-Phe-Pro-NH$_2$.

- **c)**: $\beta$-turn of Ac-Phe-Cys-NH$_2$ ($\beta$-turn type I).

- **d)**: $\beta$-turn of Ac-Phe-Pro-NH$_2$ ($\beta$-turn type VIa).
Because the conformational assignments have been already obtained, we were able in this paper to present a comparison between the experimental spectra and the BLYP-D3-MD theoretical spectra without questioning the 3D structural assignment. Extremely well resolved spectra are found (as for our previous investigations in the far IR/THz domain, see chapter 5) and all the experimental peaks positions are reproduced by the theoretical spectra within an average 6 cm$^{-1}$ (calculated over all 9 dipeptides investigated and over all peaks in the 90-800 cm$^{-1}$ range). This is remarkable and it shows that our far IR/THz DFT-MD spectra provide the same information as the experiments.

Once the theoretical method for anharmonic vibrational spectroscopy has been validated, it is possible to perform unambiguous vibrational assignment. A separation between two kinds of modes present in the far infrared is provided in our work, i.e., 'local' modes for which one single internal coordinate dominates the motions in the normal mode and 'delocalised/collective' modes for which several internal coordinates dominate the motions in the mode. In this latter case, motions are therefore much more spatially delocalised over the molecule. This observation is supported in first place by normal modes calculated via the harmonic approximation from which this information is easily extracted. And it is confirmed by analysing a big amount of ICDOS spectra (see section 4.3 of chapter 4 for definitions). This second step is crucial since normal modes recombine as soon as mode couplings are taken into account at higher order in anharmonic vibrational methods (here in the DFT-MD method, or in a VSCF method as used by Gerber or Bowman's groups, for example).

For the Ac-Phe-'AA'-NH$_2$ series, 'local' modes are found to be out of plane motions (wagging) of hydrogen atoms and their rather localised character probably arises from the lightness of these atoms. ICDOS of dihedral angles are well suited to extract such wagging motions because one specific internal coordinate has one major contribution into the associated modes (see more in chapter 9).

This is not the case for delocalised modes for which several motions (i.e. several internal coordinates) combine. The ICDOS decompositions show these combinations, but in this work PCA (Principal Component Analysis) are also introduced and used. It appears that this tool is much better than ICDOS for analysing and quantifying these delocalised modes. See the attached paper for definition of PCA method.

A mapping of the vibrational modes is therefore presented in the attached paper. Modes dominated by the $\omega$(CH) motions are found between 700 and 800 cm$^{-1}$. They are not conformation selective because they have the same signatures whatever the environment of the aromatic ring of the phenylalanine (free, interacting with a NH function, ...). The $\omega$(NH) motions of the two NH amide groups dominate the modes between 400 and 650 cm$^{-1}$ and a correlation is found between the strength of the hydrogen bond into which the NH function is involved and the frequency of the mode. The stronger the interaction the more blue-shifted the wagging signature. When this NH function is free of interaction as it is found in the $\beta$-turn of Ac-Phe-Cys-NH$_2$, the $\omega$(NH) wagging mode is found inbetween 410 and 490 cm$^{-1}$. This certainly provides an absolute reference for $\omega$(NH) of backbone free NH groups. The case of the CH$_3$
function is also discussed since different behaviours can be observed as function of their local environment in the dipeptides investigated here. If the CH$_3$ group is linked to a sp$^3$ hybridised carbon, the signatures are found local (i.e. CH$_3$ hindered rotation is the main contribution into the mode) between 200 and 250 cm$^{-1}$. If the CH$_3$ group is linked to a sp$^2$ hybridised carbon, the hindered rotation is now involved into more delocalised modes (where backbone and lateral chain motions are involved) and signatures are found below 100 cm$^{-1}$.

In this dipeptide series, the $\gamma$ or $\beta$-turn motifs are formed through the hydrogen bond between the NH$_2$ terminal function and the CO group of the first or second amino acid of the peptide backbone. One question asked in this work was whether one specific signature of one of these two motifs ($\gamma$ or $\beta$-turn) of the peptide could be found. We did not manage to get specific signatures for $\beta$ vs $\gamma$-turn structures unfortunately. The far infrared signatures of the two $\beta$-turn conformations are completely different (this might be explained by different types of $\beta$-turn, we have type I for Ac-Phe-Cys-NH$_2$ and type VIa for Ac-Phe-Pro-NH$_2$; in other words different dihedral angle values of the backbone of the amino acids involved into the motif). For all the $\gamma$-turns investigated here we find one common $\omega$(NH) signature between 470-510 cm$^{-1}$ but we found this signature related to a common C5 interaction and not to a C7 interaction. The signatures of the NH$_2$ group (involved into the C7 interaction) are different and we actually observe a good correlation between the length of the hydrogen bond and the position of the signatures of the asymmetric motions of $\omega$(NH$_2$).

For the dipeptide series, local modes, i.e. $\omega$(CH), $\omega$(NH) and $\omega$(NH$_2$), dominate the range 400-800 cm$^{-1}$ while they are absent below 400 cm$^{-1}$. This does not mean that delocalised modes are not present in the same range. These can be found in the whole 0-800 cm$^{-1}$ and mainly composed of bending motions along the backbone in the range 100-800 cm$^{-1}$ while the contributions of the dihedral angles of the backbone arise below 100 cm$^{-1}$. The local modes are mainly composed of wagging motions that induce strong variations of the dipole moment of the peptide, which explains the rather important intensities of these modes in the spectra. Delocalised modes are active in the whole range 0-800 cm$^{-1}$ even though their intensities are weak. This can be explained by the combination between some internal coordinates inducing strong variations of the dipole moment and some others inducing only a small or no variation.

An extended mapping is presented in chapter 9, including all the systems investigated in the context of this thesis.
Mapping gas phase dipeptide motions in the far-infrared and terahertz domain†

Jérôme Mahè,ab Daniël J. Bakker,c Sander Jaeqx,c Anouk M. Rjisc and Marie-Pierre Gaigeot c,b

Vibrational signatures of Ac-Phe-AA-NH₂ dipeptides are recorded and analysed in the far IR/THz spectral domain (100–800 cm⁻¹, 3–24 THz), with the ‘AA’ amino acid chosen within the series ‘AA’ = Gly, Ala, Pro, Cys, Ser, Val. Phe stands for phenylalanine. IR-UV ion dip experiments are conducted on the free electron laser FELIX and combined with DFT-based molecular dynamics simulations for the calculation of the dynamical anharmonic vibrational spectra. The excellent agreements between the experimental and theoretical spectra of the Ac-Phe-AA-NH₂ series allow us to make detailed and unambiguous mapping of the vibrational motions into three main domains: 700–800 cm⁻¹ for C–H waggings, 400–700 cm⁻¹ for N–H waggings, with a one-to-one signature per amide N–H backbone group, 0–400 cm⁻¹ for delocalized and large amplitude collective motions over the dipeptide backbone, with backbone torsional motions arising <100 cm⁻¹.

1 Introduction

Infrared (IR) vibrational spectroscopy is one of the main analytical tools widely used to unravel structures and isomers of molecules and clusters in the gas phase and of liquids and (bio)molecules immersed in liquids, solids, solid/liquid and liquid/air interfaces. Although the whole IR spectrum (0–4000 cm⁻¹) can in principle be probed, the 1000–2000 cm⁻¹ (30–60 THz) and the 3000–4000 cm⁻¹ (90–120 THz) spectral domains have certainly been the most employed in the literature. This is especially true for gas phase IR (action) spectroscopy, where the 3000–4000 cm⁻¹ O–H/N–H stretching motions have been largely studied for their diagnosis indication of the formation of hydrogen bonds in structures of e.g. peptides, nucleic acids, sugars, and clusters.1–11 The 1000–2000 cm⁻¹ spectral region, including amide I and II vibrations recognized as molecular fingerprints, has also been usefully employed, see for example ref. 12–16. In the case of spectral congestion (resulting from large complex molecular systems, temperature, intra-/inter-molecular interactions broadening the peaks, or simply many vibrations within small frequency intervals) one is left with, at best, identification of the structural families only, losing the conformational details hidden underneath the IR features.17,18 More examples exist in the literature illustrating the difficulty in assigning a definitive structure to one given molecular system when there are either subtle differences in geometry19,20 or for larger peptides.20,21

To go beyond spectral congestion and/or to circumvent the limitations of local-structure information inherent to the 1000–4000 cm⁻¹ vibrations, without adding up one more dimension to vibrational spectroscopy (as in e.g. ion mobility22,23 or isotopic substitution24), one can probe vibrational signatures in the far-IR/THz domain (<800 cm⁻¹, <24 THz).‡ In this domain, direct information on collective and delocalised motions, on long-range H-bond network(s) organization can be obtained, which are only indirectly probed in other spectral domains.

To date far IR/THz vibrational spectroscopy has been mostly developed and employed to characterize crystals, semiconductors, liquids and biomolecules, see ref. 25–28 and references therein, especially ref. 27 compiling earlier works in this field. A certain number of set-ups have been developed for probing gas phase molecular systems in the far-IR/THz domain, typically based on classical IR sources,29 or alternatively based on Raman,30–32 LIF (Laser Induced Fluorescence)33 or microwave34–36 spectroscopies. These methods are however limited in the size of the molecular systems that can be probed and some of them are not well adapted to get conformational assignment(s). We have recently demonstrated that far-IR vibrational spectroscopy of gas phase molecules,

‡ Although there are no ‘exact’ boundaries in the literature, mid-infrared radiation covers the region between 400–800 cm⁻¹ and far-infrared covers 800–1000 cm⁻¹. The THz region is nowadays broadly defined below 300 cm⁻¹ (<9 THz), extending the more conventional <100 cm⁻¹ (<3 THz) domain used in certain communities.
in particular peptides, can be achieved using the Free Electron Laser (FEL) at the FELIX laboratory. With such a radiation source of high fluence, one can access to the vibrational spectroscopy of much larger systems. The same strategy has been applied for metal- or water-clusters in other groups. In our previous works, this far-IR experiment has been successfully coupled to theoretical anharmonic spectroscopy (through Born–Oppenheimer DFT-based molecular dynamics simulations, denoted DFT-MD hereafter). We have for example demonstrated that the far-IR signatures were mandatory for distinguishing subtle differences between two structural conformers of the Ac-Phe-Gly-NH2 peptide (i.e. C7eq/C7ax H-bond). The definitive assignment of the C7eq conformation of Ac-Phe-Gly-NH2 was not possible in the mid-IR or in the 3000–4000 cm⁻¹ domains.

It is one thing to be able to measure experimental spectra of gas phase molecules in the far-IR/THz domain, it is another issue to be able to interpret the spectral signatures in terms of vibrational motions and relate the spectra to specific conformations. This is where theoretical calculations of vibrational spectra are essential. We combine here far-IR experimental IR-UV ion dip spectroscopy of gas phase Ac-Phe-AA-NH2 dipeptides with DFT-MD simulations (DFT-based Molecular Dynamics) for anharmonic dynamical theoretical spectroscopy. ‘AA’ stands for one of the following amino acids, Gly, Ala, Pro, Cys, Ser, Val. The dipeptide series always contains the Phenylalanine chromophore used for UV excitation and ionisation in the experiment.

The dipeptide series investigated here has been chosen especially because the 3D structures have been well characterized in the literature using the ‘classical’ 1000–2000 and 3000–4000 cm⁻¹ spectral signatures. The present investigations confirm these structural assignments by adding-up spectral analyses in the far-IR domain. What the present analyses further provide is a general map of the motions that are responsible for the spectroscopic signatures recorded in the far-IR domain. In the particular case of the Ac-Phe-Cys-NH2 peptide, the literature could not give a definitive conformational assignment, which is given here with far-IR signatures. Experiments explore the 100–800 cm⁻¹ region (3–24 THz), the simulations also provide the signatures below 100 cm⁻¹. The vibrational assignment map is obtained from the DFT-MD dynamic non harmonic spectra, thus taking into account temperature effects, anharmonic motions (including mode couplings), and anharmonic dipole moments. It has been shown in the last decade that DFT-MD provides good to excellent vibrational spectra in the prototypical 1000–4000 cm⁻¹ spectral range when compared to experiments.

This paper presents the molecular systems and known data for their 3D structures in Section 2, the experimental and theoretical methods in Section 3. The first assessment of the quality of the DFT-MD dynamical anharmonic spectra with respect to the experimental spectra is presented in Section 4 for the Ac-Phe-AA-NH2 dipeptide series investigated here. Dissection of the vibrational motions at play in the 100–800 cm⁻¹ (3–24 THz) far-IR/THz domain is discussed in Section 5, thus providing a vibrational map to be used by the community. Conclusions and perspectives are given in Section 6.

2 Investigated dipeptide systems

One reason to choose these particular dipeptides is that they have already been structurally well characterised using the ‘classical’ 1000–2000 and 3000–4000 cm⁻¹ spectral ranges within the last decade. These dipeptides are folded in either a γ-(C7) or a β-(C10) turn geometry, however the exact conformation adopted by Ac-Phe-Cys-NH2 is still not solved.

From the literature, only one conformer was found for Ac-Phe-Gly-NH2, Ac-Phe-Ala-NH2 and Ac-Phe-Ser-NH2 dipeptides, corresponding to a γ-turn geometry where one internal hydrogen bond leads to the formation of a seven membered ring (denoted C7 interaction and leading to a γ-turn structure). Such a conformation is illustrated in Fig. 1(a). Two conformers of Ac-Phe-Val-NH2 were identified, both with a γ-turn structure and differing by the orientation of the CH(CH3)2 residue with respect to the backbone. For Ac-Phe-Cys-NH2, two main conformations were reported, one γ-turn and one β-turn (an internal hydrogen bond leading to the formation of a ten membered ring, denoted C10 interaction and β-turn geometry), see Fig. 1(a) and (c). For the β-turn, the precise orientation of the S–H group is still an open issue (interacting either with one backbone C=O group or with the aromatic ring), which we believe is solved in our present work (see the ESI†: S–H interacts with C=O). The Ac-Phe-Pro-NH2 dipeptide was also shown to display two structural isomers, one γ- and one β-turn. The proline residue forms a ring and γ/β-turns are illustrated separately in Fig. 1(b) and (d) for the sake of clarity.

Fig. 1 Generic structures of Ac-Phe-AA-NH2 dipeptides, ‘AA’ stands for amino acid, and ‘AA’ = Glycine (Gly), alanine (Ala), proline (Pro), cysteine (Cys), serine (Ser) or valine (Val) in the present work. (a) γ-Turn structure of Ac-Phe-Gly-NH2, Ac-Phe-Ala-NH2, Ac-Phe-Pro-NH2, Ac-Phe-Val-NH2, Ac-Phe-Ser-NH2, Ac-Phe-Cys-NH2 and Ac-Phe-Cys-NH2 dipeptides. For Ac-Phe-Gly-NH2, Ac-Phe-Ala-NH2 and Ac-Phe-Val-NH2, the amino acid residue does not interact with its immediate environment. For Ac-Phe-Ser-NH2 and Ac-Phe-Cys-NH2, the residue interacts with one amide C=O group. (b) γ-Turn of Ac-Phe-Pro-NH2. (c) γ-Turn structure of Ac-Phe-Cys-NH2 (β-turn type I). (d) β-Turn structure of Ac-Phe-Pro-NH2 (β-turn type II). See the text for more details on the structures.
3 Experimental and theoretical methods

All dipeptides have been purchased from GeneCust (Dudelange, Luxembourg) with a purity exceeding 95% and used without further purification. The experimental set-up and spectroscopic methods have been described elsewhere. Briefly, the molecules were desorbed from a sample/carbon black mixture applied on a sample bar using a pulsed YAG laser operating at 1064 nm (1 mJ pulse). The cold (≥ 10–20 K) gas phase neutral molecules are prepared in a supersonic jet expansion by a pulsed valve and a backing pressure of 3 bar Argon.

For the IR-UV dip experiments, the molecules are ionised via a one color Resonance Enhanced Multi-Photon Ionization ((1 + 1) REMPI) scheme. The S0 → S1 transitions are selected for each conformation of each single molecule. The UV frequency was set at 37486 cm−1 for Ac-Phe-Gly-NH2, 37464 cm−1 for Ac-Phe-Ala-NH2, 37435.5 cm−1 and 37409 cm−1 for the γ and β-turn conformations of Ac-Phe-Pro-NH2, respectively, at 37390 cm−1 for Ac-Phe-Ser-NH2, at 37325 cm−1 and 37450 cm−1 for the γ and β-turn conformations of Ac-Phe-Cys-NH2, respectively, for the A1 conformation of Ac-Phe-Val-NH2 and 37472 cm−1 for the A2 conformation of Ac-Phe-Val-NH2. The values of these transitions can be found in the literature and have been confirmed using the REMPI spectra (spectrum of Ac-Phe-Val-NH2 is presented as an example in SM, Fig. 2). The UV pulses are produced using a frequency doubled dye laser pumped by a frequency doubled Nd:YAG laser operated at 10 Hz with a typical energy pulse of 1–2 mJ.

The generated ions are detected using a reflectron time of flight mass spectrometer. The UV pulses are preceded by far infrared light produced by the Free Electron Lasers FELIX located at the FELIX laboratory at the Radboud University. Whenever the radiation is resonant with a vibrational transition of the system, the ground state is depleted and the UV photons are not resonant anymore with an electronic transition of the system. We therefore observe a depletion in the mass spectrometer signal. During the experiment, IR-on and IR-off signals are alternated to correct the long term UV power drifts and the evolution of the source conditions by operating the IR laser at 5 Hz and the UV laser at 10 Hz. Since the experiments are performed over a very wide frequency range, the IR intensities of the measured transitions are corrected for the FELIX macropulse wavelength dependent energy and for the number of photon(s) present in the FELIX macropulse at different wavelength(s) for identical pulse energy. The experimental spectra presented here are obtained as an average over a minimum of three scans with a minimum of 30 averages per point. Macropulses of the used FELs typically have a time duration of 6 to 10 μs. The wavelength-dependent intensity of the infrared light of FELIX varies from 20 mJ per macropulse at the extremes of the measured range to 80 mJ per macropulse at 650 cm−1.

Theoretical vibrational spectroscopy is performed through DFT-MD simulations (DFT based Molecular Dynamics) using the CP2K package, within the Born–Oppenheimer framework. The BLYP-D3 GGA functional has been selected based on our previous works on gas-phase peptides, sugars and ionic clusters in different spectral ranges, including the far-IR domain. The simulations are based on the GPW (Gaussian Plane Waves) method, with a dual basis set representation constructed here on a plane wave basis set with a 450 Ry energy cutoff and the aug-TZVP Gaussian basis set. Pseudopotentials of the GTH type (Goedecker-Teter-Hutter) are used. Simulations are done in a cubic box of 20 Å3. The time-step for the dynamics is 0.4 fs.

DFT-MD is performed at 50 K, as previously determined. Initial structures for the dynamics are taken from the literature as summarized in the introduction. Geometries have been re-optimised at our level of theory, and were used as starting conformation(s) for the dynamics. Considering the low-temperature of the dynamics, no conformational change is observed along the trajectories. In the experiment, a single conformer is selected by fixing the UV laser on a unique UV transition frequency, while the scanning of the IR laser is used in order to record a mass and conformer selective IR spectrum. The simulations show that the isomeric structure chosen at the initial state of the dynamics (γ or β) does not isomerize over time. Each DFT-MD infrared spectrum is obtained as an average over three separate trajectories (the same initial geometry, different initial velocities) of 20 ps each (after thermalisation over 1 ps) in order to obtain converged relative intensities of the absorption peaks. Spectra are therefore calculated over a total 60 ps trajectory.

The comparison of absolute intensities between theory and IR-UV ion dip experiment is not trivial. The experiment measures the absorption of an IR photon through IR-UV action spectroscopy while the theoretical signal results from direct one-IR photon absorption. One however expects ratios of intensities to be comparable between theory and the experiment conducted here, which is the basis of our spectral analyses. Note that for the sake of clarity, all theoretical intensities are multiplied by 100 so that the most intense theoretical bands match the most intense experimental ones (this multiplicative factor is constant for all dipeptide systems). In a one-to-one comparison of band ratios, discrepancies can still arise because of inherent experimental error-bars, because of too short trajectories and equipartition not fully reached in the gas phase trajectories despite the care taken.

All further details for the calculation and analyses of the dynamical anharmonic IR spectra from the DFT-MD simulations are given in the ESI. Theory is based on the time correlation formalism of the fluctuating dipole moment vectors, reviewed in ref. 48 and 49.

4 How well does DFT-MD perform for far-IR spectroscopy?

All far-infrared spectra measured experimentally and extracted from the present DFT-MD theoretical calculations are reported in Fig. 3. In Fig. 3–5, the dipeptide spectra are organised from top to bottom, with the order of the γ-turn A1 conformation of Ac-Phe-Val-NH2, the β-turn of Ac-Phe-Cys-NH2, γ-turns of Ac-Phe-Gly-NH2, Ac-Phe-Ala-NH2, (A2) Ac-Phe-Val-NH2, Ac-Phe-Ser-NH2, Ac-Phe-Cys-NH2, Ac-Phe-Pro-NH2 and the β-turn of...
Ac-Phe-Pro-NH₂. We report the 0–800 cm⁻¹ far-IR spectra of these peptides in Fig. 3, split into two spectral regions for clarity reasons (left for 0–400 cm⁻¹, and right for 400–800 cm⁻¹). Experimental spectra are plotted in color while the theoretical anharmonic dynamical DFT-MD spectra are plotted in black, just below each experiment.

The first remarkable result is the average 6 cm⁻¹ difference between experimental and theoretical peaks positions, calculated over all 9 dipeptides and over all peaks in the 90–800 cm⁻¹ range. A maximum deviation of about 20 cm⁻¹ is systematically obtained (in all systems) for the peak located at 740 cm⁻¹ experimentally and for the peak located at around 720 cm⁻¹ theoretically. This peak is part of a triplet that is not conformational selective, as it is present for all systems investigated (highlighted in blue in Fig. 4). This excellent match between theory and experiment validates the previous 3D structural assignments from the literature using other spectral domains.

Theoretical bands are furthermore found as well resolved as the experimental ones, i.e. showing the same number of absorption peaks, the same narrow band-widths and the same band-shapes in general also including subtle shoulders in some broader peaks. There are some exceptions in the 400–500 cm⁻¹ range with less well resolved peaks in the calculations. Although experimental band-intensities are globally rather well reproduced in our dynamic spectra (without any adjustable parameter and/or model), some peaks however display either too low or too high intensities in comparison to the experiments. See for instance the peaks located at around 700 cm⁻¹ for the A1 conformation of the Ac-Phe-Val-NH₂ dipeptide or the γ-turn of Ac-Phe-Cys-NH₂. Band-intensities from MD trajectories are sensitive to temperature, the length of the trajectories and the number of trajectories, and presumably some of these elements participate here to such deviations from experiments. They, however, do not prevent a fair assignment and interpretation of the experimental spectra. Note also that relative intensities between peaks are usually well reproduced by the dynamic spectra.

5 Mapping vibrational motions in the far-IR

We have applied the two theoretical tools described in Section S3 of the ESI, i.e. ICDOS and PCA, in order to unravel vibrational motions of gas phase dipeptides in the 0–800 cm⁻¹ far-IR/THz domain. Although the far infrared range is generally expected to be mostly dominated by delocalized and collective vibrational motions, we will see below that local N–H and C–H wagging motions dominate the 400–800 cm⁻¹ range for the Ac-Phe-AA-NH₂ dipeptides studied here, with strong absorptions. The ICDOS local method is therefore well adapted to unravel such movements. Delocalized and collective motions are present in the whole far infrared range, displaying less intense peaks, but they especially dominate (in intensity) the 0–400 cm⁻¹ domain. PCA analysis is well adapted to characterize these motions. The values of the band positions discussed in the rest of the text are the ones obtained from the experimental spectra, keeping in mind that the average difference between the experimental and theoretical spectra is 6 cm⁻¹ (see Section 4).

5.1 Local wagging motions

C–H waggings of the phenylalanine ring (700–800 cm⁻¹). Three absorption bands are systematically found in the 700–800 cm⁻¹ spectral domain of the gas phase Ac-Phe-AA-NH₂ dipeptides, located at 700, 730 and 740 cm⁻¹ and highlighted in blue in Fig. 3 and 4. These bands arise from the ω(C–H) out-of-plane motions (wagging) of the hydrogen atoms that belong
to the phenylalanine ring, systematically present in all the
dipeptides studied here. The assignment is supported by the
ICDOS analysis presented in Fig. 4, where spectra of the dihedral
angles $\Phi = C-C-C-H_{1,2,3}$ are plotted (see Fig. 2). The phenylalanine
aromatic ring can be found in three different environments within
the dipeptides of interest here, i.e. (i) interacting with the amide
N–H group of the ‘AA’ amino acid in the $\gamma$-turn geometries of
Ac-Phe-Gly-NH$_2$, Ac-Phe-Ala-NH$_2$, Ac-Phe-Cys-NH$_2$ ($\gamma$-turn),
Ac-Phe-Ser-NH$_2$, Ac-Phe-Val-NH$_2$, (Fig. 1(a)); (ii) in the $\beta$-turn
conformation of Ac-Phe-Cys-NH$_2$, this interaction is weaker; (iii)
it interacts with the proline ring via a $\pi$-interaction in the $\gamma$- and
$\beta$-turn conformations of the Ac-Phe-Pro-NH$_2$ peptide. Despite
these different environments, the same spectral signatures for
the aromatic C–H wagging motions are found, therefore being
non sensitive to the environmental details. Note that equivalent
spectral signatures are also found in systems containing a Phenol
ring with chemical substitution, therefore being signatures of
C–H wagging motions in aromatic rings.

N–H waggings (400–700 cm$^{-1}$). The range 400–700 cm$^{-1}$ is
dominated by the backbone amide N–H out-of-plane wagging
motions (amide V mode) in the Ac-Phe-AA-NH$_2$ dipeptide
series. Because these dipeptides have one NH$_2$ terminal,
wagging signatures from this moiety are also present in this
spectral domain. The ICDOS analyses used to unravel these
movements are reported in Fig. 5, where the ICDOS spectra of the
two H$_{\text{NH2}}$–N$_{\text{NH2}}$–C$_a$–C and H$_{\text{NAA}}$–N$_{\text{NAA}}$–C$_a$–C dihedral angles are
plotted on the left-panel, and the symmetric/antisymmetric
N–H ICDOS spectra of the NH$_2$ moiety are plotted in the
right-panel.

Two separate amide V wagging signatures are obtained
for the two backbone N–H amide groups. They are labelled
$\omega$(N–H)$_{\text{Phe}}$ and $\omega$(N–H)$_{\text{AA}}$ in the following, related to the
phenylalanine amide and the ‘AA’ residue amide respectively.

$\omega$(N–H)$_{\text{Phe}}$ is located between 470 and 510 cm$^{-1}$, as high-
lighted by the green box in Fig. 5 (left panel). This specific
amide group has a common signature within the 40 cm$^{-1}$
interval for all Ac-Phe-AA-NH$_2$ dipeptides, which is not surprising
as this N–H$_{\text{Phe}}$ is always involved in a weak C5 interaction with
C$_Q$OPhe (see illustrations of $\gamma$- and $\beta$-turn structures in Fig. 1).
The 470–510 cm$^{-1}$ region can therefore be used as a reference for
$\omega$(N–H)$_{\text{Phe}}$ wagging signatures of N–H amide groups engaged in a
C5 backbone interaction. When this N–H$_{\text{Phe}}$ is free of interaction as
is found in the $\beta$-turn of Ac-Phe-Cys-NH$_2$ (only exception in the
series, see Fig. 1), $\omega$(N–H)$_{\text{Phe}}$ is found at around 410–490 cm$^{-1}$.
This certainly provides an absolute reference for $\omega$(N–H)$_{\text{Phe}}$
of the backbone free N–H groups.

$\omega$(N–H)$_{\text{AA}}$ displays various signatures between 400 and
640 cm$^{-1}$ depending on the chemical composition of the residue
and interactions of this backbone N–H with its surrounding.

For Ac-Phe-Gly-NH$_2$, Ac-Phe-Ala-NH$_2$, and the two Ac-Phe-
Val-NH$_2$ $\gamma$-turn structures, there is an interaction between
N–H$_{\text{AA}}$ and the Phe aromatic ring, see Fig. 1(a). The associated
wagging signatures are therefore located between 530 and
550 cm$^{-1}$ for these dipeptides, as illustrated by the orange
box in Fig. 5 (left panel). This represents up to a maximum of \( \approx 90 \text{ cm}^{-1} \) blue-shift from the free \( \nu(N-H)_{\text{free}} \) wagging signature characterised above. For the Ac-Phe-Ser-NH\(_2\) and Ac-Phe-Cys-NH\(_2\) \( \gamma \)-turns, the same weak phenylalanine/N-H\(_{AA}\) interaction is present. However, due to the different chemical nature of the side chain containing either OH or SH groups, respectively, a supplementary blue shift of \( \nu(N-H) \) is observable, now located at 555 and 620 cm\(^{-1}\) for Ac-Phe-Ser-NH\(_2\) and Ac-Phe-Cys-NH\(_2\), respectively. See the corresponding yellow and violet boxes in Fig. 5 (left panel). In the \( \beta \)-turn conformation of Ac-Phe-Cys-NH\(_2\), N-H\(_{AA}\) is only weakly interacting with the aromatic ring (Fig. 1), and the corresponding wagging \( \nu(N-H)_{\text{AA}} \) is found at around 565 cm\(^{-1}\). In Ac-Phe-Pro-NH\(_2\), there is no N-H\(_{AA}\), since the hydrogen atom is replaced by the proline ring.

The amide V mode is found to be diagnostic for the local environment around the N-H backbone group, and can be used as a valuable tool for the conformational assignment of gas phase peptides, providing information equivalent to the 3000–4000 cm\(^{-1}\) NH stretching range. Furthermore, we have just shown that each backbone amide N-H group gives rise to a distinct wagging signature in the 400–700 cm\(^{-1}\) domain, hence reflecting backbone interactions with the surrounding. This is reminiscent of the separate \( \nu(N-H) \) stretching signatures observed in other works\(^{2,9,46,61–64}\) in the 3000–4000 cm\(^{-1}\) domain under cold and isolated conditions in either a cold ion trap or a molecular beam environment. Although here the separate \( \nu(N-H) \) peaks are slightly broader than the ones observed for \( \nu(N-H) \), one can still distinguish one peak per N-H group without any ambiguity. One would therefore be able to experimentally separate \( \nu(N-H) \) signatures by applying isotopic substitution, as done in the \( \nu(N-H) \) stretching region\(^6\) or more recently for the equivalent \( \nu(O-H) \) wagging.\(^4\)

**NH\(_2\) wagging (400–700 cm\(^{-1}\))**. The NH\(_2\) terminal moiety in the Ac-Phe-AA-NH\(_2\) dipeptide series is known to induce folding\(^6\) as observed here for the \( \beta \)-turn geometries where this moiety is involved in one H-bond with one backbone carbonyl group. In \( \gamma \)-turn structures, NH\(_2\) can H-bond to the C=O\(_{AA}\) neighbour group. The NH\(_2\) moiety gives rise to symmetric and antisymmetric N-H out-of-plane wagging motions, extracted from the trajectories as the sum (sym) and the difference (asym) between the signatures of the H\(_{Bond}–N–C–C_{\beta}\) and H\(_{Free}–N–C–C_{\beta}\) dihedral angles (H\(_{Bond} \) and H\(_{Free} \) are, respectively, the hydrogen bonded and free N–H of the NH\(_2\) group). The right panel of Fig. 5 presents the ICDOS spectra for these NH\(_2\) motions.

As highlighted in blue in Fig. 5, a strong correlation is observed between the NH\(_2\)–O=C H-bond length and the frequency of the asymmetric out-of-plane wagging motion. The stronger the hydrogen bond the more blue-shifted the wagging signature. Note that such a correlation also exists for the N–H backbone amide V mode dissected above, and similar blue-shifts have also been reported for OH ••• OH H-bonds.\(^10\) We thus find asymmetric \( \nu(NH_2) \) wagging between \( \approx 620 \text{ cm}^{-1} \) for the weakest hydrogen bond formed in the \( \gamma \)-turn conformer A1 of Ac-Phe-Val-NH\(_2\) and \( \approx 695 \text{ cm}^{-1} \) for the stronger H-bonds formed in the \( \gamma \)-turn and \( \beta \)-turn structures of Ac-Phe-Pro-NH\(_2\). Such a correlation between \( \nu(NH_2) \) and NH\(_2\)–O=C H-bond strength is not observed for the symmetric wagging motion, as highlighted in red in Fig. 5. Signatures of this motion are systematically found in the 400–510 cm\(^{-1}\) interval. \( \nu_{\text{sym}}(NH_2) \) is however found at 595 cm\(^{-1}\) for the \( \beta \)-turn conformation of Ac-Phe-Pro-NH\(_2\).

**OH and SH wagging**. Other wagging motions can be found in the dipeptides series, *i.e.* \( \nu(O-H) \) and \( \nu(S-H) \), respectively, present in Ac-Phe-Ser-NH\(_2\) and Ac-Phe-Cys-NH\(_2\). The \( \nu(O-H) \) vibration is found at 557 cm\(^{-1}\), while \( \nu(S-H) \) is found at 340 and 369 cm\(^{-1}\) in the \( \gamma \)-turn Ac-Phe-Cys-NH\(_2\) (average H-bond of 2.45 ± 0.11 Å) and at 288 and 310 cm\(^{-1}\) in the \( \beta \)-turn conformation (average H-bond of 2.90 ± 0.21 Å). These latter values show again a blue-shift of the wagging motion with an increase in the strength of the hydrogen bond.

**CH\(_3\) rotational motions**. We find two spectral features for CH\(_3\) hindered rotational motions related to the two methyl populations in the dipeptide series investigated here. These motions are extracted by ICDOS of angle coordinates, *i.e.* H\(_{CH_3}–C–N_{\text{Ph}}\), H\(_{CH_3}–C–C_{\beta}–N_{AA}\), H\(_{CH_3}–C–C_{\beta}–C_{\beta}\) angles, respectively, for the backbone CH\(_3\) and for CH\(_3\) in Ala and Val residues. For one given methyl group, we look at the simultaneous signatures
of the three H atoms, which indeed give us the (hindered) rotational motion of CH$_3$ as a ‘solid body’. Signatures of the rotation of the terminal CH$_3$ group of the dipeptides are found below 100 cm$^{-1}$, and are systematically coupled with delocalised modes that will be described in the next section. For residues alanine and valine in Ac-Phe-Ala-NH$_2$ and Ac-Phe-Val-NH$_2$ dipeptides, rotation of the methyl group is found at 220 and 242 cm$^{-1}$ (Ala) and at 221, 239, 257 cm$^{-1}$ (Val). Distinct rotational signatures for side chain methyl groups and backbone methyl groups have already been observed for peptides in the condensed phase within the same vibrational range. For our gas phase peptides, we clearly observe that the terminal methyl group has more amplitude in its rotational motion because of its sp$^2$ carbon hybridization. When the methyl is part of the peptide residues, its carbon being now sp$^3$ hybridized, a different environment is provided to the hydrogen atoms when the methyl group is rotating, the methyl is thus more constrained in its rotation, leading to a rotational motion at a higher frequency.

5.2 Collective and delocalised backbone modes (0–400 cm$^{-1}$)

We find that collective vibrational motions delocalized over the peptide backbone are present in the entire far infrared range (<800 cm$^{-1}$), but while these modes are hidden by intense local wagging modes in the 400–800 cm$^{-1}$ range, their intensity dominates the range 0–400 cm$^{-1}$. These motions are evidenced by the ICDOS signatures of internal coordinates based on angles and dihedral angles. We find that while the backbone bending motions are predominant in the 100–400 cm$^{-1}$ range, the backbone dihedral motions dominate below 100 cm$^{-1}$. At the time of writing, the FELIX Free Electron Laser laboratory facility can be tuned down to 80–90 cm$^{-1}$ and is limited for lower frequencies. Obviously, our theoretical calculations do not have such limitations, and the theoretical dynamical spectra presented here provide the expected spectral signatures below 100 cm$^{-1}$, to be validated by experiments when they will become available.

In Fig. 6, we illustrate the limitations of the local ICDOS analyses and the strength of the PCA analyses in revealing delocalized modes in the far-IR domain. We take here the Ac-Phe-Ser-NH$_2$ dipeptide as an example. The left panel of Fig. 6 presents the individual ICDOS signatures of the dihedral angles of this dipeptide, while right panel reports the PCA (Principle Component Analyses) based on the combination of all backbone dihedral internal coordinates (see Fig. 2 for the
definition of the 8 dihedral internal coordinates used in this PCA analysis). The top blue line is the theoretical IR spectrum. One can immediately observe that each dihedral angle has intense signatures from 0 to 100 cm$^{-1}$, spread over multiple peaks. Several dihedral angles provide comparable signatures at the same given frequency (the left panel of Fig. 6). This indeed reveals the delocalized character of the normal modes in this frequency range. If we now turn to the PCA analysis (the right panel of Fig. 6), for Ac-Phe-Ser-NH$_2$ the signatures arising from the first principal component provide almost all bands observed in the IR spectrum between 0 and 100 cm$^{-1}$, the third principal component provides the two missing supplementary bands close to 100 cm$^{-1}$. Looking more into detail at the motions that constitute the 1st principal component, we find that torsions around the two dihedral angles $\Phi_{\text{AA}}$ and $\Psi_{\text{Phe}}$ (see Fig. 2) are the two systematic main contributors, for all dipeptides investigated here. The percentage of participation of these two motions varies within the dipeptide series, but on average these two motions amount to $\sim 50\%$ of all dihedral motions.

6 Conclusion and outlook

We have mapped the vibrational motions arising from gas phase dipeptides in the far IR/THz spectral domain (100–800 cm$^{-1}$, 3−24 THz). The dipeptides investigated here are of the sequence Ac-Phe-AA-NH$_2$ with the amino acid ‘AA’ being chosen from the series Gly, Ala, Pro, Cys, Ser and Val. Their 3-dimensional organisations are $\gamma$-C7 or $\beta$-C10 turn geometries, known from previous works in the literature$^{45-47}$ and from other spectral domains, whose assignments are confirmed by the present work. The experimental spectra are remarkably well resolved, and we have shown that DFT-based molecular dynamics simulations (DFT-MD) are an excellent tool to apply for the calculation of the far IR spectra of these dipeptides. An average 6 cm$^{-1}$ deviation is obtained between theoretical anharmonic dynamical spectra and experiments, the theoretical bands are found as well resolved as the experimental ones, showing the same number of absorption peaks, the same narrow band-widths and the same band-shapes in general also including subtle shoulders in some broader peaks. By combining gas phase far IR experiments with DFT-MD anharmonic dynamical theoretical spectroscopy, it is therefore possible to make definitive structural assignments and provide reliable mapping of the expected vibrational motions.

We have shown the existence of three separate vibrational domains between 0 and 800 cm$^{-1}$ (0–24 THz). 700–800 cm$^{-1}$ is due to C–H waggings, here belonging to the phenylalanine ring of the Ac-Phe-AA-NH$_2$ dipeptides. The same domain of C–H waggings has been observed for phenol derivative molecules in our previous work.$^{41}$ This domain is not conformational selective. The 400–700 cm$^{-1}$ domain is dominated by the N–H wagging motions (amide V), there are as many signatures as the number of amide N–H backbone groups. The absolute reference for a free N–H wagging has been found at 410–490 cm$^{-1}$, while any H-bonding of the N–H blue-shifts its frequency motion, up to 640 cm$^{-1}$ for the present investigated dipeptides. The stronger the H-bond formed, the larger the blue-shift of $\omega$(N–H). The N–H wagging motion is found to be diagnostic for the local environment and therefore conformer selective. Because of the presence of the NH$_2$ terminal moiety in these dipeptides, there are asymmetric $\omega$(NH$_2$) motions observed in the 620–695 cm$^{-1}$
part of the spectra, located at a rather high frequency as a result of the strong H-bond formed between the terminal NH$_2$ and the neighbouring amide C=O group. While the asymmetric $\omega$/NH$_3$ motions have distinct signatures from the amide backbone $\omega$/NH ones, the symmetric $\omega$/NH$_3$ overlaps with $\omega$/NH of the amide free N-H groups in the 400–510 cm$^{-1}$ domain.

While the 400–800 cm$^{-1}$ domain is dominated by localized wagging motions, 0–400 cm$^{-1}$ is dominated by bending (100–400 cm$^{-1}$) and torsional (0–100 cm$^{-1}$) large amplitude and collective motions, coupled and delocalized over the dipeptide backbone. Some systematic features can be observed, i.e. three peaks between 700 and 750 cm$^{-1}$, a peak between 420 and 450 cm$^{-1}$ and a doublet located at around 500 cm$^{-1}$. While the doublet is characteristic of a weak C5 interaction, no signature of a C7 or C10 interaction was found.

Next steps into characterizing far IR/THz spectral signatures combining the experiments and DFT-MD based theoretical spectra would be the investigation of large peptides and peptidic assemblies. It will in particular be relevant to unravel the lower frequency delocalized torsional signatures and the inter-molecular H-bonded signatures, quantifying the extent of delocalization over the peptidic units.

Acknowledgements

This work was supported by funding from the European Community’s Seventh Framework Programme CFP7/2007-2013 under agreement no. 312284. We gratefully thank the FELIX staff for their experimental support and acknowledge NWO support of the FELIX laboratory under project no. N2300N. JM acknowledges financial support for traveling and living expenses through the EOLE program. This work was performed using HPC resources from GENCI-France (CINES/IDRIS Grant 072484). MPG also acknowledges that part of this work was developed during her period at Institut Universitaire de France (IUF).

References

Mapping gas phase dipeptides motions in the far-infrared and terahertz domain.

Jérôme Mahé, a Daniël J. Bakker, b Sander Jaeqx, b Anouk M. Rijs, *b and Marie-Pierre Gaigeot* a

Supplementary Material
### Contents

1. Experimental method ........................................ 3  
2. REMPI spectrum of Ac-Phe-Val-NH₂ .......................... 5  
3. Theoretical details for vibrational spectroscopy .............. 6  
4. Conformational assignment of Ac-Phe-Cys-NH₂ .............. 7
1 Experimental method

A schematic overview of the set-up that is used to conduct IR-UV ion-dip spectroscopy of neutral and cold molecules at the FELIX laboratory is displayed in Figure S1. The key components of the setup are highlighted, starting with the molecular beam expansion which is produced using a pulsed supersonic valve by expanding the gas into vacuum. The sample molecules are seeded into the expansion using the laser desorption technique. The sample molecules are desorbed from a graphite sample bar and are brought intact into the gas phase. A skimmer transmits the central part of the expansion into the differentially pumped interaction chamber, where the molecules are irradiated with IR and UV light to vibrationally excite and consecutively ionize them via resonant enhanced two photon ionization (RE2PI). Finally, the ions are directed into the reflectron TOF-MS detector assembly, where they are directed towards a microchannel plate (MCP) charged particle detector.

The experimental cycle is synchronised to the FELIX pulse generation timing using a digital delay generator, to ensure that the sample molecules are irradiated by the UV laser pulses directly following their interaction with the IR radiation.
**Fig. 1** Schematic representation of the set-up used for IR-UV double resonance spectroscopy. The key components that make up the Free Electron Laser (FEL), laser desorption sample preparation and Time-of-Flight mass spectrometer are highlighted. The translation stage used to precisely position the sample bar with respect to the desorption laser, and the relative positions of the pulsed valve nozzle and the skimmer are displayed in the photos in the inset (Reproduced with permission from "Gas phase IR spectroscopy and structure of biological molecules"1).
2 REMPI spectrum of Ac-Phe-Val-NH₂

In table 1 are gathered the wavelengths of the \( S_0 \to S_1 \) transitions used for the IR-UV ion dip spectroscopy experiments conducted here.

As an example, the REMPI spectrum of the Ac-Phe-Val-NH₂ system is presented in figure S2.

![UV excitation spectrum of Ac-Phe-Val-NH₂](image)

**Fig. 2** REMPI spectrum of Ac-Phe-Val-NH₂. Note that this spectrum has not been averaged and has been scaled of +18 cm\(^{-1}\) to compensate the lack of calibration of the UV laser.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Conformation</th>
<th>( \lambda_{UV} ) (cm(^{-1}))</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ac-Phe-Gly-NH₂</td>
<td>( \gamma )-turn</td>
<td>37486</td>
<td>REF(^2)</td>
</tr>
<tr>
<td>Ac-Phe-Ala-NH₂</td>
<td>( \gamma )-turn</td>
<td>37464</td>
<td>REF(^2)</td>
</tr>
<tr>
<td>Ac-Phe-Pro-NH₂</td>
<td>( \gamma )-turn</td>
<td>37435.5</td>
<td>REF(^3)</td>
</tr>
<tr>
<td></td>
<td>( \beta )-turn</td>
<td>37409</td>
<td>REF(^3)</td>
</tr>
<tr>
<td>Ac-Phe-Cys-NH₂</td>
<td>( \gamma )-turn</td>
<td>37325</td>
<td>REF(^4)</td>
</tr>
<tr>
<td></td>
<td>( \beta )-turn</td>
<td>37450</td>
<td>REF(^4)</td>
</tr>
<tr>
<td>Ac-Phe-Ser-NH₂</td>
<td>( \gamma )-turn</td>
<td>37390</td>
<td>REF(^4)</td>
</tr>
<tr>
<td>Ac-Phe-Val-NH₂</td>
<td>( \gamma )-turn A1</td>
<td>37472</td>
<td>Present work</td>
</tr>
<tr>
<td></td>
<td>( \gamma )-turn A2</td>
<td>37410</td>
<td>Present work</td>
</tr>
</tbody>
</table>

**Table 1** \( S_0 \to S_1 \) transition used to select each conformation of each single molecule investigated in the work.
3 Theoretical details for vibrational spectroscopy

Within Statistical Mechanics, the Fermi Golden Rule for calculating an infrared spectrum can be re-formulated using Linear Response Theory\textsuperscript{5,6}, and can thus be rewritten as the Fourier Transform of the time correlation function of the fluctuating dipole moment vector of the absorbing molecular system\textsuperscript{7,8}.

\[
I(\omega) = \frac{2\pi\beta\omega^2}{3cV} \int_{-\infty}^{\infty} dt \langle \delta\mathbf{M}(t) \cdot \delta\mathbf{M}(0) \rangle \exp(i\omega t) \tag{1}
\]

where \(\beta = 1/kT\), \(T\) is the temperature, \(c\) is the speed of light in vacuum, \(V\) is the volume. The angular brackets represent a statistical average of the correlation function of the molecular dipole moment vector \(\mathbf{M}(t)\), where \(\delta\mathbf{M}(t) = \mathbf{M}(t) - <\mathbf{M}>\) is the dipole fluctuation, with \(<\mathbf{M}>\) the time average of \(\mathbf{M}(t)\). The calculation in equation 1 is done in the absence of an applied external field. For the prefactor in eq. 1, we have taken into account an empirical quantum correction factor multiplying the classical line shape of the form \(\beta\hbar\omega/(1 - \exp(-\beta\hbar\omega))\), which was shown previously to give accurate results on calculated IR intensities\textsuperscript{9–11}. For more detailed discussions on quantum corrections, see for instance refs.\textsuperscript{12–14}. In eq. 1, there are no harmonic approximations made (on the dipole moment, on the potential energy surface, and vibrational modes are coupled by construction), this equation gives the whole infrared spectrum of a molecular system in one single calculation (i.e. band positions, band intensities and band shapes), all conformations populated over time are taken into account in the final spectral calculation.\textsuperscript{7,8}

Assignment of the absorption peaks in terms of internal motions is done with two complementary analysis tools, ICDOS (Internal Coordinates Density of States) and PCA (Principal Component Analysis). ICDOS analysis was already applied in our previous work on the far-IR spectroscopy of the Ac-Phe-Pro-NH\textsubscript{2} peptide\textsuperscript{3}. It is based on the Fourier transform of the time-dependent autocorrelation of given internal coordinates denoted IC (chosen as a distance, an angle or a dihedral angle):

\[
I_{\text{ICDOS}}(\omega) = \int_{-\infty}^{\infty} \langle IC_i(t) \cdot IC_i(0) \rangle \exp(i\omega t) dt \tag{2}
\]

thus providing signatures arising from one given IC\textsubscript{i} internal coordinate. These reflect rather localized motions.

As the far infrared domain is also expected to be dominated by more delocalised modes and/or combination of large amplitude motions (over the peptide backbone), which are not expected to be well described by one single local internal coordinate, PCA analysis has also been applied. This is a dimension reduction technique from which few principal components are extracted and explain a large proportion of the total sample variance of the \(n\) chosen variables (here selected internal coordinates). To that end, a covariance \((n,n)\) matrix is built upon the chosen internal coordinates, with each matrix element being \(\text{cov}_{ij} = \langle (IC_i(t) - \langle IC_i(t) \rangle)(IC_j(t) - \langle IC_j(t) \rangle) \rangle\) (\(t\) is time, \(\langle IC_i(t) \rangle\) is the time-average of the time-dependent internal coordinate \(IC_i(t)\), \(< ... >\) is a time average over the trajectory). The matrix is diagonalized and provides \(n\) eigenvalues also known as principal components (PC). Only few PCs are relevant for the description of the total motion built on the chosen IC\textsubscript{i} coordinates, and only the first and second PCA are commonly used without redundancy. Trajectories are subsequently projected onto the identified PCs, and the resulting projections \(\rho(t)\) are time-correlated and Fourier transformed.
in order to provide the final spectral assignment in terms of delocalized and large amplitude motions.

4 Conformational assignment of Ac-Phe-Cys-NH$_2$

In two recent papers from Yan et al.$^4$ and Alauddin et al.$^{15}$, the structure of the $\beta$-turn conformation of Ac-Phe-Cys-NH$_2$ is discussed. It is assigned to a $\beta$-turn conformation by both groups, but they do not agree about the orientation of the CH$_2$-SH residue with respect to the backbone, i.e. SH interacts with the C=O function in ref.$^{15}$ (denoted conformer FC1) while it interacts with the aromatic ring in ref.$^4$ (denoted conformer FC2). These two structures are shown in Figure S3.

**Fig. 3** Structures labelled FC1 (left) and FC3 (right) of the Ac-Phe-Cys-NH$_2$ dipeptide.

DFT-MD simulations have been performed for both conformers (following the details of section 2 in the paper), and the two IR dynamical theoretical spectra are found almost identical in both the 3000-4000 cm$^{-1}$ (Fig. S4) and 800-1800 cm$^{-1}$ (Fig. S5) ranges. The same comparison has been made in the far IR range (<800 cm$^{-1}$) and is presented in figure S6. The far-IR experimental spectrum has been recorded in the present work, and is also reported in the main text of the paper.

One can observe small but workable differences between the two theoretical spectra in the far-IR range, reinforcing the idea, discussed elsewhere$^2$ that the far-IR could indeed be a stronger tool than the other spectral ranges for conformational assignment. One can note two main differences between the two theoretical DFT-MD spectra of FC1 and FC3 conformers. The out of plane wagging motion of the SH function gives rise to two active IR peaks located at 288 and 311 cm$^{-1}$ for the FC1 conformer, while there is only one peak at 295 cm$^{-1}$ for the FC3 conformer. The second difference is visible for the peaks located between 425 and 525 cm$^{-1}$ arising from the out of plane motion of both NH$_2$ and backbone NH (belonging to the amide group of the phenylalanine residue). For these peaks, differences are mainly in the band shapes. The two peaks located at 288 and 311 cm$^{-1}$ observed for FC1 are found in excellent agreement with the two experimental peaks located respectively at 289 and 304 cm$^{-1}$, thus providing strong evidence that FC1 is the conformer probed in the experimental conditions. This is the conformer used in the analyses in the paper.
Fig. 4 In green and purple, the DFT-MD theoretical spectra of FC1 and FC3 conformers of Ac-Phe-Cys-NH$_2$ peptide, respectively, in the range 3000-4000 cm$^{-1}$. The experimental frequencies reported come from the publication of Alauddin et al$^{15}$. See structures in figure 3.
Fig. 5 In cyan, the experimental spectrum of Ac-Phe-Cys-NH$_2$ dipeptide (β-turn conformation) measured in the range 1400-1800 cm$^{-1}$. In green and purple, DFT-MD theoretical spectra of FC1 and FC3 conformers of Ac-Phe-Cys-NH$_2$, respectively. See structures in figure 3.
Fig. 6 In cyan, the experimental spectrum of Ac-Phe-Cys-NH$_2$ dipeptide (β-turn conformation) measured in the present work in the range 100-800 cm$^{-1}$. In green and purple, the DFT-MD theoretical spectra of FC1 and FC3 conformers of Ac-Phe-Cys-NH$_2$, respectively. See structures in figure 3.
References
Chapter 7

Anharmonic, dynamical and functional level effects in far-infrared spectroscopy. Test case on phenol derivatives.

The paper "Mapping gas phase dipeptide motions in the far-infrared and terahertz domain"\textsuperscript{65}, presented in chapter 6 has shown the quality of the BLYP-D3-MD method to reproduce far infrared spectra of peptidic systems with in particular the excellent result of an average \(6\ \text{cm}^{-1}\) difference between experimental and theoretical peaks positions, calculated over a series of 9 dipeptides and over all peaks in the 90-800 cm\(^{-1}\) range.

In the attached paper: "Anharmonic, dynamic and functional level effects in far-infrared spectroscopy: phenol derivatives"\textsuperscript{108}, a series of phenol derivatives has been studied with the aim of extending the domain of systems investigated and to check validity and robustness of theoretical methods (not only DFT-MD) in the far infrared spectral domain.

The far-infrared spectra of phenol and four ortho-substituted phenol derivatives, including three deuterated analogs are presented, \textit{i.e.} phenol, catechol, salicylic acid, saligenin and 2-nitrophenol. The systems investigated are presented in figure 7.1. The spectra have been measured using the same strategy used for the dipeptide series (chapter 6), combining IR-UV ion dip spectroscopy and the free electron laser FELIX (The Netherlands). The experimental method is presented in chapter 3 of this thesis and BLYP-D3-MD applied for theoretical spectra and spectral assignments is presented in chapter 4.

Due to the rather simplicity of these systems, the conformational assignment is not especially challenging, which allowed us to discuss theoretical representations for far infrared spectroscopy without questioning the 3D structural assignment. Indeed, only one possible conformation is possible for the phenol, the catechol and the nitrophenol molecules. Saligenin and salicylic acid conformational assignments have already been done respectively by Kumar \textit{et al.} using Microwave and IR-UV ion dip spectroscopy\textsuperscript{112} and by Yagahi \textit{et al.} using IR-UV ion dip
CHAPTER 7. FAR INFRARED SPECTROSCOPY OF PHENOL DERIVATIVES.

Figure 7.1: Molecular structures of the phenol derivatives species, optimized at the B3LYP-D3/6-311+G(d,p) level of theory. Black spheres represent carbon, white hydrogen, red oxygen and purple nitrogen. The hydrogen bonded OH bond-length $L_{OH}$ is listed as a measure of the strength of the hydrogen bond when present. This figure has been copied from the following paper: "Anharmonic, dynamic and functional level effects in far-infrared spectroscopy: phenol derivatives".[108]

We observe a good agreement between BLYP-D3-MD and experimental spectra at the exception of some missing peaks in the BLYP-D3-MD spectrum for at least phenol and saligenin. The BLYP-D3-MD method has been compared with spectra obtained using the harmonic approximation at the BLYP-D3 and B3LYP-D3 level of theory and with VPT2 anharmonic spectra at the B3LYP-D3 level, both methods are described in sections 4.5 and 4.6 of chapter 4. This allows us to evaluate whether the mismatch between experiment and theory for specific vibrational modes originates predominantly from functional errors or from harmonic/anharmonic effects, and whether some methods are better than others for anharmonicities. This paper (as well as chapter 10 presenting comparisons of methods for all the systems investigated in this thesis) shows the the DFT-MD spectra provide good results for all systems, harmonic spectra (computationally far less costly) are often found good except for the highly anharmonic $\omega$(OH) modes, while VPT2 spectra provide theoretical spectra less good than the DFT-MD spectra for a prize yet equivalent.

The torsional motions (or wagging) related to the OH groups in these molecules are specifically discussed in our paper because of the observed anharmonic character of these modes. Indeed, large differences are observed in the frequencies for the torsion of the hydrogen bonded OH between the harmonic and the MD spectra (we observe for example a small deviation of $< 10$ cm$^{-1}$ for the $\omega$(OH) modes for saligenin but a deviation of $\approx 100$ cm$^{-1}$ for one of the
\( \omega(OH) \) modes of catechol between these two methods). Note that while the anharmonic VPT2 spectra do not improve the predictions for the OH torsional motions, as the frequency shifts and peak intensities are often overestimated, the BLYP-D3-MD spectra constantly provide improved positions and intensities for these vibrational modes with respect to the experiment.

On the other hand delocalised modes appear to be harmonic and the four theoretical representations (BLYP-D3-MD, BLYP-D3 harmonic, B3LYP-D3 VPT2 and B3LYP-D3 harmonic) provide equivalent signatures. The B3LYP-D3 spectra appear to be slightly blue shifted in comparison to the BLYP-D3 spectra for these modes. The blue shift slightly improves the match between experiment and theory.

Moreover, the experimental far infrared spectra of phenol and saligenin show overtones and combination bands as proven by the measurements of their deuterated analogs. These bands are unfortunately missing in the MD spectra while they are reproduced by the VPT2 representation (even though positions of the associated bands are shifted with respect to the experiment). We find, for instance, a shift of \( \simeq 100 \text{ cm}^{-1} \) for the \( \omega(OH) \) \( \nu_0 \rightarrow \nu_2 \) overtone of phenol with respect to the experimental spectrum. In the BLYP-D3-MD spectra, these bands are missing without affecting the other bands, suggesting a small coupling between overtones or combinations bands and fundamental transitions. Once these missing peaks are excluded from the theory-experiment comparison, we find a good agreement between BLYP-D3-MD and experimental spectra, comparable to the average mean deviation of \( 6 \text{ cm}^{-1} \) for band positions found for the dipeptide series in chapter 6 (for example, we found a peak-to-peak mean frequency deviation of \( \simeq 8 \text{ cm}^{-1} \) for phenol, \( \simeq 10 \text{ cm}^{-1} \) for catechol and \( \simeq 8 \text{ cm}^{-1} \) for saligenin).

The experimental deuterated phenol spectrum is in fact much more complex to analyse than the phenol spectrum (although deuteration should in principle help the analysis). Using the vibrational modes of phenol-OH as reference, vibrational modes analyses of the deuterated system shows that the OD torsion couples with an out of plane deformation of the ring mode (located at \( 226 \text{ cm}^{-1} \) in phenol-OH) to produce the two peaks at 213 and 247 cm\(^{-1} \) in the phenol-OD spectrum, see section 9.2 of chapter 9 where the complete vibrational mode analysis is presented for the deuterated spectrum. The peak at 588 cm\(^{-1} \) in the phenol OH spectrum can be assigned to the \( \nu_0 \rightarrow \nu_2 \) overtone of the OH wagging. While the 588 cm\(^{-1} \) peak disappears in the phenol-OD spectrum, two \( \nu_0 \rightarrow \nu_2 \) overtones of the two modes OD torsion + out of plane ring deformations appear at the same time. The transition associated to the CCO in plane bending mode is also red shifted by \( 20 \text{ cm}^{-1} \) after deuteration due to the increase in the mass of the OH function. After deuteration, the larger weight of the deuterium atom makes the coupling with the other internal coordinates much easier and we thus observe different normal modes, which does not simplify our explanation/understanding of the spectroscopy of phenol unfortunately.
Anharmonic, dynamic and functional level effects in far-infrared spectroscopy: Phenol derivatives

Daniël J. Bakker, Qin Ong, Arghya Dey, Jérôme Mahé, Marie-Pierre Gaigeot, Anouk M. Rijs

1. Introduction

The molecular structure of gas phase neutral (bio)molecules, ranging from small molecules to more complex ones, is routinely determined by applying mid-infrared (mid-IR), ultraviolet (UV) double resonance spectroscopy in combination with density functional theory (DFT) [1–6]. With mid-IR spectroscopy, the resonant frequencies of local oscillators such as NH, OH or CO groups are probed, which show shifts in frequency in case they are involved in secondary interactions [7,8]. However, when a molecule has an extended apolar group, subtle changes in the orientation of this apolar group will not noticeably change its mid-IR spectroscopic features [9]. Moreover, when considering larger molecules with additional similar functional groups, these will each give rise to a unique, but close lying absorption frequency, so that spectral congestion is often observed [10]. These two issues limit the information on the molecular structure that can be extracted from the mid-IR spectra.

The far-infrared (far-IR) region of the spectrum (<800 cm⁻¹) has shown to be very useful for spectral characterization, specifically in these cases. Since the vibrations in the far-IR involve large parts of the molecule, subtle structural differences can be probed [9]. Moreover, shifts in vibrational modes due to electrostatic interactions are larger in the far-IR than in the mid-IR [11], making the far-IR more sensitive to secondary structure. The far-IR part of the spectrum, however, poses challenges both experimentally and computationally. Especially to retrieve structural information from the experimental spectra by computational methods requires theoretical approaches that go beyond the harmonic approximations, as was shown previously [12,13].
The far-IR spectra of solids and solvated molecules measured at room-temperature show broadening of absorptions to an extent where the assignment of individual normal modes becomes mostly impossible due to spectral congestion, interactions with solvent or lattice and absorptions thereof [14,15]. To assign these spectral features, similar harmonic static calculations are applied as used for gas phase experiments [16,17]. Measurements of solids in ultracold (~4 K) environment and gas phase samples both produce well-resolved features in the far-IR, using mainly dispersed fluorescence [18], direct absorption [19], terahertz time-domain spectroscopy [20] and Fourier transform IR spectroscopy [21]. The observed match between theory and experiment is however not as good as the match observed in the mid-IR part of the spectrum, dictating that more sophisticated quantum chemical calculations are necessary to unlock the full potential of far-IR spectroscopy.

Focusing on double resonance spectroscopy of neutral, gas phase molecules, we see that structural specificity and clear separation of features can be achieved in the far-IR spectrum [11,22]. However, application of standard quantum chemical techniques to interpret and assign these spectra usually does not yield conclusive results in the far-IR region of the spectrum. Even for small and medium-sized molecules more sophisticated methods are required [9]. This raises the questions: What characteristics of molecules and methodology present difficulties? Which methods provide the desired molecular information? Molecules can have a plethora of electrostatic, dispersion and ionic interactions in addition to various degrees of flexibility. Additionally, the couplings between oscillators in the far-IR should induce more delocalized motions in the far-IR and be the source of supplementary anharmonicities.

In this paper, we analyze these four aspects, i.e. flexibility, electrostatic interactions, couplings and anharmonicities, that complicate the comparison of theory and experiment, focusing on phenol and molecules derived from phenol. Hydrogen bonds (H-bonds) of increasing strength are introduced, as well as substituents of increasing flexibility. This dataset is used to benchmark one of the most frequently used levels of static harmonic calculations (the B3LYP-D3 electronic representation, with the 6-311+G** basis set [23,24], as well as BLYP-D3 representation) against a more sophisticated method for theoretical spectroscopy, namely Born-Oppenheimer Molecular Dynamics (BOMD using the BLYP-D3 electronic representation) [25] based anharmonic spectra calculations. BOMD takes anharmonicities of the potential energy surface and of the dipolar surface directly into account. Additionally, second order perturbation theory (VPT2) is applied here to account for anharmonicities at the static level, as it has been shown to be sufficient at reproducing normal modes of small to medium size gas phase molecules and clusters in the mid-IR [26,27]. This harmonic correction was previously shown to work well for specific far-IR modes [12]. However, here we show that this correction fails for the highly anharmonic, large amplitude OH torsional modes. Our paper systematically provides a comparison of static harmonic, static anharmonic and dynamical anharmonic spectra for phenol and four phenol derivatives, while also trying to assess relative performances of the hybrid B3LYP-D3 electronic representation and the pure gradient corrected BLYP-D3 one.

2. Experimental and theoretical methods

2.1. Experiment

The samples, namely phenol (PHNL), catechol (CAT), salicylic acid (SA), saligenin (SLG) and 2-nitrophenol (NP), were purchased from Sigma Aldrich with a purity exceeding 99%, and used without further purification. The deuterated samples were produced by dissolving the molecular compounds in a large molar excess of deuterated methanol (CH$_3$OD). After the exchange reaction, the methanol was removed from the sample under reduced pressure using a rotary evaporator. The solvation and drying steps were repeated three times.

The experimental set-up and spectroscopic methods have been explained extensively elsewhere [28]. The REMPI spectra were measured using UV light from a frequency doubled dye laser (Radian dyes) pumped by a YAG laser (Spectra Physics). The pulsed laser system delivers typical output energies of 1 mJ with a pulse length below 5 ns, with Coumarin 153 as laser dye, except in the case of SA, where the laser dye DCM was used. SA additionally required a two-color ionization scheme, where the 193 nm light from an excimer laser (Neweks, typical output power 1 mJ) is used to ionize the molecules upon excitation. NP was measured in a co-expansion with PHNL, which provided a tenfold increase of the NP signal level. The improved signal is expected to be a result of freezing point depression of the mixture of the two solids that improves the vapor pressure. The REMPI spectra of the expansion of pure NP and of the co-expansion are presented in Fig. SI.1 of the supplementary material. Both REMPI spectra show identical features showing that the increase in signal is not caused by cluster formation. Cold, gas phase molecules were produced in a skinned supersonic jet expansion, using a heated sample compartment and heated pulsed valve (Parker general valve), with either helium or argon as seed gas with a backing pressure of 3 bar. The ions produced in the REMPI process are detected using a reflectron time-of-flight mass spectrometer (TOF-MS, Jordan) equipped with 40 mm microchannel plates (MCP). The full experiment runs at a repetition rate of 20 Hz.

For the IR-UV ion dip experiments, the wavelength of the UV laser is set to resonantly ionize the molecules via excitation from the $S_0$ to the $S_1$ state, producing a constant ion signal. The characteristics of the experiments (heatable source temperature, UV wavelength, seed gas) are displayed in the supplementary information in Table SI.1. The UV pulses are preceded by tunable far-IR light pulses produced by the free electron laser FELIX, which will deplete the ground state when its frequency is resonant with a vibrational transition in the molecule and thereby resulting in a dip in the ion signal. The IR light is scanned within the range from 1100 cm$^{-1}$ down to 100 cm$^{-1}$, and has a macropulse length of 6–10 µs with a typical pulse energy of 50 mJ [29]. The absorption signal is normalized using alternating shots with and without IR light by operating the molecular beam valve and the UV laser at a repetition rate of 20 Hz, while FELIX runs at a repetition rate of 10 Hz. At least 75 individual measurements were averaged for every presented spectrum. The absorbance is calculated by taking the logarithm of the ratio of the intensities with and without IR light present. All absorbance spectra were subsequently corrected for the wavelength dependent photon count of the IR pulses.

2.2. Theoretical methods

The DFT-based molecular dynamics simulations in the Born-Oppenheimer representation (BOMD) were performed using the CP2K package [30]. The dynamics part of the simulation consisted of solving the Newton’s equations of motion at a finite temperature, treating the nuclei classically and the electrons quantum mechanically using the DFT electronic representation. The BLYP functional [31,32] was applied with the Grimme D3 correction [24], with a mix of a plane-wave basis set (kinetic energy cut-off at 450 Ry) and a Gaussian basis set of the aug-TZVP type. Periodic boundary conditions were applied to a cubic simulation box of 13Å$^3$ (neutral molecules). The optimal kinetic energy cut-off, basis set size and cubic box sizes were all determined from energy convergence tests, while also compromising between accuracy and computational cost.
An initial 4–8 ps of each trajectory was used for thermalisation of the gas phase molecule, i.e. redistribution of energy into the different vibrational modes of the molecules, as the atoms initially get a randomized Boltzmann-based velocity distribution centered on the simulation temperature (50 K). Changes in temperature that occurred during this equilibration were controlled by collectively rescaling all atoms' velocities, so that the average temperature converged to the targeted temperature step by step. The temperature in the simulation is chosen to be 50 K based on previous work, where this temperature provided spectra that showed a good match to the low-temperature experiments [9,13,33]. The choice of temperature in the simulation has to be comparable to the (expected) experimental one, while keeping in mind that MD temperature and curvature of the potential energy surface are correlated. A too high temperature would lead to sampling of too anharmonic parts of the PES, that would not be reasonable for the final spectral comparison to experiment. After the equilibration, trajectories were accumulated for 20 ps with a time step of 0.4 fs, without rescaling velocities, i.e. dynamics were conducted in the NVE microcanonical ensemble.

We calculated three separate trajectories for each molecule in order to get a statistical representation of its theoretical gas phase IR spectrum. Each trajectory differed by the initial randomized velocities chosen within a Boltzmann distribution, but all started from the same optimized geometry. In Fig. S1L2 of the supplementary information, the spectra of these three different temperatures are presented for five representative systems. Differences in between the intensities of individual features in each spectrum indicate that there is not a complete equilibration of the internal energy amongst all modes (i.e. equipartition not completely achieved). This affects only a few peaks in the spectra, where the ratio of intensities is the most affected. However, the differences in intensities are small, and one can see that once (here) three trajectories are averaged, these fluctuations are spread out, ensuring that a final theoretical spectrum averaged over three separate trajectories is already a good approximation.

The IR spectrum was calculated by taking the Fourier transform of the time correlation function of the fluctuating dipole moment vector of the absorbing molecular system [33–35]. Every trajectory included not only vibrations, but also end-over-end rotations of the molecule that can couple to the vibrations. As the rotational motions of the molecules were not quantized in the simulations, the interpretation of this effect was possibly questionable, so that the rotations were removed from the trajectories at each time-step of the dynamics (done a posteriori, solely for spectral analyses). To that goal, for all consecutive time steps of the trajectory, rotation transformations were defined that maximally overlap the positions of the atoms, effectively removing the overall rotation from the trajectory. By applying the same rotation transformations to the dipole vector at each time step, the effect of the rotation was also removed from the calculated spectrum. An example of a far-IR spectrum including and excluding the molecular rotational motion is presented in the supplementary information as Fig. S1L3.

The assignment of the active bands into local modes was performed by taking the Fourier transform of intramolecular coordinate (IC) time correlation functions, denoted ICDOS in our previous work [9]. No scaling factors were applied to the vibrations extracted from the dynamics since vibrational anharmonicities are intrinsically included in these simulations. Any remaining mismatches between the BOMD and experimental spectra should therefore be due to the choice of the BLYP-D3 functional and the combination of basis sets, as DFT-based dynamics can only be as good as the chosen level of theory. Moreover nuclear quantum effects were not taken into account in our trajectories, which could also induce some band shifts where light hydrogen atoms are involved [36,37].

The calculated static harmonic DFT spectra presented for every conformer were produced using the Gaussian software package [38], at the BLYP-D3/6-311+G** and B3LYP-D3/6-311+G** levels of theory. Second order vibrational perturbation theory (VPT2) was also applied at the B3LYP-D3/6-311+G** level of theory. It has been established that this method struggles to predict the intensity of heavily anharmonic modes [39]. An overview of the modes for which the harmonic intensity is displayed instead of the anharmonic one because of this effect is given in Table S1L2 in the supplementary information. All optimizations were performed using the tight keyword, and with an ultrafine mesh grid. No scaling factors were applied to the calculated spectra except when explicitly indicated (case of phenol). All calculated lines from static theory were broadened using a Gaussian line-shape with an FWHM of 0.5% of the wavenumber, which is the lower limit of what can be expected of the FELIX bandwidth.

For the assignment of vibrational bands in the measured spectra, all levels of theory were considered simultaneously to avoid a biased assignment of any specific level of theory, keeping in mind that each of the levels of theory has its own strengths and weaknesses, which will be elaborately discussed in this paper. In several cases, this approach led to multiple possibilities for the complete assignment of the spectrum, which will be described in the following sections if applicable.

3. Far infrared spectra

An overview of the structures of the studied molecules is presented in Fig. 1. Phenol (PHNL, Fig. 1a) is selected as a reference molecule to probe the accuracy of the theoretical methods for reproducing delocalized normal modes in a rigid, planar molecule in the absence of H-bonds or other electrostatic interactions. Catechol (CAT, Fig. 1b) is characterized by a second hydroxyl group in the ortho-position. A weak H-bond is present in CAT closing a five-membered ring [40]. The two rotamers of salicylic acid (SA-R2, Fig. 1c and SA-R1, Fig. 1d), which have a carboxylic acid group at the ortho-position, provide an intermediate hydrogen bond and a strong hydrogen bond, respectively [41]. Saligenin (SLG, Fig. 1e), the only non-planar molecule has a flexible hydroxymethyl group substituted in the ortho-position [42]. In the molecule 2-nitrophenol (NP, Fig. 1f), the less common NO₂ group is substituted at the ortho-position [43,44]. In both SA-R1 and NP, the strong hydrogen bond is caused by resonance assisted hydrogen bonding. This effect is present when two stable isomers, called resonance structures, can be formed where hydrogen bond donor and acceptor reverse roles [45].

We will discuss the vibrational modes related to the OH groups specifically because of the observed anharmonic character of these modes. In particular, we study the evolution of their frequency depending on the intramolecular H-bond formed, and how the different theoretical levels presented here are able to reproduce the experimental features [11]. Although these are highly relevant and active modes, other active far-IR vibrational modes can also be crucial for the structural characterization. Assignment of the vibrational bands has been done from the knowledge of the normal modes in the harmonic static calculations and from the careful analysis of the motions participating to the active bands in the dynamical anharmonic calculations (ICDOS spectra [13]). We find that both computational approaches usually provide very similar assignments of the atomic motions involved in the bands, with obvious differences in frequency depending on harmonic/anharmonic calculations. Large differences are observed in the frequencies for the hydrogen bonded OH torsion, of which the position is not correctly obtained by the harmonic calculations. Therefore,
the assignment of these bands relies on the anharmonic dynamical spectra.

We will discuss the static harmonic spectra at the B3LYP-D3 level, as this level of calculation is employed regularly in the literature of vibrational spectroscopy of gas phase molecules and clusters. We also employ the BLYP-D3 electronic representation in harmonic and anharmonic dynamical spectral calculations, as this generalized gradient corrected functional is frequently used (also in this paper) for the DFT-based molecular dynamics. This functional is computationally far less expensive than any hybrid functional (especially for MD simulations), and has been shown extremely robust and transferable for vibrational spectroscopy (gas phase, liquid phase, and interfaces) [33,37,46–48]. The B3LYP-D3 representation is also combined with non-harmonic static spectral calculations using the VPT2 method. Harmonic and anharmonic spectra are discussed with the two different DFT functionals, trying to appraise effects arising from the chosen DFT and effects arising from anharmonicities (treated differently in the two methods).

The following sections of the paper detail the large amplitude OH torsional motions arising in the far-IR spectral domain of the phenolic molecules investigated here. Red- and blue-shifts due to hydrogen bonding are discussed with respect to the strength of the H-bonds formed. The associated bands are systematically highlighted in blue and green in all spectra reported in Figs. 2–9.

![Molecular structures of the measured species, optimized at the B3LYP-D3/6-311+G** level of theory. Black spheres represent carbon, white hydrogen, red oxygen and purple nitrogen. The H-bonded OH length $L_{OH}$ is listed as a measure of the strength of the H-bonds when present. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image1.png)

![Far-IR spectrum of phenol (PHNL). Black: experiment, Red: BOMD BLYP-D3 dynamical anharmonic spectrum, Blue: static harmonic BLYP-D3 spectrum, Green: static harmonic B3LYP-D3 spectrum, Teal: static anharmonic VPT2 correction on the static harmonic B3LYP-D3 spectrum. The rescaled BLYP-D3 spectrum is shown as the dashed blue line, with a scaling factor of 1.0304, as determined from spectral lines of PHNL, PHNL-D and CAT (see text). The overtones and combination bands predicted by the VPT2 method are included as the dashed teal line. The out-of-plane torsional free OH mode is highlighted in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image2.png)
Beside these modes, the far-IR spectra of all the phenolic molecules investigated here show supplementary similar features, which we will also describe in the following. In a nutshell, bands appearing in the 700–800 cm\(^{-1}\) domain result from out-of-plane (oop) wagging motions of the CH ring groups (CH oop). Such motions systematically appear in this particular domain and are maintained from one molecule to the next one. Depending on the type of OH substitutions on the ring, more than one active band of CH oop will be observed in the spectrum, reflecting specific symmetry selection rules that we will not discuss here. Coupled delocalized motions over the ring (typically C–C–C bending vibrations and C–C–C–C oop torsional motions) typically provide two active signatures located in the 400–550 cm\(^{-1}\) interval for the PHNL molecule, while there is one single band located roughly at 300 cm\(^{-1}\) for CAT. Such delocalized motions of the ring can also be coupled to CH oop motions, as this appears with the 752 cm\(^{-1}\) (in-plane (ip) ring deformation) band observed for CAT.

Although we mainly focus on the 200–800 cm\(^{-1}\) region, for saligenin and nitrophenol the spectra are extended to 1100 cm\(^{-1}\), as calculations revealed these regions to be quite distinctive. Here, the COH bending motions systematically give rise to active bands located beyond 1000 cm\(^{-1}\), typically around 1200 cm\(^{-1}\) for an H-bond accepting OH group and around 1400 cm\(^{-1}\) when the OH group is involved as an H-bond donor (although 1400 cm\(^{-1}\) is not displayed here). Saligenin is an exception to these rules, as the supplementary bands arising from the COH bending motion are also observed in the 900–1000 cm\(^{-1}\) region.

### 3.1. Phenol

Phenol (PHNL) is a rigid molecule, resulting in a modest number of vibrational modes in the far-IR spectrum, as can be seen in Fig. 2. In the absence of a hydrogen bond and with little flexibility within the molecule, the static harmonic B3LYP-D3 spectrum reproduces the experimental spectrum in the far-IR very well. All modes match within a deviation of 10 cm\(^{-1}\) and the relative intensities match the measured spectrum as well, except for the overestimation of the intensity of the OH torsional normal mode measured at 309 cm\(^{-1}\) (highlighted in blue in Fig. 2). Note that the peak measured at 588 cm\(^{-1}\) is not present in this calculated spectrum (nor in the others), suggesting that this peak may result from an overtone of the OH torsional normal mode, as will be discussed in more detail in Section 4.

The static harmonic BLYP-D3 spectrum results in the same normal modes and relative intensities as the B3LYP-D3, with the frequencies being systematically red-shifted with respect to B3LYP-D3. This shift increases for increasing wavenumbers, which implies that a linear scaling can be used to overlap the BLYP-D3 and B3LYP-D3 spectra for PHNL. We have found that a scaling factor of 1.0304 for the harmonic BLYP-D3/6-311+G** data should be applied in order to rescale band positions, which has been derived from 16 different lines for PHNL, PHNL-D and CAT data. The rescaled BLYP-D3 spectrum is displayed as the dashed line in Fig. 2 improving the agreement between theory and experiment. Normally, scaling factors are used to include anharmonicity and are smaller than unity. Therefore, the necessity for the application of a scaling factor larger than unity to the static harmonic BLYP-D3 data makes that this functional is not recommended for use with static harmonic DFT for far-IR applications. Since the BOMD dynamical anharmonic spectra do not require any scaling (as anharmonicity is intrinsically taken into account), we will present all further data without scaling factors to simplify the comparison.

The BOMD BLYP-D3 dynamical anharmonic spectrum (red trace) shows a light red shift with respect to the experimental spectrum for all bands, except for the OH oop vibration. It mostly follows the static harmonic BLYP-D3 (blue) spectrum, although the intensities deviate as can for example be seen for the CH oop motions between 650 and 800 cm\(^{-1}\). In the BOMD dynamical spectrum, the band above 800 cm\(^{-1}\) (not observed experimentally as it lies outside the studied wavelength range) is predicted to be the most intense, while in the static BLYP-D3 calculation the middle peak dominates.

The VPT2 B3LYP-D3 anharmonic spectrum (teal line, Fig. 2) provides two significantly shifted modes with respect to the harmonic B3LYP-D3 spectrum, decreasing the quality of the agreement with the experiment. In all calculated spectra, the three intense peaks observed in the experimental spectrum, resulting from the C–H oop and OH torsional motions, are well reproduced in intensity by the calculations. The 309 cm\(^{-1}\) (experimental value, highlighted in blue) oop OH torsion is slightly overestimated in position by both the static and dynamic anharmonic spectra, respectively by +22 cm\(^{-1}\) for the BLYP-D3 dynamical spectrum and by +24 cm\(^{-1}\) for the B3LYP-D3 VPT2 spectrum, while the harmonic band (either BLYP-D3 or B3LYP-D3) matches the experiment well.

The other, much smaller, bands observed in the far-IR spectrum for PHNL result from another CH oop motions located at 501 cm\(^{-1}\) in the experiment (besides the 700–800 cm\(^{-1}\) bands). The 400 cm\(^{-1}\) band is due to CO ip bending motion while the lowest frequency band at 226 cm\(^{-1}\) arises from collective oop deformations of the ring (normal mode 10b in the Varsanyi notation for
mono-substituted benzenes [49]). This band will be discussed in more detail for the other molecules as it systematically appears in the experiments with relevant intensity. Note that all calculated spectra for PHNL show an extremely low intensity for that particular mode, while it will appear with significant intensity for the other molecules.

3.2. Catechol

In catechol (CAT), the formation of one H-bond is introduced by the addition of a second hydroxyl group to PHNL in the ortho position, while keeping the rigidity of the molecule very comparable to PHNL [40]. The number of vibrational modes present in the far-IR is still modest. All these modes are reproduced well by the static harmonic B3LYP-D3 spectrum (green trace), except for the H-bond accepting OH torsional normal mode at 220 cm\(^{-1}\), which is predicted at 139.6 cm\(^{-1}\) at this level of calculation. In contrast, the H-bond donating OH torsional mode, observed at 414 cm\(^{-1}\), is reproduced rather well and is red-shifted by only 10 cm\(^{-1}\) from the experiment. The two static harmonic spectra calculated with the BLYP-D3 and B3LYP-D3 electronic representations are very similar, with similar unreliable positioning of the hydrogen bond accepting OH torsional motion. The B3LYP-D3 calculation provides a better match with the experiment for the CH oop motions and ip ring deformations (700–800 cm\(^{-1}\)).

The BOMD spectrum shows a very good match with the experiment, although here the bands are also slightly red shifted, except for the H-bonded OH modes. The BOMD dynamical anharmonic spectrum very convincingly shows its strength in simulating the effects of anharmonicity in large amplitude motions. The two OH torsional vibrations, respectively located in the experiment at 220 cm\(^{-1}\) (acceptor of the H-Bond, highlighted in blue) and 414 cm\(^{-1}\) (donor of the H-Bond, highlighted in green), are indeed precisely matched by the BOMD BLYP-D3 spectrum. The torsional motion of the OH donor is blue-shifted by only 8 cm\(^{-1}\). Such an excellent agreement is not observed for the harmonic BLYP-D3 spectrum, therefore we can conclude that the spectral mismatch observed for the static calculations (BLYP-D3 and B3LYP-D3) results from the application of the harmonic approximation to these large amplitude torsional motions.

The vibrational CH oop and ip deformation modes observed at 735 and 763 cm\(^{-1}\) are described well by all levels of theory with shifts in the order of about 15 cm\(^{-1}\). The calculated spectra using the B3LYP-D3 functional (both the harmonic as the anharmonic VPT2 spectra) match the experiment in this particular domain better than the calculations based on the BLYP-D3 functional (static or dynamic). We therefore expect that the BOMD dynamical anharmonic spectrum will provide a better match to experiment when a hybrid functional is used. Unfortunately, this is not tested yet, as such representation is still computationally too costly for the available resources. These normal modes are very indicative for the phenyl group, yet are not diagnostic of its structural surroundings. Therefore we did not investigate the possibility of this more computationally demanding representation. Apart from the excellent agreement for the CH oop and ip deformation modes, the VPT2 anharmonic B3LYP-D3 spectrum shows a poor agreement with the experiment.

Once a H-bond has been formed in CAT, the donating oop torsional OH mode (at 414 cm\(^{-1}\)) is blue-shifted by roughly 100 cm\(^{-1}\) with respect to the free OH observed in PHNL (309 cm\(^{-1}\)), while the accepting OH mode (at 220 cm\(^{-1}\)) is red-shifted by roughly the same amount. Note that the donation of an H-bond by an OH group systematically leads to a blue-shift of the associated torsional mode frequency in the far-IR domain [11,13]. The band observed at 309 cm\(^{-1}\) in the experimental spectrum of CAT arises from the two CCO bending motions, is also red-shifted by 100 cm\(^{-1}\) from the CCO bending mode in PHNL where the OH was free of hydrogen bonding. All calculated spectra predict a low intensity mode around 180–210 cm\(^{-1}\) due to the oop deformations of the ring, which is barely visible in the experiment (contrary to PHNL).

3.3. Salicylic acid

The two stable conformers for Salicylic acid (SA), depicted in Fig. 1c (SA-R2) and 1d (SA-R1), have very different H-bond strengths [41].

3.3.1. Salicylic acid: Conformer SA-R2

In SA-R2 (Fig. 1c), the hydroxyl group donates an H-bond of intermediate strength to the acidic OH group. As the far-IR spectrum of Fig. 4 shows, the mode density increases in comparison to PHNL and CAT due to the increased size and complexity of the substituent, despite the fact that the studied molecule remains

![Fig. 4. Far-IR spectrum of salicylic acid rotamer 2 (SA-R2). Black: experiment, Red: BOMD BLYP-D3 dynamical anharmonic spectrum, Blue: static harmonic BLYP-D3 spectrum, Green: static harmonic B3LYP-D3 spectrum, Teal: static anharmonic VPT2 correction on the static harmonic B3LYP-D3 spectrum. Both experimental and theoretical intensities are multiplied by 5 in the region below 435 cm\(^{-1}\). The out-of-plane torsional free OH mode is highlighted in blue; the out-of-plane torsional mode of the hydrogen bonded OH group is highlighted in green. Note that although only a single feature is highlighted for the out-of-plane torsional free OH motion (blue) for each level of theory, this motion is present as a part of all normal modes positioned inbetween 390 and 570 cm\(^{-1}\). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image-url)
rigid. Assigning the observed vibrational peaks using the static DFT spectra becomes significantly more challenging: There are numerous peaks in the far-IR spectrum and most calculated peaks deviate significantly from the experiment in both position and intensity. Additionally, the experimental peaks seem broader than the expected resolution determined by the bandwidth of the FELIX free electron laser, which complicates the assignment.

In this respect, the agreement between the experimental spectrum and the harmonic BLYP-D3 and B3LYP-D3 spectra is rather poor. For example, both predict the presence of an intense band between 450 and 500 cm\(^{-1}\) (resulting from the oop OH torsional motion), which is not observed experimentally in this spectral range. The anharmonic VPT2 correction does not improve the agreement with the experiment. The BOMD spectrum significantly improves the match between theory and experiment, similarly as we observed for CAT, although the dynamical spectrum largely overestimates the intensity of the bands in the 400–600 cm\(^{-1}\) region. Additionally, all bands above 500 cm\(^{-1}\) are red-shifted by 20–30 cm\(^{-1}\) compared to the experiment, but all the observed features provided by the dynamical spectrum follow the experiment directly. The triplet resulting from the OH oop vibrations between 700 and 800 cm\(^{-1}\) is well reproduced by all levels of theory. However, the BOMD spectrum also reproduces the broadened peak at 620 cm\(^{-1}\) and the broadened triplet in between 500 and 600 cm\(^{-1}\) that both have a heavy oop OH torsional character. This supports the result we acquired for CAT, namely that the OH torsional vibrations are well reproduced by the dynamical anharmonic spectra, but poorly by static calculations.

In more detail, the H-bond accepting OH oop torsional mode is coupled to three other skeletal modes in the dynamical anharmonic spectrum, respectively located at 416, 532 and 548 cm\(^{-1}\), matching the experimental features within 10–30 cm\(^{-1}\), however too intense. Replacing the simple OH substituent in CAT by the acidic COOH moiety in SA-R2 results in a splitted band pattern and in a blue shift of 200–300 cm\(^{-1}\) in frequency as a result of the acidic environment of the OH group. Concurrently, the H-bond donating OH oop torsional motion is now found at 600 cm\(^{-1}\) in the dynamical anharmonic spectrum, to be compared to the 628 cm\(^{-1}\) band in the experiment: this mode is also blue-shifted by 200 cm\(^{-1}\) from its signature in CAT, reflecting the intramolecular H-Bond formed in SA-R2.

The triplet of bands in between 700 and 800 cm\(^{-1}\) in the experimental and dynamical spectrum corresponds to the CH oop motions (the two bands at the extremities of the triplet) and ip deformations of the ring (middle peak). As for PHNL and CAT, the computational spectra predicts two bands at around 230 cm\(^{-1}\), the lower frequency one originates from the oop deformations of the ring, while the higher frequency one is associated to bending motions of the acid CCC sequence. Both peaks are barely distinguishable in the experimental spectrum.

### 3.3.2. Salicylic acid: Conformer SA-R1

In SA-R1, the hydroxyl group acts as the H-bond donor, showing a strong, resonance assisted H-bond to the acidic oxygen atom \([11]\). As can be seen in Fig. 5, the increased strength of the H-bond and the co-existence of two resonance structures reduce the complexity of the far-IR spectrum. This results in an improved agreement between experiment and theory below 700 cm\(^{-1}\) for the two static methods compared to SA-R2. The increased rigidity introduced by the stronger H-bond benefits static theories.

Contrary to SA-R2, the BOMD spectrum shows only a slight improvement over the static spectra. However, BOMD does a remarkable job at predicting the intensity and the shape of the 567 cm\(^{-1}\) measured band, while the static harmonic and anharmonic B3LYP-D3 spectra reproduce only the position of this peak. This band, which arises from the oop torsional motion of the free OH group of the acid, requires the description of the subtle dynamical behavior in order to reproduce the experimental pattern.

When we compare the static and dynamic BLYP-D3 spectra, one can furthermore see that the agreement for the quadruplet between 700 and 800 cm\(^{-1}\) is improved in the BOMD spectrum. The two higher frequency peaks in that quadruplet are due to the oop motion of the OH donor in the intramolecular HBond: they are blue-shifted by 100–120 cm\(^{-1}\) compared to SA-R2. Thereby confirming the large increase in the hydrogen bond strength in the SA-R1 conformer. The two other peaks around 700 cm\(^{-1}\) are the standard CH oop and ip ring deformation modes. This assignment is not in line with the B3LYP-D3 results, as the bound OH torsion is predicted to be the second from the left peak of the quadruplet, which would then be assigned to the measured feature at 719 cm\(^{-1}\). A doublet is observed at 241 and 247 cm\(^{-1}\) in the experimental spectrum resulting from the oop ring deformations and bending of the acid CCC sequence, which is predicted by all explored levels of theory. Due to the small spacing between the peaks, this band is observed as a single peak in the BOMD spectrum, while 0 K line spectra of the static calculations can discriminate between these two close lying peaks.

To summarize, the VPT2 anharmonic correction on the harmonic B3LYP-D3 spectrum does not bring significant overall improvements, some positions improve some worsen. For example

![Fig. 5. Far-IR spectrum of salicylic acid rotamer 1 (SA-R1).](image-url)
the agreement for the acidic OH oop torsional mode (567 cm\(^{-1}\)) has improved, while the shifts in the quadruplet peaks worsen the match by VPT2. The BOMD method slightly improves the overall agreement with the experiment, allowing to assign all vibrational modes of this complex spectrum.

3.4. Saligenin

In saligenin (SLG), the flexible CH\(_2\)OH moiety is introduced, which not only accepts the intramolecular H-bond donated by the hydroxyl group, but that also bends oop to form a weak dispersion interaction with the pi-cloud of the aromatic ring \[42\]. The IR spectrum of SLG, see Fig. 6, was measured up to 1100 cm\(^{-1}\). The different levels of theory all show strong IR activity in the range from 800 to 1100 cm\(^{-1}\), but predict different positions for the involved vibrational modes. Overall, the four theoretical spectra present very similar spectral features from 200 cm\(^{-1}\) up to 800 cm\(^{-1}\) for both the peak positions and peak intensities. In more detail, the BOMD spectrum performs better in the 700–800 cm\(^{-1}\) region than the other methods (harmonic or VPT2 anharmonic), while all methods perform equally moderate in the 550–630 cm\(^{-1}\) domain (zoom in Fig. 6). The doublet of the oop ring deformation and CCC bending of the CH\(_2\)OH moiety above 200 cm\(^{-1}\) is consistently present in all methods.

The assignment of the signatures of the oop torsion of the two OH groups is mainly based on the BOMD spectrum. The OH group that acts as H-Bond acceptor has its torsional motion peaked at 336 cm\(^{-1}\) and 384 cm\(^{-1}\) in the experiment (340 cm\(^{-1}\) and 400 cm\(^{-1}\) in the BOMD spectrum). This is blue-shifted by roughly 30–100 cm\(^{-1}\) from the free OH torsion in PHNL, while the H-bond accepting OH torsion in CAT displays a red-shift compared to the free OH torsion in PHNL. The blue-shift in SLG is thus indicative that the CH\(_2\)OH group indeed engages in a weak OH–\(\pi\) interaction with the aromatic ring. The oop torsional motion of the H-bond donating OH moiety absorbs at 690 cm\(^{-1}\) (experimental spectrum). The blue-shift of this mode is large with respect to the reference mode observed in CAT (414 cm\(^{-1}\)), while the mode is moderately red-shifted from the corresponding modes obtained for SA-R1 (doublet around 750 cm\(^{-1}\)). The band below 800 cm\(^{-1}\) (band-shape is remarkably given by the dynamical spectrum) is associated to the CH oop motions.

The three bands in the experiment between 900 and 1000 cm\(^{-1}\) are remarkably well reproduced by the harmonic B3LYP-D3 spectrum, while the anharmonic correction worsens the correctly predicted peak positions. These bands are all assigned to systematically mixed normal modes which consist of CH oop, CH\(_2\) wagging and COH bending motions, with a rather strong component of CO stretching in the CH\(_2\)OH moiety in the higher frequency mode of that triplet. As previously reported, the BOMD spectrum shows a reduced agreement above 800 cm\(^{-1}\) and shows only one intense band around 900 cm\(^{-1}\).

VPT2 provides an alternative assignment for the oop OH torsion of the H-bond donor, as it predicts this OH torsional mode at 591.5 cm\(^{-1}\), which matches quite closely to the more intense (compared to the peak observed at 690 cm\(^{-1}\)) experimental feature at 586 cm\(^{-1}\). However, static anharmonic VPT2 theory based on B3LYP-D3 has shown to exaggerate the anharmonic shifts in the OH torsion features in the other molecules. Therefore, this agreement could be a coincidence. Moreover, comparing the measured to the calculated spectra for SLG, we see that the intensity of the measured spectrum is significantly reduced, complicating assignments. To confirm our assignments and go beyond difficulties in theories, deuterated analogs of SLG were measured, which will be discussed in Section 4.

3.5. Nitrophenol

The NO\(_2\) moiety in the molecule nitrophenol (NP) is expected to introduce strong, anharmonic behavior. Fig. 7 shows the IR spectrum of NP, which was measured in the range from 250 to 600 cm\(^{-1}\) using the high-signal NP-PHNL mixture, while the rest of the spectrum up to 1100 cm\(^{-1}\) was recorded using pure NP (see experimental section). Smoothing was applied to the measured spectrum (Savitzky-Golay, \(n = 5\)) to increase visibility of individual features; the unsmeothed data is shown as grey dots.

Both harmonic spectra are in reasonable agreement with the experiment, surprisingly with a better match for the BLYP-D3 spectrum especially for the hydrogen bonded OH vibrations highlighted in green. However, none of the harmonic calculations shows the intense doublet of peaks measured at 254 and 284 cm\(^{-1}\), predicting only a single band. The corrected VPT2 B3LYP-D3 spectrum also has the same deficiency for this doublet. Only the BOMD spectrum shows a doublet (oop ring deformation at lower frequency and bending of the CCN sequence at higher frequency). The BOMD spectrum shows an overall better match to the experimental features than the harmonic spectra, both in peak positions as in intensities, especially, for the 200–800 cm\(^{-1}\) region as was observed before \[9\].
The doublet located at 666 cm$^{-1}$ and 683 cm$^{-1}$ in the experiment arises respectively from the CH out-of-plane motions and H-bonded OH in-plane torsional motion. These vibrations are only slightly red shifted in the BOMD spectra, as we observed also for the other molecules, but BOMD provides the same peak spacing as the experiment. Both NP and SA-R1 provide the most blue-shifted torsional OH modes amongst the molecules studied in this paper, showing the strongest intra-molecular H-Bond. This is corroborated by the fact that both molecules form resonance assisted H-bonds.

The triplet in the dynamical anharmonic spectrum just beyond 800 cm$^{-1}$ results from the ONO bending motions mixed with the large amplitude OH torsional motions, the middle peak originates from a pure OH torsion. The bands measured at 710 and 840 cm$^{-1}$ receive contributions from coupled CH out-of-plane and O–H in-plane torsional motions. Ring CC stretching start to appear in the 1000 cm$^{-1}$ experimental peak, blue-shifted at 1050 cm$^{-1}$ in the BOMD spectrum, but well calculated in the VPT2 B3LYP-D3 spectrum. These stretching motions are systematically blue-shifted whenever a BLYP representation is applied, which has also been shown previously [13].

4. Overtones and deuterated molecules

4.1. Phenol-D

Not all features in the far-IR spectrum of PHNL (Fig. 2) could be assigned. Therefore, deuterated PHNL, of which the hydrogen atom of the OH group is exchanged by a deuterium atom, was measured to support the assignment. The REMPI spectra of the deuterated sample, presented in Fig. S.l.4, were recorded for the PHNL and the PHNL+1 amu mass channels. The REMPI spectrum in the PHNL+1 amu mass channel shows a peak at 36,346 cm$^{-1}$ resulting from PHNL-D, which is shifted with respect to the collection of origin peaks of the PHNL isomer with a $^{13}$C atom at 36,352 cm$^{-1}$. This allowed us to record a pure PHNL-D UV excitation spectrum, and thus an unperturbed IR absorption spectrum of PHNL-D.

Based on the increased value of the reduced mass for PHNL-D at the OD group, the modes involving the OD moiety will red-shift with respect to the equivalent modes involving the OH moiety in PHNL. When comparing the IR spectrum of PHNL (dotted line) and PHNL-D (solid line), both presented in Fig. 8, three large differences are observed. The intense peak of PHNL at 309 cm$^{-1}$, resulting from the large amplitude OH torsional motion, is red-shifted to

![Fig. 7. Far-IR spectrum of 2-nitrophenol (NP). Black: experiment, Red: BOMD BLYP-D3 dynamical anharmonic spectrum, Blue: static harmonic BLYP-D3 spectrum, Green: static harmonic B3LYP-D3 spectrum, Teal: static anharmonic VPT2 correction on the static harmonic B3LYP-D3 spectrum. The measured spectrum was smoothed (Savitzky-Golay, n = 5), grey points represent the unsmoothed data. The out-of-plane torsional mode of the hydrogen bonded OH group is highlighted in green. Two quite intense features at 1123 and 1141 cm$^{-1}$ in the dynamical spectrum are displayed with a reduced intensity to enable a better comparison of the spectra. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image1)

![Fig. 8. Far-IR spectrum of deuterated phenol (PHNL-D). Black: experiment, Red: BOMD BLYP-D3 dynamical anharmonic spectrum, Blue: static harmonic BLYP-D3 spectrum, Green: static harmonic B3LYP-D3 spectrum, Teal: static anharmonic VPT2 correction on the static harmonic B3LYP-D3 spectrum. The far-IR spectrum of PHNL is displayed as a dashed line for reference (in grey). The overtones and combination bands predicted by the VPT2 method are included as the dashed teal line. The out-of-plane torsional free OH mode is highlighted in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image2)
247 cm\(^{-1}\) in PHNL-D. Additionally, the low intensity feature of PHNL at 226 cm\(^{-1}\) gains intensity by coupling to the large amplitude OD torsion, which results in two intense peaks for PHNL-D at 213 and 247 cm\(^{-1}\). Secondly, the peak at 588 cm\(^{-1}\) present for PHNL disappears for PHNL-D, showing it can indeed be assigned as an overtone of the intense 309 cm\(^{-1}\) peak of PHNL, as was also shown previously in literature [50]. Thirdly, two new, small peaks appear in the spectrum of PHNL-D at 419 and 446 cm\(^{-1}\). These peaks can be assigned as overtones of the two coupled OD torsional vibrations (at 213 and 247 cm\(^{-1}\)). Both the OH torsional vibration and the coupled OD torsions show measurable overtones.

The BOMD spectrum of PHNL does not show the observed overtone. The VPT2 static anharmonic spectrum does predict its existence, however highly blue-shifted from the experiment (588 cm\(^{-1}\) in the experimental spectrum of PHNL, 687 cm\(^{-1}\) predicted by VPT2, visible in the dashed cyan line in Fig. 2). The position of the overtone in the spectrum of PHNL has been predicted with surprising accuracy, to within 2 cm\(^{-1}\) by Bist et al. using an analytical local mode hindered rotor calculation [51]. The OD oop torsional motion overtones in PHNL-D are predicted by the anharmonic B3LYP-D3 VPT2 method at 445 cm\(^{-1}\) (overtone of the 213 cm\(^{-1}\) band) and 501 cm\(^{-1}\) (overtone of the 247 cm\(^{-1}\) band), visible in the dashed cyan line in Fig. 8. The local mode analysis of Bist predicts only one single overtone at 476 cm\(^{-1}\) for the OD torsional local mode. This does not match either of the measured features as well as it did for PHNL, since the OD torsion feature is more strongly coupled to the collective oop deformations of the ring positioned at 226 cm\(^{-1}\). An alternative assignment for the two experimental peaks at 419 and 446 cm\(^{-1}\) is provided by the combination band analysis of VPT2, since the combination band of the two intense low-frequency bands is predicted to be at 474 cm\(^{-1}\), with a higher intensity than predicted for the overtones. Disentangling these alternative assignments is however outside the scope of this paper.

### 4.2. Saligenin-D

Deuterated isotopologues were also measured for SLG and presented in Fig. 9. Two OH groups are present in SLG, and each of the hydrogen atoms can be exchanged for deuterium. Therefore, two singly deuterated isomers can be observed, i.e. SLG deuterated on the free OH group (SLG-FD) and SLG deuterated on the bound OH group (SLG-BD), see inset of Fig. 9a and b respectively. Doubly deuterated SLG was only very weakly present. The REMPI spectra measured in the SLG and the SLG+1 amu mass channels are presented in the supplementary information as Fig. SI.5. The REMPI spectrum shows a clear signature for both the SLG-FD and the SLG-BD species. Therefore, molecule selective IR spectra of the two deuterated isomers, presented in Fig. 9, could be recorded.

The fingerprint region shown here from 700 to 1100 cm\(^{-1}\) containing the COH bending vibrations, shows large changes for both isomers. For both SLG-BG (Fig. 9a) and SLG-FD (Fig. 9b), these vibrations are well reproduced by the harmonic B3LYP-D3 spectrum and within acceptable agreement by the B3LYP-D3 anharmonic VPT2 spectrum. As seen before with the other molecules,
the BOMD spectrum does not reproduce these features above 700 cm\(^{-1}\) well.

Most part of the far infrared spectra \((<800\text{ cm}\,\text{cm}^{-1})\) remain unchanged after deuteration, which indicates no strong coupling of the backbone vibrational modes with the torsional OH modes once deuteration is applied. In more detail, the oop motions of the hydrogen atoms of the aromatic ring of deuterated-SLG appear at similar frequencies as normal SLG, namely at 727 and 756 cm\(^{-1}\). Modes involving the ip deformation of the cycle can be found at 561 and 621 cm\(^{-1}\) in SLG, and the equivalent modes are located at 568 and 625 cm\(^{-1}\) for SLG-BD although with a very weak intensity. Only one peak at 618 cm\(^{-1}\) is observed for SLG-FD. For the CCO bending involving the oxygen atom of the OD/OH group that acts as the H-bond donor, a shift of 20 cm\(^{-1}\) is observed for SLG-BD \((421\text{ cm}\,\text{cm}^{-1}\) for SLG versus 446 cm\(^{-1}\) SLG-BD), while the signature of SLG-FD remains the same \((424\text{ cm}\,\text{cm}^{-1})\).

Only few peaks in the spectra of SLG and its isomers are really affected by this deuteration suggesting that these OH torsional modes are mostly local vibrations. For SLG-FD, the free OD torsion does not couple strongly with other local modes, resulting in a single large peak at 300 cm\(^{-1}\) in the BOMD spectrum. The bound OH torsional motion is mainly localized and observed experimentally at 682 cm\(^{-1}\) \((\text{BOMD at } 690\text{ cm}^{-1})\). The shift of only 9 cm\(^{-1}\) upon deuteration indicates that the H-bond donor and acceptor OH groups are only weakly coupled to each other.

The small peak found at 690 cm\(^{-1}\) for SLG \((\text{dotted line in Fig. } 9)\) is not observed for SLG-BD, while it is observed (albeit slightly red-shifted) for SLG-FD. This indicates that this feature is related to the H-bonded OH group. For SLG-BD, both torsional CCOH motions \((\text{of the H-bonded and the free OH moieties})\) show up in a local peak at 44 cm\(^{-1}\) in BOMD data, indicating that they are strongly coupled together. The BOMD spectrum also predicts that both OH torsional vibrations also have a contribution in the doublet at 532 and 539 cm\(^{-1}\), although these bands are not observed experimentally, or at least not at those frequencies. The coupling is probably due to the strong red shift due to the deuteration allowing these two motions to show up in the same normal mode. We assigned these modes to the experimental features at 378 cm\(^{-1}\) and 449 cm\(^{-1}\) despite their low intensity and the large frequency shift.

The spectral feature measured at 586 cm\(^{-1}\) for SLG is not predicted by the static harmonic levels of theory or by the spectra constructed from BOMD based on BLYP-D3. The spectrum of SLG-BD in Fig. 9a does not show this peak, while it is observed for SLG-FD. This implies that this peak is strongly related to the bound OH torsional local mode. Since this peak is not an overtone of the H-bonded OH torsion, we expect that this peak is a combination band. The most likely candidate is the combination band of the bound OH torsional vibration and the CCCO torsion of the CH\(_2\)OH moiety \((\text{see Fig. } 9a, \text{the dashed line in the VPT2 spectrum})\). Moreover, the VPT2 method provides an alternative assignment for this H-bonded OH torsion feature for SLG-FD \((586\text{ cm}^{-1})\), which predicts this mode at 591 cm\(^{-1}\). Although this assignment seems to match the experimental data well, caution should be applied when assigning large amplitude OH torsional modes based on the anharmonic static VPT2 spectrum. Previously, this VPT2 method has shown to produce unreliable results for large amplitude anharmonic vibrations.

The expected red-shift behavior of the OH torsional modes after deuteration is observed both in the calculated and experimental spectra. These spectra are however equally complicated as SLG since the intensities predicted by all the theoretical methods for the OH torsion are overestimated in comparison with the experimental features in the far-IR, making it hard to assign all modes unambiguously. As discussed previously, the theoretical intensities, which were reasonable throughout the rest of the dataset, are particularly overestimated in the calculations of the SLG isomers.

5. Discussion

A comparison of the results of our computations of the dynamical anharmonic spectra extracted from Born-Oppenheimer DFT-based molecular dynamics simulations \(\text{(using here the BLYP-D3 functional)}\) and the static anharmonic spectra calculated at the B3LYP-D3 VPT2 corrected level to the static harmonic spectra employing both the BLYP-D3 and B3LYP-D3 functionals, allows us to evaluate whether the mismatch between experiment and theory for a specific vibrational mode originates predominantly from functional errors or from the improper treatment of anharmonicity \((\text{which arises either from potential energy surfaces or dipole moments})\). In short, static harmonic spectra are expected to perform poorly on bands where anharmonicities will be most relevant, while anharmonic methods should perform better. We also tested here whether there is added advantage in choosing a hybrid functional over a simpler GGA functional. The B3LYP-D3 electronic representation has the reputation of performing well at the static harmonic level for spectral calculations in the mid-IR regions where mainly localized stretching motions are involved, while we have systematically demonstrated a robustness of the BLYP-D3 functional in spectral predictions in the far-IR once coupled to Born-Oppenheimer dynamical exploration of the potential energy and dipolar surfaces \([9,13]\). In the gas phase molecules investigated here, the anharmonic features are the large amplitude oop OH torsional motions. These OH groups are hydrogen bonded in all present \(^1\) molecules except for PHNL, while the choice of molecules provided a wide range of variations for the intramolecular OH hydrogen bond formed.

As expected, we have shown that the harmonic spectra calculations do not consistently predict the large amplitude OH torsions, regardless of the functional used. While the static anharmonic VPT2 spectra do not improve the predictions for the OH torsional motions, as the frequency shifts and peak intensity are often overestimated, the BOMD spectra consistently provide improved positions and intensities for these vibrational modes. This is due to the dynamical exploration of both the potential energy surface and the dipolar surface, exploring zones that are mandatory to the modulation of the OH torsional motions, and including the dynamical correlation in the motions that are necessary elements for providing reliable spectral motions.

The static harmonic B3LYP-D3 spectra calculated here show that the performance of that level of theory is especially good for the frequency range above 600 cm\(^{-1}\), where the ring modes of the phenyl ring are present, as well as C–O–H bending vibrations. Below 600 cm\(^{-1}\), where mainly large amplitude torsional motions of the ring, its functional groups and more delocalized motions are present, the agreement between harmonic B3LYP-D3 spectra and the experiments becomes less good. Furthermore, if one looks only at the predicted band-positions, the errors in the frequencies predicted by the harmonic B3LYP-D3 spectra are not systematic in the far-IR, so that using one single scaling factor to the whole spectrum would not improve results. This makes far-IR spectra of significantly larger molecules not trivial to assign using only harmonic static theoretical spectra. In these cases, the addition of anharmonic spectra derived from BOMD trajectory simulations is required for achieving full assignments and understanding the structure and dynamics underlying the experimental results. The conclusions reached here for phenol derivative molecules is similar to our previous conclusions for peptides \([9,13]\) in that respect.

Because of the non-simultaneous explorations of PES and dipolar surfaces in the harmonic static methods, the anharmonic VPT2
spectra are usually not able to correct for large anharmonicities in a sufficiently consistent way in the far-IR domain, as the dynamic anharmonic BOMD method can. For modes with large anharmonic constants, such as the OH torsional motions, the frequency shifts introduced by the VPT2 method are often exaggerated, while intensities are overestimated. In that respect, the dynamic anharmonic spectra perform better, showing that the dynamical sampling of the PES and dipoles is important for a good assessment of experimental features in the far-infrared. 

Catechol (CAT) is a good example of the different strengths and weaknesses observed in current computational methods as described above. The vibrational modes that are observed in the experimental spectra are all assignable from any level of theory. As shown in Fig. 3, the BLYP-D3 harmonic spectrum has red-shifts for all modes, while the harmonic B3LYP-D3 spectrum shows both red and blue shifted modes. Both functionals fail in predicting the position of the large amplitude OH torsional motion of the hydrogen bond acceptor when they are applied for harmonic spectrum calculations. The shift is large enough that for any of the other molecules in this dataset it could result in a false assignment of the exact vibrational features due to a reversal of the order of the modes. VPT2 overestimates the shifts for the two OH torsional modes in CAT (Table SI.2), as well as their intensities, showing that this anharmonic correction does not work well for large amplitude motions. The dynamical spectrum on the other hand provides an excellent overall agreement to the experiment, both in terms of frequencies and intensities.

The effect of the strength of the intramolecular hydrogen bond on the spectral match can be observed most directly when comparing SA-R1, which has a strong hydrogen bond, and SA-R2, which has a hydrogen bond of intermediate strength [111]. For both structures, all the spectra produced by static levels of theory and the dynamical anharmonic spectrum are capable of providing consistent results, which enable assignments of all observed features. This shows that the presence of the intramolecular hydrogen bond does not further complicate assignments within this dataset. Flexibility also does not play a significant role in the match between the different theories and experiment within this dataset. Although the presented molecules are not highly flexible, we see that the quality of the match for SLG, the most flexible molecules in the dataset, is similar to that of any of the other molecules. However, the flexibility of the molecules presented here is small when compared to biomolecules, such as peptides [52] and sugars [53].

6. Conclusions

We have recorded conformer and mass selective far-IR spectra of phenol and a set of phenol derivatives using the wide tunability and spectral brightness of the free electron laser FELIX. By comparing these experimental far-IR spectra with theoretical far-IR spectral calculations at different levels of theory, we have seen that within these rather rigid molecules the added flexibility does not have a large influence on the agreement between theories and experiment. The same holds for the strength of a formed intramolecular hydrogen bond (when present). A complicating factor is the coupling between two degenerate local modes such as the large amplitude OH torsion, which is by itself already challenging to predict theoretically. Anharmonicity plays a large role in the dataset, as both intensities and positions are often incorrectly calculated by the static harmonic theories (regardless of the DFT functional used). Also second order perturbation theory is not capable of dealing with the strong anharmonic OH torsional modes present in the molecules. The molecules present in this dataset therefore provide a good test for alternative theories, such as the here discussed anharmonic spectra calculations in the far-IR from Born-Oppenheimer Molecular Dynamics (BOMD). Spectra extracted from anharmonic BOMD trajectory simulations were shown here to reproduce experimental results well (although not without faults) in the far-IR domain and were shown to consistently provide the proper blue and red-shifts of the large amplitude OH torsional motions in the far-IR.

Deuteration is a helpful technique to assist the assignment of the vibrational modes in far-IR spectra, especially when overtones or combination bands are present and theories still struggle at reproducing these effects. These overtones and combination bands can be quite significant, as shown for PHNL and SLG, as well as for other common functional groups such as NH₂ [112]. The static harmonic spectra cannot predict these anharmonic couplings, while the VPT2 anharmonic method was able to, although with very large shifts in positions. The dynamical anharmonic spectra reported here, which in principle include these specific anharmonic couplings, were also unable to calculate these peaks probably due to too weak anharmonic couplings that did not show up properly in the dynamical representation. This might in part be due to the choice of the BLYP-D3 functional, too short trajectories for such weak couplings to correctly be taken into account, or too low temperatures in the trajectories for these low couplings to be fully sampled.

Acknowledgments

We gratefully thank the FELIX staff for their experimental support. All authors acknowledge the Stichting voor Fundamenteel Onderzoek der Materie (FOM) for the support of the FELIX laboratory (project no. N2300 N). The calculations were sponsored by NWO Physical Sciences (EW) by granting access to the supercomputer facilities at SurfSara (project no. MP-264-13) and by HPC resources from GENCI-France through Grant O82484. We also thank Daria Ruth Galimberti for helpful discussions and coding the removal of rotational motions from the dipolar trajectories.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jms.2017.02.004.

References

M.-P. Gaigeot, Theoretical spectroscopy of floppy peptides at room

V. Barone, Anharmonic vibrational properties by a fully automated second-

V. Dixit, R. Yadav, Experimental IR and Raman spectroscopy and DF T methods

S. Ebbinghaus, S.J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D.M. Leitner,


V. Yatsyna, D.J. Bakker, R. Feifel, A.M. Rijs, V. Zhaunerchyk, Aminophenol

G. Varsányi, Normal vibrations of benzene and its derivatives, in: Vibrational

J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, J. Hutter,

M.-P. Gaigeot, R. Spezia, Theoretical methods for vibrational spectroscopy and

A.D. Becke, Density-functional exchange-energy approximation with correct

D. Oepts, A. Van der Meer, P. Van Amersfoort, The free-electron-laser user

P. Gruene, D.M. Rayner, B. Redlich, A.F.G. van der Meer, J.T. Lyon, G. Meijer, A.

B. Kirchner, P.J. di Dio, J. Hutter, Real-world predictions from ab initio molecular

M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman,

S. Kumar, S.K. Singh, C. Calabrese, A. Maris, S. Melandri, A. Das, Structure of

A. Witt, S.D. Ivanov, M. Shiga, H. Forbert, D. Marx, On the applicability of

J. Evans, The vibrational spectra of phenol and phenol-OD, Spectrochim. Acta

W. Caminati, S. Di Bernardo, L. Schäfer, S.Q. Kulp-Newton, K. Siam,


Cunningham, Probing temperature- and solvent-dependent protein dynamics using

S. Ebbinghaus, S.J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D.M. Leitner,

M. Havenith, An extended dynamical hydration shell around proteins, Proc.


K. Fumino, A. Wolf, R. Ludwig, The cation-anion interaction in ionic liquids

emission of disperse red 1 in solution at room temperature, J. Phys. Chem. B


D.F. Plusquellic, K. Siegrist, E.J. Heilweil, O. Esenturk, Applications of terahertz

B. Fischer, M. Walther, P.J. Jepson, Far-infrared vibrational modes of DNA

M. Rozenberg, G. Shoham, I. Reva, R. Fausto, Low-temperature Fourier
transform infrared study of hydrogen bonding in polycrystalline i-

P. Gruene, D.M. Rayner, B. Redlich, A.F.G. van der Meer, J.T. Lyon, G. Meijer, A.
Fielicke, Structures of neutral Au7, Au19, and Au20 clusters in the gas phase,

P.J. Stephens, F.J. Devlin, C.F. Chabalowski, M.J. Frisch, Ab initio calculation of
vibrational absorption and circular dichroism spectra using density functional

S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio
parametrization of density functional dispersion correction [DFT-D) for the 94

M.-P. Gaigeot, M. Bićevski, J. Blono, Fully anharmonic IR and Raman spectra of
medium-size molecular systems: accuracy and interpretation, Phys. Chem.

W. Caminati, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, et al., Gaussian
09, Gaussian Inc., Wallingford, CT, USA, 2009.

M. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman,
M. Scalfani, G. Scalmani, V. Barone, N. Boccia, M. Mengucci, G. Petersson, et al.,
Gaussian 09, Gaussian Inc., Wallingford, CT, USA, 2009.

V. Barone, M. Bićevski, J. Blono, Fully anharmonic IR and Raman spectroscopy of
medium-size molecular systems: accuracy and interpretation, Phys. Chem.

C. Chen, S.-F. Shyu, Conformers and intramolecular hydrogen bonding of the

S. Kumar, S.K. Singh, C. Calabrese, A. Maris, S. Melandri, A. Das, Structure of
saligenin: microwave, UV and IR spectroscopy studies in a supersonic jet
combined with quantum chemistry calculations, Phys. Chem. Chem. Phys. 16

A. Kovács, I. Ivékó, G. Keresztury, G. Pongor, Vibrational analysis of 2-
Nitrophenol. A Joint FT-IR, FT-Raman and scaled quantum mechanical study,

V. Dixit, R. Yadav, Experimental IR and Raman spectroscopy and DFT methods
based material characterization and data analysis of 2-nitrophenol, Biochem.


B. Kirchner, P.J. di Dio, J. Hutter, Real-world predictions from ab initio molecular
interactions, in: B. Kirchner, J. Vrabec (Eds.), Multiscale Molecular Methods in Applied Chemistry, Springer, Berlin, Heidelberg, 2012,

M.-P. Gaigeot, M. Sprik, Ab initio molecular dynamics computation of the
10358.

M. Sulpiri, M. Salanne, M. Sprik, M.-P. Gaigeot, Vibrational sum frequency
generation spectroscopy of the water liquid-vapor interface from density
functional theory-based molecular dynamics simulations, J. Phys. Chem. Lett. 4

H.D. Bist, J.C.D. Brand, D.R. Williams, The vibrational spectrum and torsion of

S. Jaque, W. Du, E.J. Meijer, J. Oomens, A.M. Rijs, Conformational study of Z-
Glur-OH and Z-Arg-OH: dispersion interactions versus conventional hydrogen

P. Çarçabal, R.A. Jockusch, I. Hüning, L.C. Snoek, R.T. Kroemer, B.G. Davis, D.P.
Camberlin, I. Compagnon, J. Oomens, J.P. Simons, Hydrogen bonding and
cooperativity in isolated and hydrated sugars: mannose, galactose, glucose,
Anharmonic, dynamic and functional level effects in far-infrared spectroscopy: phenol derivatives

Daniël J. Bakker†, Qin Ong†, Arghya Dey†, Jérôme Mahé‡, Marie-Pierre Gaigeot‡*, Anouk M. Rijs†*

†Radboud University, Institute for Molecules and Materials, FELIX Laboratory, Toernooiveld
7c, 6525 ED Nijmegen, the Netherlands

‡ LAMBE CNRS UMR8587, Université d’Evry val d’Essonne, Blvd F. Mitterrand, Bât
Maupertuis, 91025 Evry, France

Corresponding Authors
*a.rijs@science.ru.nl, mgaigeot@univ-evry.fr

Contents

Figure SI.1. REMPI spectrum of nitrophenol

Table SI.1. Tabulated experimental parameters for all molecules.

Figure SI.2. Three BOMD spectra of five representative molecules

Figure SI.3. Example of BOMD spectrum including and excluding rotations

Table SI.2. Anharmonic normal modes for which the VPT2 method provides an unreasonably high or low intensity.

Figure SI.4. REMPI spectrum of phenol and deuterated phenol.

Figure SI.5. REMPI spectrum of saligenin and deuterated saligenin.
Figure SI.1. REMPI spectra of nitrophenol, both in an isolated expansion (black) and when co-expanded with phenol (red). The co-expansion gives rise to a signal increase of an order of magnitude.

Table SI.1. Tabulated experimental parameters of all molecules. The frequencies of the UV transitions are referenced to the values found in literature if possible.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$T_{source}$ (°C)</th>
<th>$\lambda_{UV}$(cm$^{-1}$)</th>
<th>Seed gas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenol$^1$</td>
<td>25</td>
<td>36348.5</td>
<td>Ar</td>
</tr>
<tr>
<td>Phenol-D$^1$</td>
<td>25</td>
<td>36346</td>
<td>Ar</td>
</tr>
<tr>
<td>Catechol$^2$</td>
<td>90</td>
<td>35648</td>
<td>He</td>
</tr>
<tr>
<td>Salicylic acid – rotamer 1$^3$</td>
<td>90</td>
<td>30940</td>
<td>He</td>
</tr>
<tr>
<td>Salicylic acid – rotamer 2$^3$</td>
<td>90</td>
<td>32098</td>
<td>He</td>
</tr>
<tr>
<td>Saligenin$^4$</td>
<td>80</td>
<td>35503</td>
<td>Ar</td>
</tr>
<tr>
<td>Saligenin-FD</td>
<td>80</td>
<td>35502.5</td>
<td>Ar</td>
</tr>
<tr>
<td>Saligenin-BD</td>
<td>80</td>
<td>35509.5</td>
<td>Ar</td>
</tr>
<tr>
<td>Nitrophenol</td>
<td>25</td>
<td>35940</td>
<td>He</td>
</tr>
</tbody>
</table>
Figure SI.2. Three BOMD spectra of five representative molecules, for which the trajectories were started from the same optimized geometry, yet differed by the randomized velocity initialization. Differences in between the intensities of individual features in each spectrum indicate that the energy distribution amongst the normal modes is not completely relaxed to the Boltzmann distribution.

Figure SI.3. Far-infrared spectrum of deuterated phenol, as calculated using BOMD. In red, the spectrum resulting from the trajectory including end-over-end rotations is shown. In black, the spectrum resulting from the trajectory of which the end-over-end rotations were removed is shown.
Table SI.2. Anharmonic normal modes for which the VPT2 method provides an unreasonably high or low intensity. Listed are the harmonic and anharmonic frequencies ($\nu_h$ and $\nu_a$) and intensities ($I_h$ and $I_a$), together with the description of the motion of each normal mode, using the normal mode labels taken from Varsanyi$^5$ when possible.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$\nu_h$(cm$^{-1}$)</th>
<th>$\nu_a$(cm$^{-1}$)</th>
<th>$I_h$(km/mol)</th>
<th>$I_a$(km/mol)</th>
<th>Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>139.6</td>
<td>869.5</td>
<td>153.4</td>
<td>6393.1</td>
<td>free OH torsion</td>
</tr>
<tr>
<td></td>
<td>189.4</td>
<td>207.5</td>
<td>8.1</td>
<td>731.7</td>
<td>10b</td>
</tr>
<tr>
<td></td>
<td>404.1</td>
<td>485.9</td>
<td>73.3</td>
<td>303.7</td>
<td>bound OH torsion</td>
</tr>
<tr>
<td>SA-R2</td>
<td>52.3</td>
<td>187.1</td>
<td>1.2</td>
<td>73.5</td>
<td>acidic CCCO-torsion</td>
</tr>
<tr>
<td></td>
<td>423.6</td>
<td>480.8</td>
<td>32.4</td>
<td>506.0</td>
<td>acidic OH torsion coupled to 16a</td>
</tr>
<tr>
<td></td>
<td>488.4</td>
<td>814.9</td>
<td>121.2</td>
<td>1.7</td>
<td>acidic OH torsion &amp; bound OH torsion, symmetric</td>
</tr>
<tr>
<td></td>
<td>539.8</td>
<td>566.9</td>
<td>47.7</td>
<td>692.3</td>
<td>bound OH torsion coupled to 16b</td>
</tr>
<tr>
<td></td>
<td>583.4</td>
<td>652.5</td>
<td>18.9</td>
<td>335.1</td>
<td>acidic OH torsion &amp; bound OH torsion, asymmetric</td>
</tr>
<tr>
<td>SLG</td>
<td>424.7</td>
<td>395.5</td>
<td>72.0</td>
<td>219.5</td>
<td>free OH torsion</td>
</tr>
<tr>
<td></td>
<td>1187.5</td>
<td>1157.5</td>
<td>5.1</td>
<td>160.6</td>
<td>free COH-bending</td>
</tr>
<tr>
<td>SLG-FD</td>
<td>1264.0</td>
<td>1235.0</td>
<td>36.6</td>
<td>438.5</td>
<td>bound COH-bending coupled to ring vibration</td>
</tr>
<tr>
<td></td>
<td>1265.2</td>
<td>1234.9</td>
<td>117.4</td>
<td>451.2</td>
<td>bound COH-bending coupled to bound CO stretching</td>
</tr>
<tr>
<td>NP</td>
<td>689.5</td>
<td>748.9</td>
<td>2.8</td>
<td>195.4</td>
<td>4 coupled to NO$_2$ wagging</td>
</tr>
<tr>
<td></td>
<td>783.3</td>
<td>784.1</td>
<td>24.3</td>
<td>0.0</td>
<td>11 coupled to NO$_2$ wagging and OH torsion</td>
</tr>
</tbody>
</table>
Figure SI.4. REMPI spectra measured in the mass channels of (a) PHNL and (b) PHNL + 1 amu. Deuterated PHNL and PHNL with one C13 atom incorporated give rise to clearly separated peaks.
Figure SI. 5. REMPI spectra measured in the (a) SLG and (b) SLG+1 amu mass channels. Every peak from the SLG spectrum matches to two separate peaks in the SLG+1 amu spectrum, showing the two different deuterated species. The peak indicated with the red arrow is assigned to SLG-FD, the peak indicated by the blue arrow is SLG-BD. The less intense, broadened features that are observed as shoulders of the sharp features are tentatively assigned to the $^{13}$C isotopologue of SLG.

Literature


Chapter 8

Fingerprints of Inter- and Intramolecular Hydrogen Bonding in Saligenin-Water Clusters Revealed by Mid- and Far-Infrared Spectroscopy.

The micro-solvation of bare molecules that are stabilised by intramolecular hydrogen bonds can be complicated due to the competition between the existing intramolecular hydrogen bonds and the formation of intermolecular hydrogen bonds with the interacting solvent molecules. Within the field of gas-phase infrared spectroscopy, many experiments are performed to study the intrinsic properties of isolated and microsolvated complexes, and the investigations presented in this chapter are part of this field of research.

However, many of the vibrations that directly involve hydrogen bonds deformations are positioned in the far infrared part of the infrared spectrum (<800 cm\(^{-1}\)), most notably the hydrogen bond stretching vibrations. This is exactly the spectral domain of probing in our work.

In this context, bare saligenin and saligenin-(H\(_2\)O)\(_n\) (n=1-3) clusters have been chosen for investigation, especially because a large diversity of hydrogen bonds is expected. The systems have been measured using infrared (IR) ultraviolet (UV) ion-dip spectroscopy in the far- and mid-IR regions of the spectrum, and different theoretical methods have been applied to obtain vibrational spectra. Both harmonic and anharmonic quantum chemical calculations were applied to assign cluster geometries to the measured spectra, and to assign vibrational modes to the spectral features measured. Two anharmonic methods were applied: one method is the coupled local modes developed by Prof Sibert’s group applied here in the mid-infrared spectral domain, the other is our preferred dynamical DFT-MD method, used in the far infrared range. The structures of saligenin-water clusters were assigned based predominantly on the comparison of the mid-infrared signatures (experiments and calculations) while the far-infrared spectra proved crucial for disentangling between different structures when the mid-infrared spectra were not sufficiently diagnosis, showing again that the far-infrared is sensitive to subtle structural changes\(^{105}\). All results are presented in the attached paper: "Fingerprints of Inter-
and Intramolecular Hydrogen Bonding in Saligenin-Water Clusters Revealed by Mid- and Far-Infrared Spectroscopy\textsuperscript{111}.

The paper: "Anharmonic, dynamic and functional level effects in far-infrared spectroscopy: phenol derivatives" presented in chapter 7 has shown a good agreement between experimental and BLYP-D3-MD spectra for the phenol derivatives series. In particular, the OH wagging modes, both for free OH and OH engaged into an hydrogen bond have been correctly reproduced by the BLYP-D3-MD method with deviations < 20 cm\textsuperscript{-1} with respect to the experiment. Overtones and combination bands were however missing from our DFT-MD spectra, but no missing peak has been reported for saligenin-water clusters of interest here. The BLYP-D3-MD theoretical method thus provides good results for the investigated clusters.

In the work reported in this chapter, another anharmonic method has been employed specifically for the 3000-4000 cm\textsuperscript{-1} spectral domain (on top of the harmonic signatures also calculated in this domain). It is the local modes model (LM) developed by the group of Prof Edwin L. Sibert in the US. This method has already been successfully applied to reproduce the mid infrared spectra of similar systems\textsuperscript{193–196}. The LM spectral signatures in the range 3000-4000 cm\textsuperscript{-1} offer most of the time an improvement in comparison with the harmonic ones, when compared to experiment.

The assigned structure of saligenin-(H\textsubscript{2}O)	extsubscript{1}, \textit{i.e.} the lowest electronic energy conformer at the MP2/6-311++G(2d,p) level, consists of the native saligenin structure (as presented in chapters 2 and 7) in which the water molecule is inserted inbetween the CH\textsubscript{2}-OH group and the \textpi-cloud of benzene. One OH group of the water molecule is hence directed towards the benzene \textpi-cloud. See figure 8.1 (SLG-1w). This is similar to the structure of the benzyl alcohol-(H\textsubscript{2}O) complex\textsuperscript{197}.

In saligenin-(H\textsubscript{2}O)	extsubscript{2}, the water molecules now form a dimer chain spanning the gap from the methanolic OH group to the benzene ring of the saligenin molecule in its native conformation. See figure 8.1 (SLG-2w). This structure, which has the lowest electronic energy at the MP2/6-311++G(2d,p) level, is therefore also similar to the structure of the benzyl alcohol-(H\textsubscript{2}O)\textsubscript{2} complex\textsuperscript{197}.

Two different isomers co exist for saligenin-(H\textsubscript{2}O)\textsubscript{3} in our experimental conditions. For the first conformer (SLG-3w-a cluster, figure 8.1), the assigned structure has a 10-membered hydrogen bonded cycle through OH···O linkages similar to the phenol-(H\textsubscript{2}O)\textsubscript{4} structure\textsuperscript{114} and isolated (H\textsubscript{2}O)\textsubscript{5} cluster\textsuperscript{198}. This specific isomer is the lowest electronic energy structure amongst the SLG-3w clusters that were considered in our conformational search, at the MP2/6-311++G(2d, p) level. The water chain is located above the saliginin ring. For the second conformer (SLG-3w-b, figure 8.1), the assigned geometry includes a hydrogen bonded 10-membered cycle with the direction of hydrogen bonding reversed with respect to the SLG-3w-a cluster. The hydrogen bonded cycle is distorted, causing this geometry to be energetically less favorable than SLG-3w-a (by 48 kJ/mol at the MP2/6-311++G(2d, p) level of theory).
Note that SLG-3w-b apart, the geometry of saligenin(H₂O)_(n−1) is always a good guess for the structure of saligenin(H₂O)_n.

Figure 8.1: Assigned structures of the measured saligenin-(H₂O)₁⁻₃ species, optimized at the MP2/6-311+G(2d,p) level of theory. This figure has been copied from the following paper and the colors are used to identify each OH function in the context of this work.

Three hydrogen bond deforming modes, namely OH stretching, OH torsion and hydrogen bond stretching were specifically studied in our work and their frequencies plotted as a function of the hydrogen bond strength, represented by either the OH covalent bond-length or the hydrogen bond-length. OH stretching and wagging are the main contributors of the local modes of the same name and therefore their frequencies will be used. On the other hand, the hydrogen bond stretching is involved in several delocalised normal modes, therefore the mode with the biggest participation of this internal coordinate has been chosen for analysis.

The OH torsion (wagging) vibration displays the largest frequency shift when the OH is hydrogen bonded, with signatures found over a range of 550 cm⁻¹ (and more specifically inbetween 200 and 750 cm⁻¹), followed by the covalent OH stretching vibrations (over a range of 500 cm⁻¹, inbetween 3220 and 3720 cm⁻¹) and finally the hydrogen bond stretching frequency (over a range of 300 cm⁻¹, inbetween 50 and 350 cm⁻¹).

We found that the shifts in the frequencies of these three modes correlate linearly with the OH covalent bond-length. But the frequency shifts of these hydrogen bond deforming modes behave non-linearly as a function of the hydrogen bond length, asymptotically approaching the frequency expected for the non hydrogen bonded modes (i.e. for long hydrogen bonds). The nonlinear behaviour was quantified using exponential functions. A more detailed discussion on hydrogen bond signatures in the far infrared/THz domain for these (floppy) systems and other (more rigid) systems can be found in chapter 9.

Despite the small amount of water molecules (3 maximum here), one can also observe
collective motions of the water molecules in the far infrared/THz that one can compare with bulk water collective vibrational modes (even if it has not been discussed in this work) as observed for example by the group of Martina Havenith in liquid water in the same spectral range as ours\textsuperscript{199}. These motions involve in particular the hydrogen bond stretching motions of all the OH functions coupled together.
Fingerprints of inter- and intramolecular hydrogen bonding in saligenin–water clusters revealed by mid- and far-infrared spectroscopy†

Daniël J. Bakker,⁠a Arghya Dey,⁠a Daniel P. Tabor,⁠a,b Qin Ong,⁠a Jérôme Mahé,⁠c Marie-Pierre Gaigeot,⁠d,e Edwin L. Sibert III* and Anouk M. Rijss*²

Saligenin (2-(hydroxymethyl)phenol) exhibits both strong and weak intramolecular electrostatic interactions. The bonds that result from these interactions compete with intermolecular hydrogen bonds once saligenin binds to one or more water molecules. Infrared (IR) ultraviolet (UV) ion-dip spectroscopy was used to study isolated saligenin–(H₂O)_n clusters (n = 1–3) in the far- and mid-IR regions of the spectrum. Both harmonic and anharmonic (coupled local modes and Born–Oppenheimer molecular dynamics) quantum chemical calculations were applied to assign cluster geometries to the measured spectra, and to assign vibrational modes to all spectral features measured for each cluster. The hydrated clusters with n = 1 and 2 have geometries that are quite similar to benzyl alcohol–water clusters, whereas the larger clusters with n = 3 show structures equivalent to the isolated water pentamer. Systematic shifts in the frequencies of three hydrogen bond (H-bond) deforming modes, namely OH stretching, OH torsion and H-bond stretching, were studied as a function of the hydrogen bond strength represented by either the OH bond length or the H-bond length. The shifts of the frequencies of these three modes correlate linearly to the OH length, despite both intra- and intermolecular H-bonds being included in this analysis. The OH torsion vibration displays the largest frequency shift when H-bonded, followed by the OH stretching vibrations and finally the H-bond stretching frequency. The frequency shifts of these H-bond deforming modes behave non-linearly as a function of the H-bond length, asymptotically approaching the frequency expected for the non H-bonded modes. The nonlinear behavior was quantified using exponential functions.

1 Introduction

The secondary structure of proteins is directed by the formation of intramolecular hydrogen bonds (H-bonds) leading to diverse morphologies such as turns,⁠a helices⁠a,b,d and sheets.⁠a,b,d The microsolvation behavior of molecules that are stabilized by intramolecular H-bonds is complicated due to the competition between the existing intramolecular H-bonds and the formation of intermolecular H-bonds with the interacting solvent molecules.⁠a,b,d Therefore, understanding the microsolvation of intramolecular H-bonded systems is crucial as it plays a vital role in therapeutics where it helps in maintaining lipophilicity around the cell membranes and enhancing cell permeability.⁠a,c,d 

There is a large body of literature describing the interplay between intramolecular and intermolecular hydrogen bonding. A wide range of methods have been employed such as NMR,⁠a,b X-ray crystallography⁠a,c and infrared (IR) spectroscopy.⁠b,c,⁠d Within the field of gas-phase IR spectroscopy, many experiments are performed to study the intrinsic properties of isolated, microsolvated complexes.⁠b,c,⁠d To characterize the full three dimensional geometry of neutral complexes of limited size, mid-IR spectroscopy covering the CH, NH and OH stretching vibrations (2500–4000 cm⁻¹) is routinely applied combined with ab initio quantum chemical calculations. However, many of the vibrations that directly involve H-bond deformations are positioned in the far-IR part of the spectrum (<800 cm⁻¹), most notably the H-bond stretching vibration itself. To date, gas-phase far-IR spectroscopy is not as commonly applied as mid-IR techniques, due to limited IR intensities of vibrational transitions, difficulties for quantum chemical calculations to
accurately reproduce these far-IR spectral features and the rarity of intense far-IR light sources. An intense light source is not required for far-IR Raman spectroscopy,\textsuperscript{23,24} which provides complementary information to absorption spectroscopy, but is not straightforward for gas phase experiments. Despite these limitations, it has been shown that gas phase far-IR spectroscopy is a potent technique, displaying vibrational modes that are sensitive to subtle structural changes\textsuperscript{25} and systematic frequency shifts that are sensitive to the formation of intramolecular H-bonds.\textsuperscript{23,26–33} Moreover, sophisticated quantum chemical calculations can be applied to accurately reproduce the far-IR spectra of both semi-rigid\textsuperscript{34,35} and floppy peptide molecules.\textsuperscript{25,36,37} We continue building on this work by studying the interplay of intra- and intermolecular H-bonds in saligenin (Fig. 1) – a benzene derivative – clustered with water molecules, using both mid- and far-IR spectroscopy.

Benzene and benzene derivatives (see Fig. 1 for their structures) and their microsolvated clusters have been studied elaborately as model systems for solvation chemistry by gas phase mid-IR action spectroscopy.\textsuperscript{39–44} In benzene-water clusters with up to five water molecules, the water molecules are arranged in the same geometries as they would be in isolation,\textsuperscript{45,46} although with significantly perturbed H-bond lengths. The benzene molecule is bound to one of the unoccupied OH groups of a cluster through an OH····π bond.\textsuperscript{47,48} Similar H-bond networks are observed for phenol-water clusters, although one of the water molecules in the water cluster structure is replaced by the phenolic OH group.\textsuperscript{42,43} This behavior is not observed for benzyl alcohol, which shows a weak OH····π interaction in isolation between the methanolic CH\textsubscript{2}OH group and the benzene π-cloud.\textsuperscript{50} The benzyl alcohol–water complexes form a network of OH····O linkages, where the water molecules are consistently engaged in an additional OH····π bond with the aromatic ring. The weak intramolecular OH····π bond is broken immediately upon the bonding of the first water molecule.\textsuperscript{51} The doubly substituted benzene derivative catechol, which has two hydroxyl groups, the second of which is in the ortho-position, exhibits a planar conformation with an OH····O intramolecular H-bond between the two hydroxyl groups.\textsuperscript{51,52} Upon the addition of water, the intramolecular H-bond is weakened but not broken.\textsuperscript{53} Seeking to further our understanding of the competing forces that influence the formation of intra- and intermolecular H-bonds, the molecule saligenin (SLG) is studied here, which contains the flexible CH\textsubscript{2}OH substituent of benzyl-alcohol and a phenolic OH group in the ortho-position.

SLG is a naturally occurring drug with antipyretic\textsuperscript{54} and anesthetic properties.\textsuperscript{55} Microwave and IR spectroscopic studies show that the ground state geometry of saligenin contains an intramolecular OH····O bond donated by the phenolic OH group to the oxygen of the CH\textsubscript{2}OH group and a very weak OH····π interactions between the methanolic OH group and the aromatic π-cloud of the benzene ring.\textsuperscript{56} The water molecules can either bind to the methanolic OH to produce water complexes similar to that of benzyl alcohol, breaking the OH····π bond or they can interact with the phenolic OH to form an intermolecular H-bond, breaking the intramolecular H-bond of SLG. H-bonded SLG–(H\textsubscript{2}O\textsubscript{n}) clusters (n = 1–3) are investigated by IR–UV double resonance spectroscopy using both the OH stretching region of the mid-IR (3100 to 3800 cm\textsuperscript{-1}) and the far-IR region (<800 cm\textsuperscript{-1}). The structures of the H-bonded clusters are found by comparing the experimental spectra to calculated spectra based on equilibrium structures obtained from quantum chemical calculations. The far-IR spectra of the H-bonded complexes provide crucial information for the structural assignments. Moreover, the far-IR has been highly instrumental in probing the H-bond stretching modes and other vibrational modes that are sensitive to H-bond formation.\textsuperscript{25}

2 Experimental and computational details

All the experiments were performed in a pulsed molecular beam set-up coupled with a time-of-flight mass spectrometer, explained in detail elsewhere.\textsuperscript{57} Here, a brief description of the experimental procedures is presented. SLG was obtained from Sigma Aldrich and used without further purification. The sample was heated to 85 °C and supersonically expanded into the source chamber using a pulsed valve (Series 9, Iota One; General Valve Corporation). The nozzle temperature was maintained at 90 °C to prevent condensation and subsequent clogging of the nozzle. Helium was used as a buffer gas with a backing pressure of 3 bar. To produce saligenin–water complexes, the buffer gas was bubbled through a distilled water reservoir maintained at room temperature. The molecular beam was skimmed approximately 5 cm downstream from the nozzle and entered the differentially pumped ionization chamber where it interacted with the UV and IR laser beams. The molecules and the clusters were electronically excited using the (1 + 1)-REMPI scheme\textsuperscript{58} using UV radiation generated by a frequency doubled dye laser (NarrowScan, Radiant Dyes) operating with Coumarin 153 dye. The dye laser was pumped with the third harmonic output of a pulsed Nd:YAG laser (Quanta Ray Lab series, Spectra Physics).

The IR spectra of the molecules and clusters were recorded using IR–UV ion dip spectroscopy.\textsuperscript{59,60} Both the IR and UV beams were spatially and temporally overlapped and the IR beam interacted with the molecular beam prior to the UV beam.
The wavelength of the IR beam was scanned while monitoring the ion signal obtained by fixing the frequency of the UV laser on a resonant vibronic transition of the isomer, resulting in dips in the signal when the resonant IR light depletes the vibrational ground state. For the far-infrared region (200 to 800 cm$^{-1}$), the IR radiation was produced by the free electron lasers (FEL) at the FELIX Laboratory located at the Radboud University. The FEL pulse energy is typically 80 mJ in the center of the measured range and decreases to 25 mJ at the edge of the measured range. The IR pulses of the FEL have a pulse duration of 6 to 10 μs in the used range. The end of the FEL-pulses coincided with the arrival of the UV light. The mid-IR laser beam was produced using an optical parametric oscillator (OPO, Laser Vision) fitted with a KTA crystal pumped with the fundamental of a pulsed Nd:YAG laser (SpitLight 1200, Innolas GmbH). The IR pulses of the OPO have a pulse length of a few nanoseconds and typically have a pulse energy of 10–12 mJ over the whole measured range. The OPO pulses were timed 100 ns prior to the UV laser pulse. All IR spectra were corrected for the wavelength-dependent photon flux of the IR light sources. In order to compensate for long term fluctuations in sample density and UV power, alternating IR-on and IR-off ion signals were measured by operating the IR laser at a repetition rate of 10 Hz and the UV laser at 20 Hz.

A detailed conformational search was performed using simulated annealing molecular dynamics calculations. For this, the clusters are heated to 250 K and consecutively cooled down to 0 K, in a cycle which is repeated 500 times to find the minima present in the potential energy surface. The maximum temperature of 250 K is chosen to enable isomerization, but to avoid fragmentation of the clusters. Different input geometries were used to initiate the simulated annealing calculations, based on the assigned structures of similar microsolvated clusters such as benzyl alcohol water, but present for all SLG–water clusters). The harmonic calculations rely on a classical, Newtonian treatment of the electrons, while the electrons are treated quantum mechanically. The forces on the nuclei and the energy of the system were evaluated at each time step by application of the DFT formalism, using the BLYP functional and a D3 dispersion correction and a mixed basis set consisting of a plane-wave basis set (kinetic energy cut-off at 450 Ry) combined with a Gaussian one of the aug-TZV2P type with a GTH type pseudopotential. A cubic box with 16 Å sides – based on energy optimizations as a function of the box size using the largest system – was employed to ensure that the systems did not interact with their duplicates beyond the simulation volume, due to the periodic boundary conditions. A single simulation was initialized for each system using the optimized geometry and a randomized velocity distribution based on the usual decomposition into Cartesian coordinates. The internal coordinate contributions of several normal modes of interest were studied in detail for an unambiguous quantified assignment.

The anharmonic calculations in the mid-IR were performed using a local mode (LM) model. This method has been successfully applied to reproduce the mid-IR spectra of similar systems. The procedure started with a harmonic normal mode calculation on the MP2/6-311+G(2d,p) level of theory, performed using Gaussian-09. Subsequently, the normal modes were projected on a basis of local OH stretching and HOH bending modes, creating a LM model Hamiltonian. The OH stretching local modes were all scaled using a two-parameter scaling function that is based on fitting the harmonic frequencies of the benzene–(H$_2$O)$_n$ cluster to the OH stretching experimental fundamentals of the assigned book isomer and

$$\nu_{\text{scaled}} = 0.8698\nu_{\text{local}} + 328.4$$

in units of (see Fig. SI.1, ESI†). The HOH bending overtone frequencies were determined by scaling the LM fundamental frequencies of the bending vibration with a single scaling parameter (0.99), and consecutively subtracting a constant 40 cm$^{-1}$ from twice the fundamental. The dipole derivatives were left unscaled and transformed into this local mode basis. The created model Hamiltonian initially consisted of two separate blocks belonging to the OH stretching and HOH bending overtone vibrations. The Fermi coupling matrix elements that describe the coupling between the bending overtones and OH stretching fundamentals were set to 45.3 cm$^{-1}$, based on results obtained for the benzene–(H$_2$O)$_n$ cluster. The complete LM model Hamiltonians are listed in the ESI† Tables SL2, SL.4, SL.6, SL.8 and SL.10. These coupling effects mainly the modes in the 3100 to 3300 cm$^{-1}$ region, so that variance in the Fermi couplings for higher OH stretching frequencies minimally changes the final spectra. The adjusted model Hamiltonian was diagonalized to find the coupled local mode frequencies, which provides a model Hamiltonian spectrum when combined with the transformed dipoles. A strong advantage of the LM model is that the anharmonic frequencies are not explicitly calculated, so that the computational cost is not significantly increased beyond that of the harmonic MP2 frequency calculation.

The anharmonic spectra reported for the far-IR were calculated using Born–Oppenheimer Molecular-Dynamics (BOMD), which relies on a classical, Newtonian trajectory calculation of the nuclei, while the electrons are treated quantum mechanically. The forces on the nuclei and the energy of the system were evaluated at each time step by application of the DFT formalism, using the BLYP functional and a D3 dispersion correction and a mixed basis set consisting of a plane-wave basis set (kinetic energy cut-off at 450 Ry) combined with a Gaussian one of the aug-TZV2P type with a GTH type pseudopotential. A cubic box with 16 Å sides – based on energy optimizations as a function of the box size using the largest system – was employed to ensure that the systems did not interact with their duplicates beyond the simulation volume, due to the periodic boundary conditions. A single simulation was initialized for each system using the optimized geometry and a randomized velocity distribution based on the
Boltzmann distribution for the simulation temperature of 50 K. After a thermalization of 4 ps, which is required to redistribute energy among all vibrational modes, a 20 ps trajectory was calculated with a time step of 0.4 fs.

End-over-end rotations of the cluster were removed from the trajectory a posteriori after which the time-dependent electric dipole vector was extracted from the trajectory. The IR spectrum was calculated by Fourier transforming the dipole time-correlation function. It has been shown previously that the energy redistribution between vibrational modes is slow in isolated systems and especially for smaller systems as are studied here. Therefore, 4 ps of thermalisation is not enough to reach full energy equilibration. The relative intensities of spectra based on a 20 ps single trajectory can therefore deviate from the intensities that would be found for a much longer trajectory or for the average of multiple initializations. Because of this, significant care should be taken when interpreting the relative intensities of the BOMD simulations, while the positions of the spectral features are reliable. To assign the signatures in the far-IR spectrum to molecular motions, Fourier transforms were applied to intramolecular coordinate time correlation functions, which provide signatures of the internal coordinates. This is therefore an effective tool to assign local modes involving an intra- or intermolecular coordinate. The BOMD calculations were performed using the CP2K software package.

The structural assignments were primarily based on the comparison of the experimental mid-IR spectra with the LM model MP2/6-311++G(2d,p) calculations, with support provided by the far-IR spectra. The details of the assignments are discussed in the ESL, as well as the atomic coordinates of the assigned geometries in Tables SI.1, SL.3, SL.5, SL.7 and SL.9. The reported mid-IR features were broadened using Gaussian functions with a FWHM of 2 cm$^{-1}$. The harmonic far-IR spectrum was convoluted with Gaussian functions with a FWHM of 0.5% of the IR wavenumber which is equivalent to the bandwidth of the FELIX free electron laser. The spectral features in the BOMD spectrum are intrinsically broadened.

3 Infrared spectra and assigned structures

The electronic excitation spectra recorded at the m/z channel of SLG, SLG–1w and SLG–2w are presented in Fig. SI.3 of the ESL. Despite being detected in the mass channels belonging to SLG–(H$_2$O)$_2$ clusters, the measured spectra were assigned to a single SLG–1w and SLG–2w cluster and two SLG–3w clusters. Five of the reported UV transitions were used to record the IR spectra using IR–UV ion-dip spectroscopy: $\nu_{\text{UV}} = 35,492$ cm$^{-1}$ for SLG, $\nu_{\text{UV}} = 35,676$ cm$^{-1}$ for SLG–1w, $\nu_{\text{UV}} = 35,650$ cm$^{-1}$ for SLG–2w, $\nu_{\text{UV}} = 35,977$ cm$^{-1}$ for SLG–3w-a and $\nu_{\text{UV}} = 35,855$ cm$^{-1}$ for SLG–3w-b (indicated by arrows in Fig. SL.3, ESL†). The IR spectra are recorded in the OH stretching ($3100$–$3800$ cm$^{-1}$) as well as in the far-IR region ($<800$ cm$^{-1}$), which are presented in Fig. 2. The electronic transitions in the (1 + 1)-REMPI spectrum were differentiated for the various isomers by IR–UV hole burning spectroscopy where the IR laser was fixed on a resonant vibrational transition and the wavelength of the UV laser was scanned. The hole burning spectra, presented in Fig. SL.3 (ESI†), were used to assign a specific cluster to each UV feature.

In the mid-IR spectrum of SLG two peaks are observed resulting from the two OH groups of the molecule. The peaks at 3493 and 3637 cm$^{-1}$ originate from the phenolic OH and methanolic OH stretching vibrations respectively. The position of the peaks are in agreement with the earlier reported spectrum. In the far-IR, many peaks can be discerned with low-intensity features visible due to the low noise-level. As water molecules bond to SLG, two extra OH stretching features are expected to manifest in the mid-IR spectrum per water molecule, leading to 2 + 2n features, where n is the number of water molecules (not taking into account anharmonic couplings and crossover between clusters).

In the mid-IR spectrum of SLG–1w, four intense absorptions and several low-intensity depletion- and gain features can be seen. The low-intensity ion-gain and ion-dip features result from absorptions of SLG–2w and SLG–3w-a visible in the SLG–1w IR spectrum. The UV excitation for SLG–1w is also partly resonant with the SLG–2w cluster, resulting in additional peaks of SLG–2w in the IR spectrum SLG–1w, which is indicated by green arrows in Fig. 2. Additionally, upon absorption of resonant IR light the SLG–3w-a isomers partly fragment, leading to an increase of the SLG–1w signal, which manifests as negative signal (or gain signal) in the IR spectrum of SLG–1w. This effect, indicated by the purple arrows in Fig. 2, also causes the splitting of the blue-most feature present at 3720 cm$^{-1}$ in the SLG–1w spectrum. In contrast, the far-IR spectrum shows negligible crossover from the SLG–2w and SLG–3w-a spectrum.

The SLG–2w spectrum was recorded in the SLG–1w mass channel (see Fig. SL.3, ESI†), because of fragmentation of these clusters upon irradiation with the ionizing UV radiation, as was observed previously for benzene water clusters. The structure was assigned to a two water cluster based on the six spectral features visible in the mid-IR spectrum, of which two are positioned at the blue side of the spectrum ($>3680$ cm$^{-1}$), indicating that there are two free OH groups. The far-IR spectrum shows more absorptions than the SLG–1w far-IR spectrum, with a comparable noise-level. Two different IR spectra were found for the two SLG–3w isomers, namely the SLG–3w-a and SLG–3w-b. Both show eight features in the mid-IR, which caused them to be assigned as three-water clusters, although both were measured by probing the SLG–(H$_2$O)$_2$ mass channel. The SLG–(H$_2$O)$_2$ mass channel resulted exclusively in IR spectra of even higher order SLG–water clusters. The spectrum of SLG–3w-a shows a signal to noise ratio that is comparable to that of the smaller clusters. The SLG–3w-b spectrum shows both an increased noise level, and also multiple gain features that point to fragmentation of clusters of higher mass. The far-IR spectra belonging to the SLG–3w clusters show many features, some significantly broadened resulting in several congested regions, especially around 450 cm$^{-1}$.

Saligenin

The structure, the mid-IR spectrum and the far-IR spectrum of bare SLG, presented in the top row of Fig. 3, were previously
fully assigned. In isolated SLG, the phenolic OH moiety (colored in red) acts as H-donor for an H-bond to the oxygen atom of the methanolic OH group (in orange), while the methanolic OH group is engaged in a very weak OH···π interaction with the π-cloud of the benzene ring. The vibrational modes corresponding to the colored OH groups in Fig. 3 are highlighted in the same color as the vibrating moieties, with letters labeling the different assigned local modes. In the mid-IR region, the letters B, π and F indicate an H-bonded, OH···π bonded and a free OH stretching vibration respectively. In the far-IR, the labels TB, Tπ and TF represent the torsional vibrations corresponding to H-bonded, OH···π bonded and free OH groups, while HB and Hπ are H-bond and OH···π bond stretching vibrations (colored for the H-donor OH moiety).

The shifts of the two mid-IR features of SLG to 3493 and 3639 cm\(^{-1}\) with respect to the free OH stretching vibration of phenol at 3657 cm\(^{-1}\),\(^{42}\) are ascribed to the two electrostatic interactions (dashed lines in Fig. 3) in the molecule, where the strongest red-shift belongs to the H-bonded OH group. In the mid-IR, the frequencies of the features are well reproduced by the LM model MP2 calculations, while the harmonic M06-2X level of theory slightly underestimates the red-shifts of the H-bonded OH groups. Both calculations reproduce the relative intensities of these features well. The far-IR feature with the highest intensity at 756 cm\(^{-1}\), which is shared by all clusters, is assigned to collective CH out-of-plane torsion of the benzene ring (which is coupled to a bound OH torsion vibration for several of the larger clusters). This feature, highlighted by the dashed grey line for all SLG–water clusters, hardly shifts for the larger clusters. Both calculations reproduce this feature, albeit red-shifted by BOMD and blue-shifted by M06-2X. The other notable far-IR absorptions are observed in the experiment at 690, 384, 229 and 129 cm\(^{-1}\), which are assigned to phenolic OH out-of-plane torsion, methanolic OH torsion, phenolic OH donated H-bond stretching and OH···π bond stretching, respectively. Based on previous work,\(^{26}\) these vibrational modes (excepting the collective pure CH out-of-plane vibration) are expected to be diagnostic for the characteristics of the H-bonds within the clusters.

**Saligenin–(H\(_2\)O)**

The IR spectrum of the SLG–1w cluster is presented in the middle panel of Fig. 3 together with the calculated IR spectra of the assigned structure. The assigned structure, the lowest energy isomer found in our conformational search, consists of the native SLG structure, where the water molecule is interjected in between the methanolic OH group directed towards the π-cloud of the benzene ring. The SLG–1w structure is therefore similar to the structure of the benzyl alcohol–(H\(_2\)O) complex.\(^{41}\) The IR spectrum in the OH stretching region shows four distinct transitions, as expected for a SLG–1w isomer. The transition at 3724 cm\(^{-1}\) is assigned to free OH stretching of the water molecule. This free OH stretch vibration is blue-shifted with respect to the free OH stretching normal mode in phenol at 3657 cm\(^{-1}\), and with respect to the free OH stretch vibration in for example methanol which is positioned at 3682 cm\(^{-1}\) (measured in vapour phase).\(^{27}\) This blue-shift is caused by the oxygen atom of the vibrating OH group acting as an H-acceptor. The feature at 3610 cm\(^{-1}\) belongs to the OH···π-bonded group and the two peaks at 3512 and 3397 cm\(^{-1}\) to two H-bonded OH groups. The red-shift of the phenolic OH stretching vibration with respect to the same vibration in bare SLG is a consequence of a stronger H-bond. The increased strength is caused by the additional water

---

Fig. 2  Infrared spectra of SLG water clusters, from bare SLG up to SLG–3w in both the far-IR and the mid-IR part of the spectrum. The mid-IR spectra are expected to show 2 + 2\(n\) absorptions, where \(n\) is the number of water molecules in the cluster, which is the total number of OH groups in the cluster. The mid-IR spectrum of SLG–1w shows several additional low-intensity features, where additional depletions (positive signal) are related to the co-ionization of SLG–2w clusters at the same UV wavelength and gain (negative signal) is related to fragmentation of SLG 3w-a clusters into the same mass channel, as indicated by the green and purple arrows respectively.
molecule in between the methanolic OH group and the benzene ring, which increases the degree of polarization of the methanolic OH group. The phenolic OH stretching is also significantly broadened with respect to the other features present in the spectrum. This is a well-known effect exhibited by H-bonded stretching vibrations. For SLG–1w the LM model MP2 spectrum reproduces the spectrum quite well, while the harmonic M06-2X spectrum again underestimates the shift for the more strongly bound oscillators.

The far-IR spectrum of SLG–1w shows an intense peak at 756 cm\(^{-1}\) which belongs to collective CH out-of-plane torsion, observed in the same position for bare saligenin. Moreover, high-intensity peaks are observed at 577 and 705 cm\(^{-1}\). The peak at 705 cm\(^{-1}\) is assigned to out-of-plane wagging of the phenolic OH group and the intense, broad peak at 577 cm\(^{-1}\) to the out-of-plane torsion of the methanolic OH. The broadening of this feature is not reproduced by BOMD, while the intensity is underestimated by BOMD but overestimated by M06-2X. The features on the red side of the far-IR spectrum belong to free or weakly bound OH torsion vibrations and HB stretching vibrations. The range chosen for the experimental spectrum covers the OH torsion features and the stretching vibrations of strong H-bonds, but the weaker H-bonds are shifted beyond the measured range. Therefore, these vibrations can only be pinpointed in the calculated spectra. The BOMD and M06-2X calculations predict the same general shape of the far-IR spectrum, but the BOMD simulation shows a better match to the experimental intensities.

**Saligenin–(H\(_2\)O)\(_2\)**

In the measured SLG–2w isomer, the water molecules form a chain spanning the gap from the methanolic OH group to the benzene ring of the SLG molecule in its native conformation. In this geometry, which has the lowest stabilization energy, two free OH groups are present belonging to the two water molecules, as well as three strong H-bonds and a weaker OH...π-interaction. The SLG–2w structure is therefore also similar to the structure of the benzyl alcohol–(H\(_2\)O)\(_2\) complex. These characteristics are reflected in the mid-IR spectrum, where we see a doublet of features at 3722 and 3698 cm\(^{-1}\), corresponding to the free OH moieties. A single peak is red-shifted to 3581 cm\(^{-1}\) with respect to these free OH normal modes and three peaks are more heavily shifted to 3443, 3393 and 3313 cm\(^{-1}\). The three H-bonded features, all significantly broadened, are assigned starting at the blue-most feature to the methanolic OH, phenolic OH and water OH group (green water molecule). The LM model mid-IR spectrum matches well, but shows slight overestimation of the shift for the H-bonded features. The harmonic results however underestimate the shift for the strongest H-bonds, while matching quite well for the weaker bound OH stretching vibrations.

The peak at 765 cm\(^{-1}\) in the far-IR spectrum of SLG–2w is associated with the collective CH out-of-plane torsion vibration of the benzene hydrogen atoms and is observed in the similar position for bare saligenin and SLG–1w. In the case of SLG–2w...
this specific vibration is coupled to the H-bonded OH torsion of the green water, so that it is assigned to that local mode. The shift of the methanolic OH torsion is the largest of the SLG–2w modes, determined with respect to the free OH torsion vibration in phenol which is positioned at 309 cm$^{-1}$. This peak, calculated at 897.7 cm$^{-1}$, is positioned beyond the measured range (<800 cm$^{-1}$), despite the mid-IR results which suggest that this is the weakest of the three H-bonds. Of the other two torsional features especially the peak labeled T$_I$ is reproduced excellently by BOMD, in width, position and intensity. Several intense features are measured in the 150 to 400 cm$^{-1}$ range, where the free torsion vibrations and the H-bond stretching vibrations are positioned. This range holds many transitions for the SLG–2w cluster, with large differences in the relative intensities belonging to the different type of normal modes. This effect has led to several ambiguous assignments, and possibly causes spectral congestion. Finally, not all intense features corresponding to water vibrations in the far-IR spectrum are assigned within the experimental spectrum. These features originate from several collective water vibrations, such as libration and wagging, of which the impact on the H-bond network is not easily interpreted, since many of the H-bonds are deformed simultaneously.

Saligenin–(H$_2$O)$_3$

IR spectra for two different isomers were identified for the SLG–3w clusters. The mid-IR spectra of both SLG–3w isomers (Fig. 4) show eight clearly identifiable bands. The top panel shows the spectrum and assigned structure of the SLG–3w-a cluster, which forms a 10-membered cycle through OH···O linkages similar to the phenol–(H$_2$O)$_4$ structure$^{42}$ and isolated (H$_2$O)$_5$-cluster. The addition of an OH···π bond further stabilizes the bent ring structure of the assigned geometry. The SLG molecule within the SLG–3w-a cluster is approximately in its native structure, although the methanolic substituent is rotated to enable the formation of the H-bonded cycle. This specific isomer is the lowest energy structure amongst the SLG–3w clusters that were considered in our conformational search (see Fig. SI.9, ESI). The SLG–3w-a structure deviates from the structure of the benzylic alcohol–(H$_2$O)$_3$ complex (skewed cycle with four H-bonds including an OH···π bond$^{43}$), contrary to the smaller SLG–water cluster described previously.

In the mid-IR spectrum of SLG–3w-a, eight bands are observed. Two free OH stretching modes are nearly coincident at 3720 cm$^{-1}$. The peak at 3657 cm$^{-1}$ is assigned to the OH···O bond OH stretching vibration. The position and width of the six bands below 3550 cm$^{-1}$ indicate that they are involved in H-bonding, although the water network only exhibits five OH···O H-bonds. The features at 3501, 3467, 3345, 3286 and 3204 cm$^{-1}$ are assigned to the progressively stronger H-bonded OH stretching vibrations. The broad peak at 3143 cm$^{-1}$ is assigned to a water bending overtone which gains intensity through a Fermi coupling (green, OT), which is predicted by the LM model calculations at 3204 cm$^{-1}$. The match between the
LM model spectrum and the experimental spectrum is excellent in the range from 3250 to 3750 cm$^{-1}$. However, it appears that the strength of the H-bond donated by the green water is underestimated. This is consistent with the MP2 calculations on benzene–(H$_2$O)$_6$ (see Fig. S1.1, ESI†). If the frequency of this LM OH stretching vibration would be shifted slightly, then the shift due to the Fermi coupling will increase, improving the agreement sharply. In the harmonic spectrum we see that the agreement to the experimental data is good, although the frequency shifts are underestimated. The relative intensities are comparable to the LM model intensities, as expected.

SLG–3w-a shows features throughout the full far-IR range, while the range from 350 to 600 cm$^{-1}$ showed negligible vibrational activity for the smaller clusters. The H-bonded torsion peak of the green water molecule in the BOMD simulation, shifted beyond the measured range, has been cut in intensity, because it showed an unrealistic intensity dominating the spectrum. Both models reproduce the cluster of peaks in between 700 and 800 cm$^{-1}$, where the methanolic (orange, TB), phenolic (red, TB) and central water (cyan, TB) OH torsions are positioned. The BOMD spectrum shows more congestion than is actually present in the experimental spectrum. The central part of the far-IR spectrum shows features with lower intensity, including the weakly bound OH torsion vibrations. The highest intensity feature in this range at 619 cm$^{-1}$ is assigned to water wagging coupled to a collective CCC in-plane bending vibration in the benzene ring. This vibrational mode is reproduced well by BOMD, while M06-2X underestimates the intensity. In the low-frequency range from 100 to 350 cm$^{-1}$, many vibrational modes have been assigned (see Fig. 4), of which only four peaks cannot be matched to experimental features as they are beyond the measured window. A multitude of features can be observed in this range with strong variations in intensity, complicating assignments. The relative intensities of these features are overestimated by both levels of theory. Moreover, several peaks are assigned in a different order than would be expected based on their frequency. The assignment is based on strong shifts and relative intensities of OH torsion features observed for similar molecules in the far-IR region.$^{26}$

The assigned geometry of the SLG–3w-b cluster also includes an H-bonded, 10-membered cycle, with the direction of H-bonding reversed with respect to the SLG–3w-a cluster. The H-bonded cycle is skewed, causing this SLG–3w geometry to be energetically less favorable (4.8 kJ mol$^{-1}$) relative to the SLG–3w-a structure (see Fig. SI.9, ESI†). This isomer is the only isomer where the OH · · · O H-bond of the native SLG structure is broken. The mid-IR spectrum of SLG–3w-b presented in the lower right panel of Fig. 4 shows two closely spaced transitions at 3714 cm$^{-1}$ and 3721 cm$^{-1}$ that originate from the stretching of the free OH groups of the green and cyan water molecules, respectively. The intense transition at 3664 cm$^{-1}$ is very similar to the one observed for SLG–3w-a, and is also assigned to the weak OH · · · π bonded OH stretching vibration. A low-intensity feature is measured at 3566 cm$^{-1}$, which originates from the stretching vibration of the methanolic OH group. The position of the methanolic OH stretching feature is blue-shifted with respect to the same vibration in the SLG–3w-a isomer. The inverted H-bonded cycle strains this particular H-bond, hereby limiting its strength. The four high intensity, increasingly broadened features at 3446, 3364, 3327 and 3231 cm$^{-1}$ are assigned to the H-bonded purple water, green water, blue water and phenolic OH stretching vibrations. The red-most feature in the spectrum at 3231 cm$^{-1}$ is further broadened by the presence of additional HOH bending overtones. Moreover, the peak at 3198 cm$^{-1}$ is assigned to the water bending overtone of the purple water. The highlighted overtone features (Fig. 4) gain intensity predominantly from the assigned OH stretching vibrations, however the overtones are all complex mixtures of many involved states (see Hamiltonian matrix in Table SI.10, ESI†). Both levels of theory reproduce the five blue-most features well. The functional M06-2X overestimates the shifts of the green and blue H-donor OH stretching and the phenolic OH stretching vibrations. The LM model displays a systematic blue-shift with respect to the measured features, consistent with the results for SLG–3w-a.

The OH torsion vibration, typically observed in the far-IR, is shifted to 969.3 cm$^{-1}$ for the phenolic OH group (red TB, BOMD frequency) beyond the displayed far-IR range. The far-IR spectrum of SLG–3w-b (see bottom left panel) displays a sharp peak at 749 cm$^{-1}$ and several congested, broad features covering large parts of the spectral region. The sharp peak at 749 cm$^{-1}$ (dashed grey line to guide the eye) originates from the collective CH out-of-plane torsional mode of the benzene ring. This CH mode has no significant contribution from an OH torsion vibration in the case of the SLG–3w-b isomer. The peaks assigned to the OH torsion vibrations of the H-donor OH groups of the purple and green water molecules are positioned directly to the right and left of this feature. The range from 400 to 600 cm$^{-1}$ is dominated by a ridge of features of decreasing intensity. These features are reproduced by BOMD and M06-2X, although the intensity of the methanolic OH torsion measured at 537 cm$^{-1}$ is overestimated by both methods. The experimental spectrum shows several low-intensity features at the low frequency side, with the onset of a more intense feature at the end of the measured IR range. The low intensities show that the OH torsion vibrations present there, which normally provide quite high intensities, are coupled to other modes resulting in quenched transition dipole moments.

### 4 Hydrogen bonds and local mode shifts

Three different types of vibrational modes are highlighted in the spectra presented in Fig. 3 and 4 which are expected to be highly diagnostic for the H-bond network formed in the SLG-water clusters: OH stretching (mid-IR), OH torsion and H-bond stretching (both far-IR). We have chosen to disregard all couplings which may be present in the performed assignments and to view the vibrations as local modes (stretching, bending and torsion), where the local mode with the largest amplitude displacement within the vibrating moiety is assigned. A decomposition of the normal modes into local modes shows that for all vibrations the dominant local mode
has a significant higher contribution than the other local modes. The contributions of the assigned local modes to their respective normal modes are largest for the OH stretching vibrations and smallest for the H-bond stretching vibrations.

In SLG–3w-a for example, the methanolic OH stretching local mode contributes a dominant 50.4% to the normal mode at 3329.4 cm⁻¹ in the M06-2X spectrum. The second strongest contribution to this normal mode is 11.3% of the H-donor OH stretch of the green water. The most localized normal mode at 3727.7 cm⁻¹ in the mid-IR spectrum of SLG–3w-a is assigned to the free OH stretching vibration of the green water molecule for 63.5%. Considering the OH torsion vibrations, we observe that the methanolic OH torsion local vibration contributes 25.6% to the normal mode at 781 cm⁻¹ in the M06-2X far-IR spectrum, and that the second strongest contribution to this normal mode comes from the torsion of the H-donor OH moiety of the violet water molecule at 8.6%. For the OH torsion vibrations, we also observe that the free OH groups are most localized, with the green water free OH torsion contributing 29.8% to the normal mode calculated at 214.5 cm⁻¹. This result implies that the H-bonds facilitate local mode couplings. Finally, we see that a very high number of local modes make small contributions to the normal coordinates assigned to H-bond stretching. The H-bond stretching vibration of which the green water molecule is the donor for example participates for 7.9% in the normal mode at 267.5 cm⁻¹. The second most dominant local motion in this normal motion is a rocking vibration of the green water molecule with a 5.4% contribution. The average expected contribution of a single internal coordinate for a SLG–(H₂O)₃ cluster assuming equipartition would be 100/(3N−6) = 1.39%. Normal modes become more localized for smaller clusters. The normal vibration in bare SLG at 423.5 cm⁻¹ in the M06-2X far-IR spectrum consists for 54.9% of the assigned methanolic OH torsion local mode (orange) and the methanolic OH stretch local mode dominates the mid-IR normal mode at 3637.1 cm⁻¹ with a contribution of 83.8%. As the Fourier transforms of intra-molecular coordinate time correlation functions extracted from the BOMD simulations are local mode signatures by construction, these provide valuable complimentary information to the harmonic calculations for local mode assignments.

The positions of the assigned local modes are evaluated as a function of the H-bond strength, which shows systematic behavior both for the mid-IR and far-IR features in terms of frequency shifts. Two different measures for the H-bond strength are evaluated in this work, namely the H-donor OH length ℓOH and the length of the H-bond ℓHB. It was previously established that ℓOH is a good measure for the strength of intramolecular H-bonds. The ℓOH is not heavily influenced by the covalently bound ring closed by an intramolecular H-bond, while ℓHB and various other geometrical features are strongly dependent on the geometry of the molecule. In case of intermolecular H-bonds – when this covalently bound ring is absent and the intermolecular distance is mostly determined by the H-bond network – ℓHB might properly predict the H-bond strength as well. The geometrical parameters are extracted from the structures optimized on the MP2/6-311++G(2d,p) level of theory.

The frequencies of the three selected local modes are displayed in Fig. 3 as a function of ℓOH with respect to the minimum OH bond length ℓOH,min of 0.9629 Å found in the SLG–water clusters studied here. Calculated results on the two anharmonic levels (MP2 LM model and BOMD) are displayed as squares. The grey line shows a linear fit to these calculated results. The assigned experimental bands are represented by crossed circles with a pink line. The color of the symbols indicates the interaction that the specific OH group is involved in: the free OH moieties are purple, the OH groups involved in an OH···π interaction are red, the H-bonded OH groups attached to the SLG molecule are blue, and the H-bonded OH groups in water are black.

The frequencies of all the H-bond deforming vibrations display a linear correlation with the lengthening of the optimized OH length. The free OH stretching vibrations in the water molecules are the blue-most features, blue-shifted with respect to the OH stretching vibrations in e.g. phenol and methanol because of the H-bond accepted by the oxygen atom in the OH moiety. The OH···π bonded groups show small red-shifts, and the H-bonded groups show large shifts, up to 500 cm⁻¹. The linear fit to the experimental data was calculated using linear regression, with a slope of −172 ± 7 cm⁻¹ pm⁻¹ and −169 ± 6 cm⁻¹ pm⁻¹ for calculated shifts. This is significantly smaller than the slope previously found for intramolecular H-bonding (−233 cm⁻¹ pm⁻¹). To test whether the difference was a result of the different levels of theory used to optimize the cluster geometries (MP2 vs. B3LYP-D3) we have performed the same analysis of the water clusters using B3LYP-D3 optimized geometries, which resulted in a slope of −174 ± 15 cm⁻¹ pm⁻¹. The difference between the slopes found for the two levels of theory is well within the error bar of the correlation itself, which indicates that the significant difference observed between the results presented here and the results on intramolecular H-bonds is related to differences in the formed bonds. Upon further inspection, the steeper slope for intramolecular H-bonds was found to result from including the resonance assisted H-bond in salicylic acid. Excluding this resonance assisted H-bond from the dataset significantly reduces the slope, closely matching the slope reported here. The result found for the shifts of OH stretching vibration as a function of H-bond strength is thus more widely applicable, as long as similar H-bond interactions are regarded.

The OH torsion frequency (Fig. 5b) and the H-bond stretching frequency (Fig. 5c) both blue-shift when engaged in stronger H-bonds. For these two modes, not all vibrations are assigned to experimental features as these are positioned outside the measured window. Logically, Fig. 5c shows no free OH features, as no HB stretching frequency can be assigned for a free group. Both vibrations positioned in the far-IR show a linear correlation to ℓOH similar to the OH stretching frequency, with slopes of 203 ± 18 for OH torsion and 70 ± 10 cm⁻¹ pm⁻¹ for HB stretching (238 ± 15 and 82 ± 9 cm⁻¹ pm⁻¹ for the calculated shifts). Similar to the correlations previously found for intramolecular H-bonding, the OH torsion frequency in the far-IR is more sensitive to H-bond formation than the OH stretching
The frequency of the H-bond stretching vibration displays a more gradual slope compared to the OH torsion and OH stretch vibrations. In terms of relative shifts however, the H-bond stretching vibration is shifted up to a factor of three from 100 to 300 cm\(^{-1}\), which is by far the largest relative shift within the dataset.

In Fig. 6, the frequencies of the three H-bond related modes are displayed as a function of the H-bond length \(L_{\text{HB}}\) found in the optimized structures on the MP2/6-311++G(2d,p) level of theory. The H-bond length varies from 1.71 Å belonging to the H-bond donated by the phenolic OH group of SLG–3w-b in Fig. 4 to 3.731 Å for the very weak OH–π interaction in SLG (H-donor colored in orange in Fig. 3). The H-bond length of the OH stretching and OH torsion vibrations results from the H-donor OH moiety. Notable is the nonlinear correlation between the shifts in the three presented local modes and \(L_{\text{HB}}\). For long H-bonds, \(L_{\text{HB}} > 2.25\) Å the influence on the vibrational frequencies is negligible, which explains the observed saturation behavior for long H-bonds. As the H-bonds become shorter and less strained, their bond strength will increase and with that also their influence on the related vibrational frequencies.

The correlation between the H-bond related vibrational frequencies and the H-bond lengths can be fitted by an exponential function of the form

\[
\nu(L_{\text{HB}}) = \nu_{\text{lim}} \{ 1 + e^{-(L_{\text{HB}} - L_{\text{HB,0}})} \}
\]

where \(\nu_{\text{lim}}\) is the limiting value for the frequency at infinitely large \(L_{\text{HB},0}\) is the length where the frequency crosses zero in the case of the OH stretch vibrations (minus sign in eqn (2)), and where the frequency is 2\(\nu_{\text{lim}}\) in the case of the OH torsion and HB stretching vibrations (plus sign in eqn (2)). This function is chosen to fit our data rather than directly describing the complex underlying physical processes, which is beyond the scope of this paper. As can be expected, the function does not predict the correct limiting behavior for small \(L_{\text{HB}}\). For the OH stretching vibration in Fig. 6a, \(\nu_{\text{lim}}\) can be interpreted as the free stretching frequency, which is 3682 cm\(^{-1}\) in gas phase methanol, and 3657 cm\(^{-1}\) in phenol. The result from the fit is 3637 ± 22 cm\(^{-1}\) which is lower than the expected value for phenol, but matches within the error bar. A similar behavior is found for the OH torsion vibration. For the frequency of the OH torsion vibration, we found an experimental limiting frequency \(\nu_{\text{lim}} = 328 ± 70\) cm\(^{-1}\). While the free OH torsion frequency of phenol is positioned at 309 cm\(^{-1}\), the HB stretching frequency is expected to go to zero for infinite H-bond lengths, however, the limiting frequency \(\nu_{\text{lim}}\) is 128 ± 24 cm\(^{-1}\). In all three cases, the H-bond interactions of adjacent atoms were excluded. Experiments on rotaxanes have shown that although a certain moiety itself is not involved in an interaction, but when a neighboring group is, still a frequency shift of the free moiety is observed, i.e. the secondary effect. This results in a lower (for the OH stretching) and higher (for OH torsion and H-bond stretching) limiting shift than expected.

The large difference in behavior of the local mode shifts as a function of the OH bond length and of the H-bond length shown here is caused by the large difference in bond strength in between the covalent OH bond and the electrostatic H-bond. The relative change in the OH bond length induced by the additional formation of an H-bond is a few percent and can therefore be seen as a small perturbation. The relative changes in the length are much larger for the H-bond (more than 200%), so that the nonlinear regime is reached where the H-bond is so weak that its influence becomes negligible. Because of this, H-bonds are an interesting system to study the limiting behavior of molecular interactions stretched to far beyond their optimized length. We expect that for small variations in the H-bond length with respect to the optimal length, the correlation of the presented normal modes is linear as it is for the OH bond length. In our dataset, this linear regime is emerging, although not enough strong H-bonds are present to conclusively show this behavior.

Fig. 5 Frequencies of the (a) OH stretching, (b) OH torsion and (c) HB stretching vibrations as a function of the lengthening of the vibrating OH moiety \(L_{\text{OH}}\), corrected for the minimum OH length \(L_{\text{OH, min}} = 0.9629\) Å found in this dataset. The OH bond lengthening represents the strength of the involved H-bond, where a longer OH bond corresponds to a stronger H-bond. Calculated results are shown as squares, experimental results as crossed circles. For all three vibrations a linear relation is observed between the frequency shift and the OH length, with linear fits to the experimental frequencies presented as pink lines and fits to the calculated data as grey lines.

This results in a lower (for the OH stretching) and higher (for OH torsion and H-bond stretching) limiting shift than expected.

20352 | Phys. Chem. Chem. Phys., 2017, 19, 20343–20356 This journal is © the Owner Societies 2017
and HB stretching vibrations described here. Most vibrational modes in the far-IR are more localized, and is sensitive to subtle structural changes. Spectra were not sufficiently diagnostic, showing that the far-IR disambiguation between different structure when the mid-IR static DFT calculations. The far-IR spectra proved crucial for interpretation of the shifts of these modes in terms of the formed interactions less straightforward. Assignments based solely on the far-IR spectra remain challenging, due to the large variety of modes present in combination with the deviations between experiments and calculations of far-IR spectra.26,34,35

The measured spectra were modeled using different levels of theory, namely harmonic M06-2X, anharmonic MP2 with the coupled LM model and dynamical BOMD calculations. The LM model MP2 calculations reproduce the positions of the measured features in the mid-IR well, while the harmonic M06-2X spectra systematically underestimate the shifts of strongly H-bonded OH moieties, except for SLG–3w-b. The relative intensities (surface area of the experimental features) are matched closely by the MP2 calculations as well as by the harmonic M06-2X calculations, although the match becomes progressively worse for more strongly H-bonded and/or coupled vibrations, for which both theories overestimate the intensity. The anharmonic couplings in the MP2 calculations additionally predict the appearance, shifts and intensity changes of overtone features that gain intensity through Fermi couplings for negligible additional computational cost. In the case of water clusters this is crucial because the presence of additional features can alternatively be interpreted in terms of additional OH oscillators and thus larger water clusters.

The LM model makes use of a two-parameter scaling factor that is derived from the spectrum of the benzene–(H2O)6 book conformer (Fig. SL1, ESI†). The good fit of the LM model spectrum to the experimental data shows that the employed MP2 method is transferable from the benzene to the saligenin system. A similar two-parameter scaling function was determined from the benzene–(H2O)6 dataset for the M06-2X level of theory: $\nu_{\text{scaled}} = 0.807 \nu_{\text{local}} + 526.24$ in cm$^{-1}$. The results of this two-parameter scaling are visible in Fig. SL2 (ESI†) alongside the spectrum produced using the constant scaling factor. While both OH stretch frequencies are blue-shifted in the constant scaled spectrum with respect to the experimental data, the blue-shift is larger for the two-parameter scaled spectrum. This shows that the M06-2X two-parameter scaling function is not transferable from the benzene water to the saligenin water system, unlike the scaling function for the MP2 method.

The BOMD far-IR spectrum immediately stands out for the broadened and more convoluted features when compared to the harmonic spectrum, especially for the larger clusters. Especially features originating from floppy motions with a lot of motion become broadened, which is the case for the H-bonded water network (e.g. the “T“-feature for SLG–2w in Fig. 3). Intensities of large amplitude motions such as the OH torsion vibrations are systematically overestimated by both levels of theory, but show improvements when applying dynamical BOMD calculations instead of harmonic M06-2X calculations. BOMD also results in a better agreement between the calculations and experiments in the range from 100 to 400 cm$^{-1}$, where the free and weakly bound torsional vibrations are positioned. Possible overtones and combination bands in the far-IR spectra have not been considered, because neither of the used theories reproduces
these anharmonic features in their current configuration. We do not expect large contributions from these anharmonic effects in this spectral region due to the low intensity of the features present in the 100 to 400 cm\(^{-1}\) range.

A few of the localized features in the far-IR that we have identified to be diagnostic of the H-bond network are located outside the measured range. The H-bond stretching vibrations belonging to weaker H-bond are mostly present to the red of the experimental cut-off at 220 cm\(^{-1}\), while several strongly bound OH torsion features are shifted beyond the limit at the blue-side of the experimental spectrum. To include these signatures belonging to weak and strong H-bond interactions when characterizing H-bonded networks, the experiment should range from 100 to 1000 cm\(^{-1}\), to enable assignment and characterization of the features at the extremes of the far-IR.

The assigned structures of Fig. 3 and 4 can be regarded in terms of the mechanism of formation of the SLG-water clusters. For the SLG–1w cluster, the water molecule is positioned in the hiatus in between the methanolic OH group and the phenyl ring. The SLG–2w cluster is shaped similarly, with the two water molecules bound linearly as they would be in isolation.\(^{45,46}\) The formation of the SLG–2w cluster can therefore occur both through sequential additions of the water molecules, or through the interaction of SLG with a preformed (H\(_2\)O)\(_2\)-cluster. The SLG–3w-a geometry is similar to the isolated cyclic water pentamer, forming a bent 10-membered cycle through OH···O linkages of SLG with water molecules. This structure can be formed through the addition of a water molecule to the SLG–2w cluster on the phenolic side of the water molecules, while the OH···O bound water rotates to complete the ring. Such a mechanism cannot be imagined for SLG–3w-b, which requires a significant conformational change of the involved SLG molecule. Possibly the barrier for conformational change for SLG is low, so that both conformers exist prior to supersonically expanding the molecules into the vacuum, and either can be stabilized through the formation of an H-bond network.

The correlation between the frequencies of H-bond deforming modes to calculated geometrical parameters depends on the specific level of theory used to compute the geometry. To determine these relations without the influence of computational choices, the shifts of the vibrational mode frequencies can be compared to the accurate geometric parameters found by rotational spectroscopy.\(^{46}\) This will allow for the quantification of the relation between band positions and bond lengths in molecules or molecular clusters.

6 Conclusions

The IR spectra of the SLG molecule and of SLG water complexes with up to three water molecules were presented in this paper, measured in specific parts of the far-IR (220 to 800 cm\(^{-1}\) and mid-IR (3100 to 3800 cm\(^{-1}\)) spectrum. Despite being detected in the mass channels belonging to SLG–(H\(_2\)O)\(_i\), and SLG–(H\(_2\)O)\(_2\) clusters, the measured spectra were assigned to a SLG–1w and SLG–2w cluster and two SLG–3w clusters. The assignments were principally performed on grounds of the mid-IR spectra, although the far-IR spectrum was crucial in the disambiguation of the assignment of the SLG–2w and SLG–3w-b geometries. For cluster sizes up to SLG–2w, the assigned structures show similarities to the structures found for benzyl alcohol-water clusters, while the SLG–3w geometries show H-bond networks similar to the isolated, cyclic water pentamer. All isomers except for SLG–3w-b contain the native structure of the SLG molecule.

Several levels of theory were employed to model the spectra and were consecutively compared: the LM model MP2 calculations match the experimental spectra in the mid-IR well, including overtone features that gain intensity through Fermi couplings. The harmonic M06-2X calculations show good agreement, but going beyond the harmonic approximation with a simple LM model aids in assignments. The spectral density in the far-IR region is higher than in the mid-IR region, making the assignments of the observed peaks challenging and required the use of sophisticated dynamic BOMD simulations.

We have shown that the shifts in the frequencies of three different local modes (OH stretching, OH torsion and H-bond stretching) all correlate linearly to the OH length, even when both intra- and intermolecular H-bonds are included in this analysis. The OH torsion vibration displays the largest absolute frequency shift when it becomes H-bonded, followed by the OH stretching vibrations and finally the HB stretching frequency. The H-bond stretching vibration however displays the largest relative frequency shifts. When plotted versus the H-bond length, the frequency shifts displayed non-linear behavior, saturating for highly stretched H-bonds. The nonlinear behavior is described using an exponential function, which shows that the shifts found for infinitely long H-bonds is consistently smaller than what is expected based on the absence of an H-bond.

Acknowledgements

We gratefully acknowledge the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organisation for Scientific Research (NWO) for the support of the FELIX Laboratory. Moreover, we would like to thank the entire team of the FELIX Laboratory for experimental assistance and useful discussions, and S. Bakels for assisting with the conformational search. D. P. T. and E. L. S. acknowledge support from NSF Grant No. CHE-1566108. The HPC resources used for quantum chemical calculations were provided by SurfSara (NWO Physical Sciences, project MP-264-13), by GENCI-France (Grant 082484) and by the phoenix computing cluster at UW-Madison (NSF Grant CHE-0840494). Lastly, we thank D. R. Galimberti for helpful discussions and coding the removal of rotational motions from the dipolar trajectories.

References


Fingerprints of Inter- and Intramolecular Hydrogen Bonding in Saligenin-Water Clusters Revealed by Mid- and Far-Infrared Spectroscopy

Daniël J. Bakker, Arghya Dey, Daniel P. Tabor, Qin Ong, Jérôme Mahé, Marie-Pierre Gaigeot, Edwin L. Sibert III and Anouk M. Rijs

Radboud University, Institute for Molecules and Materials, FELIX Laboratory, Toernooiveld 7c, 6525 ED Nijmegen, The Netherlands

Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.

Current address: Department of Chemistry and Chemical Biology, Harvard University, Cambridge MA, 02138 USA

LAMBE CNRS UMR8587, Université d’Evry val d’Essonne, Blvd F. Mitterrand, Bât Maupertuis, 91025 Evry, France.

Corresponding authors
a.rijs@science.ru.nl, sibert@chem.wisc.edu, mgaigeot@univ-evry.fr

Table of contents
Scaling factors
Fig SI.1. Determination of the two-parameter scaling function used for all MP2 calculations
Fig SI.2. Comparison of constant and two-parameter scaling on the SLG mid-IR spectrum

REMPI spectra
Fig SI.3. REMPI spectra of SLG-water clusters, with assignments of UV peaks to conformers

Assignment of saligenin water cluster structures
Fig SI.4. Structures and energies of lowest energy isomers of saligenin-(H2O) clusters
Fig SI.5. Mid-IR spectrum of SLG-1w, with theoretical spectra of the isomers of Fig SI.2
Fig SI.6. Structures and energies of lowest energy isomers of saligenin-(H2O)2 clusters
Fig SI.7. Mid-IR spectrum of SLG-2w, with theoretical spectra of the isomers of Fig SI.4
Fig SI.8. Far-IR spectrum of SLG-2w, with theoretical spectra of isomers sw2-1 and sw2-3
Fig SI.9. Structures and energies of lowest energy isomers of saligenin-(H2O)3 clusters
Fig SI.10. Mid-IR spectrum of SLG-3w, with theoretical spectra of the isomers of Fig SI.4

References
Atomic coordinates and local mode model Hamiltonian matrices
Scaling factors

Figure SI.1. Infrared (IR) frequencies absorbed by the benzene-(H2O)6 cluster versus the calculated frequencies at the MP2/6-311++G(2d, p) level of theory based on the “book” isomer. The linear fit to the data in red represents the two-parameter scaling function used for all MP2 calculations reported throughout this work. The experimental frequencies were originally measured by Pribble and Zwier and later reproduced at higher resolution by Zwier and co-workers.

Figure SI.2. Mid-IR spectrum of bare SLG. M06-2X/6-311++G(2d, p) calculations with a constant scaling factor (0.9426) and with a two-parameter scaling function are displayed in blue and green, respectively. The two-parameter scaling function is based on the benzene-(H2O)6 cluster: \[ \nu_{\text{scaled}} = 0.8073 \nu_{\text{local}} + 526.24 \text{ in cm}^{-1} \]. The red-shift of both OH stretch frequencies is underestimated by the constant scaled spectrum, yet more so by the two-parameter scaled spectrum, which shows that the M06-2X two-parameter scaling does not transfer from benzene to SLG, as it does for the MP2 method.
Assignment of saligenin water cluster structures
The structural assignments are primarily based on the comparison of the experimental mid-IR spectra with ab initio quantum chemical calculations, with support provided by the far-IR spectra. Mid-IR spectroscopy has been applied successfully to elucidate the structural preferences of similar systems such as phenol water\textsuperscript{4-5}, benzyl-alcohol water\textsuperscript{6} and propofol water\textsuperscript{7-8}. The isomers presented here were generated using simulated annealing molecular dynamics calculations. The level of theory applied to calculate the mid-IR spectra and optimized structures is MP2/6-311++G(2d,p)\textsuperscript{9}, using the local mode approach, which has been applied previously to successfully identify complex benzene water clusters by Tabor et al.\textsuperscript{1}.
Saligenin water 1

Figure SI.4. Structures and stabilization energies of the lowest energy isomers of the saligenin-(H₂O) cluster, calculated on the MP2/6-311++G(2d, p) level of theory. The assigned structure is sw1-1, the lowest energy isomer, which is similar to the structure of the benzyl alcohol-(H₂O) cluster.

Figure SI.5. Mid-infrared spectrum of the SLG-1w cluster, alongside anharmonic calculations performed on the MP2/6-311++G(2d,p) level of theory, using the local mode approach. The assigned structure is the lowest energy isomer sw1-1, which matches best both for the positions and intensities of the experimental features.
Saligenin water 2

Seven low-energy isomers (< 12 kJ/mol) were found in the conformational search for SLG-2w (Figure SI.4). The measured SLG-2w spectrum (Figure SI.5) matches well with the IR spectra of three different isomers, namely sw2-1 (water wire), sw2-5 (water cycle) and sw2-7 (water cycle). Sw2-1 is the most stable isomer, being 6.3 and 10.2 kJ/mol more favorable than the two other conformers. On the blue side of the mid-IR spectrum (free and loosely bound OHs), sw2-1 and sw2-5 provide a significantly better match than sw2-7. As this region is the most easy to determine quantum chemically, being the least influenced by interactions and anharmonicities, it is the most reliable part of the spectrum. In the far-IR regime, displayed for these isomers at the M06-2X/6-311++G(2d, p) level of theory in Figure SI.6, the wire isomer fits better, especially in the 600-800 cm$^{-1}$ region and around 300 cm$^{-1}$.

In the REMPI spectrum (Figure SI.2), it is observed that the origin transition of SLG-1w is shifted by 92 cm$^{-1}$ with respect to the origin transition of bare SLG, while the origin of SLG-2w is positioned closer to the bare SLG origin with a shift of only 19 cm$^{-1}$. This behavior can be linked to the formation of a cyclic water network, e.g. in the case of propofol-water clusters$^7$. This interpretation would support the assignment of a water cycle isomer. However, this pattern of UV wavelength shifts was observed for benzyl-alcohol water clusters$^6$, where the water-wire isomer was convincingly assigned for the benzyl alcohol-(H$_2$O)$_2$. Moreover, the conformation of the SLG molecule within the sw2-3 isomer is quite different from the bare SLG molecule, so that shifts of the UV transitions cannot be interpreted in terms of the assignments without further work, such as transition state calculations. Therefore, the structure sw2-1 is assigned to be present based on the energetics, UV behavior and structurally diagnostic IR spectra.

---

**Figure SI.6.** Structures and stabilization energies of the lowest energy isomers of the saligenin-(H$_2$O)$_2$ cluster, calculated on the MP2/6-311++G(2d, p) level of theory. The assigned structure is sw2-1, the lowest energy isomer, which is similar to the structure of the benzyl alcohol-(H$_2$O)$_2$ cluster$^6$. 
Figure SI.7. Experimental mid-IR spectrum of the SLG-2w cluster and anharmonic calculations performed on the MP2/6-311++G(2d,p) level of theory using the local mode approach. The assignment cannot be performed based on the mid-IR, with isomers sw2-1, sw2-5 and sw2-7 all matching well to the experimental spectrum.

Figure SI.8. Experimental far-infrared spectrum of the SLG-2w cluster, with the calculated spectra of isomers sw2-1 and sw2-3 on the harmonic M06-2X/6-311++G(2d,p) level of theory. The spectrum of the sw2-1 isomer reproduces the experimental spectrum especially well in the region from 600 to 800 cm$^{-1}$, as well as below 300 cm$^{-1}$. Based on these characteristics, structure sw2-1 is assigned.
Saligenin water 3

For the SLG-3w-a spectrum, the isomer sw3-1 clearly provides the best match, which is also the most stable structure. Therefore, we have assigned the structure sw3-1 to SLG-3w-a. For SLG-3w-b, the isomers sw3-2 and sw3-3 both provide a good match to the spectrum. At the blue side of the spectrum, above 3500 cm\(^{-1}\), where the quantum chemical calculations are perform optimally, the sw3-2 spectrum matches significantly better to the measured spectrum. Below 3500 cm\(^{-1}\), where the H-bonded OH stretching vibrations are present, the sw3-3 structure fits slightly better, although the spectrum of the sw3-2 geometry also matches well. The sw3-2 spectrum displays a blue-shift below 3500 cm\(^{-1}\) with respect to the measurements, consistent with the match of the sw3-1 spectrum to SLG-3w-a. Lastly, the sw3-2 structure is slightly more stable (0.5 kJ/mol difference). Based on this argumentation structure sw3-2 is assigned to the SLG-3w-b spectrum.

Figure SI.9. Structures and stabilization energies of the lowest energy isomers of the saligenin-(H\(_2\)O)\(_3\) cluster, calculated on the MP2/6-311++G(2d, p) level of theory. The assigned structures are sw3-1 and sw3-2, for the SLG-3w-a and SLG-3w-b spectra respectively. The water molecules in these two isomers form interactions that closely resemble the structures of isolated three and five water clusters\(^{10}\).
Figure SI.10. Experimental mid-infrared spectra of the two SLG-3w clusters, alongside anharmonic calculations performed on the MP2/6-311++G(2d,p) level of theory using the local mode approach. The SLG-3w-a IR spectrum (black) is assigned to isomer sw3-1, where the water molecules form a 10-membered H-bonded ring together with the two OH groups of SLG. The SLG-3w-b IR spectrum (pink) is assigned to the sw3-2 isomer, a 10-membered ring including five H-bonds as well, for which the H-bond direction around the ring is inverted with respect to sw3-1.
References


### Table SI.1: Atomic coordinates of the optimized geometry (MP2/6-311++G(2d,p) level of theory) of the saligenin molecule.

<table>
<thead>
<tr>
<th>Atom</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-0.81452</td>
<td>-1.21900</td>
<td>0.25471</td>
</tr>
<tr>
<td>C</td>
<td>0.56629</td>
<td>-1.35408</td>
<td>0.11203</td>
</tr>
<tr>
<td>C</td>
<td>1.36043</td>
<td>-0.23390</td>
<td>-0.18608</td>
</tr>
<tr>
<td>C</td>
<td>0.74202</td>
<td>1.00732</td>
<td>-0.36237</td>
</tr>
<tr>
<td>C</td>
<td>-0.63774</td>
<td>1.14907</td>
<td>-0.22157</td>
</tr>
<tr>
<td>C</td>
<td>-1.41153</td>
<td>0.03017</td>
<td>0.09468</td>
</tr>
<tr>
<td>H</td>
<td>-1.40181</td>
<td>2.10119</td>
<td>0.48845</td>
</tr>
<tr>
<td>H</td>
<td>1.35728</td>
<td>1.87154</td>
<td>-0.60147</td>
</tr>
<tr>
<td>H</td>
<td>-1.10377</td>
<td>2.11917</td>
<td>-0.35709</td>
</tr>
<tr>
<td>H</td>
<td>-2.48695</td>
<td>0.12594</td>
<td>0.20638</td>
</tr>
<tr>
<td>C</td>
<td>2.85186</td>
<td>-0.38668</td>
<td>-0.25894</td>
</tr>
<tr>
<td>O</td>
<td>3.25017</td>
<td>-1.49151</td>
<td>-1.09849</td>
</tr>
<tr>
<td>H</td>
<td>3.31710</td>
<td>0.54341</td>
<td>-0.59845</td>
</tr>
<tr>
<td>H</td>
<td>3.26288</td>
<td>-0.63396</td>
<td>0.72280</td>
</tr>
<tr>
<td>H</td>
<td>2.91958</td>
<td>-1.31106</td>
<td>-1.98962</td>
</tr>
<tr>
<td>O</td>
<td>1.11433</td>
<td>-2.59320</td>
<td>0.29413</td>
</tr>
<tr>
<td>H</td>
<td>1.99195</td>
<td>-2.58992</td>
<td>-0.13178</td>
</tr>
</tbody>
</table>

### Table SI.2: Local mode hamiltonian matrix of saligenin: local mode frequencies are on the diagonal, couplings off-diagonal.

<table>
<thead>
<tr>
<th></th>
<th>3640.8</th>
<th>-6.6000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-6.6000</td>
<td>3503.9</td>
</tr>
</tbody>
</table>

### Table SI.3: Atomic coordinates of the optimized geometry (MP2/6-311++G(2d,p) level of theory) of the sw1-1 cluster.

<table>
<thead>
<tr>
<th>Atom</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-1.67879</td>
<td>0.95109</td>
<td>0.70364</td>
</tr>
<tr>
<td>C</td>
<td>-0.41481</td>
<td>1.00991</td>
<td>0.11380</td>
</tr>
<tr>
<td>C</td>
<td>0.04287</td>
<td>-0.05158</td>
<td>-0.68733</td>
</tr>
<tr>
<td>C</td>
<td>-0.77735</td>
<td>-1.17149</td>
<td>-0.85871</td>
</tr>
<tr>
<td>C</td>
<td>-2.04107</td>
<td>-1.23725</td>
<td>-0.26949</td>
</tr>
<tr>
<td>C</td>
<td>-2.48980</td>
<td>-0.16577</td>
<td>0.50669</td>
</tr>
<tr>
<td>C</td>
<td>1.39341</td>
<td>0.05061</td>
<td>-1.34467</td>
</tr>
<tr>
<td>O</td>
<td>2.40435</td>
<td>0.49936</td>
<td>-0.42795</td>
</tr>
<tr>
<td>O</td>
<td>0.34789</td>
<td>2.11943</td>
<td>0.32789</td>
</tr>
<tr>
<td>H</td>
<td>-2.00607</td>
<td>1.78706</td>
<td>1.31371</td>
</tr>
<tr>
<td>H</td>
<td>-0.41662</td>
<td>-1.99629</td>
<td>1.46906</td>
</tr>
<tr>
<td>H</td>
<td>-2.66740</td>
<td>-2.11075</td>
<td>-0.41660</td>
</tr>
<tr>
<td>H</td>
<td>-3.47034</td>
<td>-0.20256</td>
<td>0.97091</td>
</tr>
<tr>
<td>H</td>
<td>1.67092</td>
<td>-0.91065</td>
<td>-1.78861</td>
</tr>
<tr>
<td>H</td>
<td>1.37288</td>
<td>0.80036</td>
<td>-2.13991</td>
</tr>
<tr>
<td>H</td>
<td>2.43130</td>
<td>-0.16420</td>
<td>0.28728</td>
</tr>
<tr>
<td>H</td>
<td>1.27378</td>
<td>1.87003</td>
<td>0.12458</td>
</tr>
<tr>
<td>O</td>
<td>1.69797</td>
<td>-1.57612</td>
<td>1.38857</td>
</tr>
<tr>
<td>H</td>
<td>1.79899</td>
<td>-1.69005</td>
<td>2.34080</td>
</tr>
<tr>
<td>H</td>
<td>0.74598</td>
<td>-1.47727</td>
<td>1.23844</td>
</tr>
</tbody>
</table>

### Table SI.4: Local mode Hamiltonian matrix of structure sw1-1: local mode frequencies are on the diagonal, couplings off-diagonal.

<table>
<thead>
<tr>
<th></th>
<th>3506.5</th>
<th>-16.360</th>
<th>-6.9300</th>
<th>-11.610</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-16.360</td>
<td>3425.4</td>
<td>-0.3700</td>
<td>-3.3500</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>-6.9300</td>
<td>-0.3700</td>
<td>3699.4</td>
<td>-61.240</td>
<td>45.250</td>
</tr>
<tr>
<td></td>
<td>-11.610</td>
<td>-3.3500</td>
<td>-61.240</td>
<td>3632.4</td>
<td>45.250</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>45.250</td>
<td>45.250</td>
<td>3146.7</td>
<td></td>
</tr>
</tbody>
</table>
Table SI.5: Atomic coordinates of the optimized geometry (MP2/6-311++G(2d,p) level of theory) of the sw2-1 cluster.

<table>
<thead>
<tr>
<th></th>
<th>Atomic coordinates</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1.72377</td>
<td>-0.66083</td>
</tr>
<tr>
<td>C</td>
<td>0.41335</td>
<td>-0.98096</td>
</tr>
<tr>
<td>C</td>
<td>-0.04267</td>
<td>-0.69765</td>
</tr>
<tr>
<td>C</td>
<td>0.82943</td>
<td>-0.08440</td>
</tr>
<tr>
<td>C</td>
<td>2.14158</td>
<td>0.23095</td>
</tr>
<tr>
<td>C</td>
<td>2.58728</td>
<td>-0.06684</td>
</tr>
<tr>
<td>C</td>
<td>-1.43688</td>
<td>-1.10745</td>
</tr>
<tr>
<td>O</td>
<td>-2.39928</td>
<td>-0.83605</td>
</tr>
<tr>
<td>H</td>
<td>0.38850</td>
<td>-1.57665</td>
</tr>
<tr>
<td>H</td>
<td>2.05062</td>
<td>-0.89969</td>
</tr>
<tr>
<td>H</td>
<td>0.47125</td>
<td>0.14260</td>
</tr>
<tr>
<td>H</td>
<td>2.81013</td>
<td>0.69232</td>
</tr>
<tr>
<td>H</td>
<td>3.60610</td>
<td>0.16587</td>
</tr>
<tr>
<td>H</td>
<td>-1.71646</td>
<td>-0.61408</td>
</tr>
<tr>
<td>H</td>
<td>-1.47415</td>
<td>-2.18857</td>
</tr>
<tr>
<td>O</td>
<td>-2.44925</td>
<td>0.14284</td>
</tr>
<tr>
<td>H</td>
<td>-1.31182</td>
<td>-1.47173</td>
</tr>
<tr>
<td>H</td>
<td>0.39465</td>
<td>2.40966</td>
</tr>
<tr>
<td>H</td>
<td>0.79075</td>
<td>1.54133</td>
</tr>
<tr>
<td>H</td>
<td>0.92043</td>
<td>2.7427</td>
</tr>
<tr>
<td>O</td>
<td>-2.19867</td>
<td>1.8486</td>
</tr>
<tr>
<td>H</td>
<td>-1.30421</td>
<td>2.13180</td>
</tr>
<tr>
<td>H</td>
<td>-2.80968</td>
<td>2.47871</td>
</tr>
</tbody>
</table>

Table SI.6: Local mode Hamiltonian matrix of structure sw2-1: local mode frequencies are on the diagonal, couplings off-diagonal.

Saligenin-(H2O)2 - Model Hamiltonian matrix

<table>
<thead>
<tr>
<th></th>
<th>3368.9</th>
<th>-25.130</th>
<th>-6.3200</th>
<th>-1.4500</th>
<th>-30.760</th>
<th>-11.350</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>-25.100</td>
<td>3351.5</td>
<td>-3.8200</td>
<td>-0.1400</td>
<td>-6.9800</td>
<td>-0.9600</td>
<td>45.2500</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-6.3200</td>
<td>-3.8200</td>
<td>3588.0</td>
<td>-49.660</td>
<td>-21.050</td>
<td>-0.0700</td>
<td>45.2500</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-1.4500</td>
<td>-0.1400</td>
<td>-49.660</td>
<td>3690.2</td>
<td>-12.060</td>
<td>-0.2600</td>
<td>45.2500</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-30.760</td>
<td>-6.9800</td>
<td>-21.050</td>
<td>-12.060</td>
<td>3405.4</td>
<td>60.730</td>
<td>45.2500</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-11.350</td>
<td>-0.9600</td>
<td>-0.0700</td>
<td>-0.2600</td>
<td>-60.730</td>
<td>3708.4</td>
<td>45.2500</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>45.250</td>
<td>45.250</td>
<td>0</td>
<td>0</td>
<td>3183.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>45.250</td>
<td>45.250</td>
<td>0</td>
<td>3241.2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table SI.7: Atomic coordinates of the optimized geometry (MP2/6-311++G(2d,p) level of theory) of the sw3-1 cluster.

<table>
<thead>
<tr>
<th></th>
<th>Atomic coordinates</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1.12144</td>
<td>-1.63045</td>
</tr>
<tr>
<td>C</td>
<td>0.67533</td>
<td>-0.44277</td>
</tr>
<tr>
<td>C</td>
<td>0.96544</td>
<td>0.79833</td>
</tr>
<tr>
<td>C</td>
<td>1.66991</td>
<td>0.81745</td>
</tr>
<tr>
<td>C</td>
<td>2.10838</td>
<td>-0.36477</td>
</tr>
<tr>
<td>C</td>
<td>1.84281</td>
<td>-1.58893</td>
</tr>
<tr>
<td>C</td>
<td>0.53857</td>
<td>2.06933</td>
</tr>
<tr>
<td>O</td>
<td>-0.78665</td>
<td>1.97053</td>
</tr>
<tr>
<td>O</td>
<td>-0.07556</td>
<td>-0.51955</td>
</tr>
<tr>
<td>H</td>
<td>0.88544</td>
<td>-2.57368</td>
</tr>
<tr>
<td>H</td>
<td>1.87187</td>
<td>1.77489</td>
</tr>
<tr>
<td>H</td>
<td>2.65569</td>
<td>-0.33145</td>
</tr>
<tr>
<td>H</td>
<td>2.18606</td>
<td>-2.51531</td>
</tr>
<tr>
<td>H</td>
<td>0.60767</td>
<td>2.91214</td>
</tr>
<tr>
<td>H</td>
<td>1.19339</td>
<td>2.27124</td>
</tr>
<tr>
<td>H</td>
<td>-1.38800</td>
<td>1.94092</td>
</tr>
<tr>
<td>H</td>
<td>-0.49954</td>
<td>0.37099</td>
</tr>
<tr>
<td>O</td>
<td>-1.99919</td>
<td>-2.18641</td>
</tr>
<tr>
<td>H</td>
<td>-2.84968</td>
<td>-2.32643</td>
</tr>
<tr>
<td>H</td>
<td>-1.45709</td>
<td>-1.67998</td>
</tr>
<tr>
<td>O</td>
<td>-1.55761</td>
<td>-0.80184</td>
</tr>
<tr>
<td>H</td>
<td>-0.60950</td>
<td>-0.93032</td>
</tr>
<tr>
<td>H</td>
<td>-1.75938</td>
<td>-1.35697</td>
</tr>
<tr>
<td>O</td>
<td>-1.99196</td>
<td>1.77162</td>
</tr>
<tr>
<td>H</td>
<td>-1.86768</td>
<td>0.84735</td>
</tr>
<tr>
<td>H</td>
<td>-2.84326</td>
<td>2.06442</td>
</tr>
</tbody>
</table>
### Table SI.8: Local mode Hamiltonian matrix of structure sw3-1: local mode frequencies are on the diagonal, couplings off-diagonal.

<table>
<thead>
<tr>
<th>Saligenin-(H₂O)₃-a - Model Hamiltonian matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>3333.6</td>
</tr>
<tr>
<td>-31.930</td>
</tr>
<tr>
<td>-0.050</td>
</tr>
<tr>
<td>-6.4200</td>
</tr>
<tr>
<td>-4.3900</td>
</tr>
<tr>
<td>-5.8300</td>
</tr>
<tr>
<td>-35.910</td>
</tr>
<tr>
<td>-12.180</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

### Table SI.9: Atomic coordinates of the optimized geometry (MP2/6-311++G(2d,p) level of theory) of the sw3-2 cluster.

<table>
<thead>
<tr>
<th>Saligenin-(H₂O)₃-b</th>
<th>N_{atoms} = 26</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-1.3971233</td>
</tr>
<tr>
<td>C</td>
<td>-0.3204962</td>
</tr>
<tr>
<td>C</td>
<td>-0.0136851</td>
</tr>
<tr>
<td>C</td>
<td>-0.8056616</td>
</tr>
<tr>
<td>C</td>
<td>-1.8864202</td>
</tr>
<tr>
<td>C</td>
<td>-2.173836</td>
</tr>
<tr>
<td>C</td>
<td>1.20654688</td>
</tr>
<tr>
<td>O</td>
<td>2.39424643</td>
</tr>
<tr>
<td>O</td>
<td>0.48993615</td>
</tr>
<tr>
<td>H</td>
<td>-1.6124239</td>
</tr>
<tr>
<td>H</td>
<td>-0.5616067</td>
</tr>
<tr>
<td>H</td>
<td>-2.4931195</td>
</tr>
<tr>
<td>H</td>
<td>-3.0073</td>
</tr>
<tr>
<td>H</td>
<td>1.3656505</td>
</tr>
<tr>
<td>H</td>
<td>1.09981725</td>
</tr>
<tr>
<td>H</td>
<td>2.28386502</td>
</tr>
<tr>
<td>H</td>
<td>0.3947894</td>
</tr>
<tr>
<td>O</td>
<td>1.85154281</td>
</tr>
<tr>
<td>H</td>
<td>1.86451959</td>
</tr>
<tr>
<td>H</td>
<td>1.8537355</td>
</tr>
<tr>
<td>O</td>
<td>1.82801282</td>
</tr>
<tr>
<td>H</td>
<td>2.15925519</td>
</tr>
<tr>
<td>H</td>
<td>0.91740576</td>
</tr>
<tr>
<td>H</td>
<td>0.38679542</td>
</tr>
<tr>
<td>H</td>
<td>0.833192126</td>
</tr>
<tr>
<td>H</td>
<td>0.91053806</td>
</tr>
</tbody>
</table>

### Table SI.10: Local mode Hamiltonian matrix of structure sw3-2: local mode frequencies are on the diagonal, couplings off-diagonal.

<table>
<thead>
<tr>
<th>Saligenin-(H₂O)₃-b - Model Hamiltonian matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>3580.0</td>
</tr>
<tr>
<td>-7.2800</td>
</tr>
<tr>
<td>-0.1700</td>
</tr>
<tr>
<td>-4.2400</td>
</tr>
<tr>
<td>-14.280</td>
</tr>
<tr>
<td>0.6500</td>
</tr>
<tr>
<td>-0.6400</td>
</tr>
<tr>
<td>-2.5200</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>
Chapter 9

Mapping vibrational modes in the far-infrared/Tera-Hertz domain.

This chapter presents a far infrared vibrational mode analysis of all systems investigated in this thesis. Our final aim is to build a general map of the motions that are responsible for the spectroscopic signatures recorded in the far infrared domain (<800 cm\(^{-1}\), <24 THz) as it exists in the more common 3000-4000 and 800-1800 cm\(^{-1}\) infrared ranges, see table 9.1. In practice, we try to find systematic spectral features and we observe the possible modifications in frequencies induced by changes in conformations. In this sense, this chapter follows the discussion from our paper "Mapping gas phase dipeptides motions in the far infrared and terahertz domain"\(^{65}\) presented in chapter 6 and the analyses have now been extended to several systems that I characterised during my PhD. We believe that such mapping might be useful for experimentalists, providing a tool to assign vibrational modes and to obtain insights on three dimensional structures without using theoretical methods. For an even more complete vibrational mode mapping, this chapter should be completed with supplementary molecular systems like water clusters\(^{96,97}\) or metal clusters\(^{98-104}\) for instance, measured by Asmis and Fielicke’s groups in the far infrared. These systems have not been included here. In our work, experiments are exploring the 90-800 cm\(^{-1}\) spectral region (3-24 THz), the simulations also provide the signatures below 90 cm\(^{-1}\). FELIX FEL is currently restricted to go down to \(\sim 90\) cm\(^{-1}\), a new source (FLARE) should be available within the next 2 years to allow probing wavelengths below 100 cm\(^{-1}\). This will be opening an exciting new avenue, where our theoretical predictions will be extremely useful. You will find a figure summarising our mapping at the end of this chapter, see figure 9.28.

In my work the infrared spectroscopy has been used as a tool for conformational assignment. Two examples can be found in chapters 8 and 11 for saligenin-water clusters and the (Ac-Phe-OMe)\(_2\) dimer respectively. Beyond conformational assignment, the identification of vibrational modes presented here is a prerequisite to identify strengths and weaknesses of theoretical methods (e.g. systematic deviations for certain vibrational modes) and to work on possible improvements, see chapter 10 for such discussion.

The discussion below will be on the assignment of motions in the vibrational modes ob-
Table of mid-infrared absorptions. Reproduced from the internet website: "https://webspectra.chem.ucla.edu/irtable.html"

<table>
<thead>
<tr>
<th>Functional Group</th>
<th>Characteristic Absorption(s) (cm⁻¹)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alkyl C-H Stretch</td>
<td>2950 - 2850 (m or s)</td>
<td>Alkane C-H bonds are fairly ubiquitous and therefore usually less useful in determining structure.</td>
</tr>
<tr>
<td>Alkenyl C-H Stretch</td>
<td>3100 - 3010 (m)</td>
<td></td>
</tr>
<tr>
<td>Alkenyl C=C Stretch</td>
<td>1680 - 1620 (v)</td>
<td></td>
</tr>
<tr>
<td>Alkynyl C-H Stretch</td>
<td>~3300 (s)</td>
<td>Absorption peaks above 3000 cm⁻¹ are frequently diagnostic of unsaturation</td>
</tr>
<tr>
<td>Alkynyl C=C Stretch</td>
<td>2260 - 2100 (v)</td>
<td></td>
</tr>
<tr>
<td>Aromatic C-H Stretch</td>
<td>~3030 (v)</td>
<td></td>
</tr>
<tr>
<td>Aromatic C-H Bending</td>
<td>860 - 680 (s)</td>
<td></td>
</tr>
<tr>
<td>Aromatic C=C Bending</td>
<td>1700 - 1500 (m,m)</td>
<td></td>
</tr>
<tr>
<td>Alcohol/Phenol O-H Stretch</td>
<td>3550 - 3200 (broad, s)</td>
<td></td>
</tr>
<tr>
<td>Carboxylic Acid O-H Stretch</td>
<td>3000 - 2500 (broad, v)</td>
<td>Primary amines produce two N-H stretch absorptions, secondary amides only one, and tertiary none.</td>
</tr>
<tr>
<td>Amine N-H Stretch</td>
<td>3500 - 3300 (m)</td>
<td></td>
</tr>
<tr>
<td>Nitrile C≡N Stretch</td>
<td>2260 - 2220 (m)</td>
<td>The carbonyl stretching absorption is one of the strongest IR absorptions, and is very useful in structure determination as one can determine both the number of carbonyl groups (assuming peaks do not overlap) but also an estimation of which types.</td>
</tr>
<tr>
<td>Aldehyde C=O Stretch</td>
<td>1740 - 1690 (s)</td>
<td></td>
</tr>
<tr>
<td>Ketone C=O Stretch</td>
<td>1750 - 1750 (s)</td>
<td></td>
</tr>
<tr>
<td>Ester C=O Stretch</td>
<td>1780 - 1710 (s)</td>
<td></td>
</tr>
<tr>
<td>Carboxylic Acid C=O Stretch</td>
<td>1690 - 1630 (s)</td>
<td></td>
</tr>
<tr>
<td>Amide C=O Stretch</td>
<td>3700 - 3500 (m)</td>
<td>As with amines, an amide produces zero to two N-H absorptions depending on its type.</td>
</tr>
</tbody>
</table>

Served in the far infrared/THz spectral range. To that end, DFT-MD trajectories are used for calculating infrared spectra by Fourier transforming the correlation function of the dipole moment of the gas phase molecule or cluster of interest (see chapter 4 for the description of the method, equation 4.62 in particular for infrared), and we analyse the vibrational motions by Fourier transform of internal coordinates named ICDOS (Internal Coordinate - vibrational Density Of States, equation 4.75 in chapter 4). We further refer the reader to chapter 10, where we discuss the quality of the DFT-MD spectra (by comparison with experimental spectra in terms of frequencies) for the systems presented here, which is crucial to be sure of our vibrational modes and conformational assignments (see also all discussions in chapters 5-8 of this thesis preceding this chapter).
Section 9.1 presents the ICDOS method employed here for unraveling vibrational motions. Strengths and weaknesses of the ICDOS will be discussed and also compared with the traditional harmonic normal modes (see section 4.5 in chapter 4 for their definition). These later were also used as guidelines in our work.

Section 9.2 presents the experimental isotopic substitution. This experimental tool is extremely powerful for vibrational mode assignments in the 2800-3800 cm$^{-1}$ range. We will show that this approach is less efficient, or at least more difficult to use in the far infrared/THz spectral domain.

In a nutshell, as presented in details in sections 9.3 and 9.4 of this chapter, we find two kinds of modes in the far infrared/THz domain, i.e. local modes and delocalised modes:

- We found 'local' modes (i.e. composed of only one major internal coordinate involved in the motion) to be out of plane/wagging motions of hydrogen atoms ($\omega$(CH), $\omega$(NH), $\omega$(OH)) and their rather localised character probably arises from the lightness of the hydrogen atom. We find that they are intense active infrared modes in the range 200-900 cm$^{-1}$, and a detailed mapping is presented in section 9.3.

- Section 9.4 presents a mapping of the delocalised modes (i.e. composed of a large number of internal coordinates) located in the whole far infrared range (below 800 cm$^{-1}$). These infrared modes are however less intense in the infrared spectrum than the local wagging modes, and they are really present only below 400 cm$^{-1}$, especially for peptides.

### 9.1 ICDOS: strengths and limitations.

Once an infrared spectrum has been extracted from DFT-MD simulations, one has to interpret the peaks in terms of vibrational motions for a complete understanding of structure-infrared fingerprints relationship for the investigated molecular system. In harmonic spectra calculations, this is traditionally done by harmonic modes. See chapter 4, section 4.5 for the method. This is well established and in a way we could have used these normal modes in order to interpret our anharmonic DFT-MD theoretical spectra. Calculated normal modes have the advantage that they include a direct knowledge of the internal coordinates involved in the vibrational motion, either visually (using the package MOLDEN for instance) or quantitatively (normal mode decomposition given by the Gaussian package directly). The % of participation of each internal coordinate into a given normal mode is hence known in the harmonic modes. The trouble behind using harmonic normal modes to interpret non harmonic peaks from DFT-MD is that we have no clue whether the harmonic representation is valuable because of the anharmonicities of the vibrational modes (in terms of composition of the motions or in terms of frequency of the modes). It is therefore logical to turn to theoretical tools typical of molecular dynamics trajectories in order to get modes assignments. Here we employ the ICDOS tool *(Internal Coordinate - vibrational Density Of States)*.

Chapter 5 introduced the ICDOS tool *(Internal Coordinates - vibrational Density Of States)*, based on the time dependent autocorrelation of internal coordinates. For a given internal
coordinate IC, for which the evolution in time is known from the DFT-MD trajectory \((IC(t))\), one can calculate the following associated spectral signature:

\[
I_{ICDOS}(\omega) = \int_{-\infty}^{\infty} \langle IC_i(t) \cdot IC_i(0) \rangle e^{i\omega t} dt
\]  

This is the Fourier transform of the autocorrelation function \(\langle IC_i(t) \cdot IC_i(0) \rangle\) of the chosen internal coordinate IC. The internal coordinate can be a bond length distance, angle or dihedral angle. See also chapter 4. A few remarks are important to make here as we will illustrate these points afterwards in this chapter.

In equation 9.1 we employ the autocorrelation function of the chosen internal coordinate. Any coupling of this coordinate with any other internal coordinate is not taken into account in the correlation (it is of course taken into account in the dynamics). This is of course a disadvantage in the analyses of the infrared spectra. We will see that such couplings can be obtained anyway, but they require the analyses of (many) separate ICDOS spectra to get that information. Nothing prevents us to calculate \(\langle IC_i(t) \cdot IC_j(0) \rangle\), the cross-correlation function between two internal coordinates, but one can easily see that this would amount to quite a lot of calculations and analyses to be done.

By calculating 9.1, one gets positions in frequency of the peaks of the chosen internal coordinate within the whole spectrum without a quantification of how much this particular internal coordinate participates to the corresponding infrared peak. What I mean is that we have no way to provide a % of participation of this particular internal coordinate into the final mode responsible for the infrared peak with expression 9.1. We will however see in this chapter that such disadvantage/drawback does not prevent a clear analysis of the vibrational motions.

One could think that by plotting the ICDOS spectra of all the non redundant internal coordinates of the system, one could obtain a quantitative view of the vibrational modes in the DFT-MD trajectory. First of all, in order to do that, one needs to define a complete set of non redundant internal coordinates, and this is not straightforward. More importantly the intensity of the peaks in any ICDOS spectrum is dependent on the energy in the vibrational mode. The equipartition of energy in all vibrational modes is not easy to reach in gas phase DFT-MD trajectories, and any deviation from equipartition has consequences on peaks intensities.

To go beyond these remarks would require to extract for instance effective normal modes\(^{200,201}\) from the trajectory, which has been done in the past in the group in Evry\(^{202,203}\). With that, a % of motions could indeed be given. Even though it sounds great, in practice the method is not so easy to apply and is also dependent from the equipartition in energy in the vibrational modes, difficult to reach in the gas phase. This method has not been applied here. D. Galimberti in the group tries to redevelop new codes with that aim.

One last remark. Expression 9.1 does not include selection rules for the infrared. As a consequence, the ICDOS spectrum does not contain the information on the infrared activity (or non activity) of the modes. This information is only contained in the infrared DFT-MD spectrum calculated using the Fourier transform dipole moment correlation function.

VDOS (velocity\(^{-}\) Vibrational Density Of States, obtained through the Fourier transform of the autocorrelation function of velocities) are more traditionally used in combination with molecular dynamics simulations\(^{52,186}\) in order to extract vibrational motions. We found that the
ICDOS are more useful to analyse vibrational modes in the far infrared/THz domain. VDOS indeed do not include any information about the nature of the motions, i.e. stretches, bends, torsions. Only the ICDOS does include this information through the internal coordinates and this knowledge is essential for collective/delocalised motions. Therefore VDOS will not be used in this chapter. However the ICDOS decomposition is not fully satisfactory either as already mentioned above and as it will be shown in this chapter.

We find two kinds of modes in the far infrared/THz domain, i.e. local modes (i.e. composed of only one major internal coordinate involved in the motion) for which ICDOS provide a clear view of the composition into motions, and delocalised modes (i.e. composed of a large number of internal coordinates) for which one would certainly like to be able to get a % of participation of each internal coordinate into each mode, but this will not be possible with the analyses done here using only equation 9.1. Despite this remark, as illustrated hereafter, one gets a good knowledge of the vibrational motions involved in the delocalised modes.

Using the phenol gas phase molecule as example (see the structure in figure 9.2), I illustrate below a few ideas related to the way we extract vibrational modes from theoretical DFT-MD spectra and some limitations intrinsic to the ICDOS method. We will therefore see below, spectral range by spectral range (or type of mode by type of mode), what we can learn about vibrational modes in the far infrared/THz domain using the ICDOS theoretical tool described above.

![Figure 9.2: 3D structure of the phenol molecule (optimised here at the BLYP-D3/6-311+G(d,p) level).](image)

### 9.1.1 Localised modes

We start our discussion with the 3000-4000 cm\(^{-1}\) range and the "OH stretching" mode in our pet-molecule phenol. The ICDOS of the OH bond length has been plotted in figure 9.3 and only one peak is found in the whole 0-4000 cm\(^{-1}\) range, which means that the mode is perfectly local. This stretching mode is typically one example of local modes for which ICDOS are well suited for unravelling.

To know whether the \(\nu(\text{OH})\) mode is perfectly local (i.e. 100% of participation of the OH motion into the stretching) is not really of interest here, what we want to know is that the
stretches internal coordinate dominates the mode. It is legitimate however to want to quantify
the remaining contributions that might couple with the OH stretching in that 3680 cm$^{-1}$ band.
To be honest, this question is relevant for all vibrational modes, but I will only illustrate this
for the OH stretching mode.

To answer this question we plot in figure 9.4 the ICDOS signatures of the OH covalent bond
length (in black), of one CH covalent bond length (in purple), of the CO covalent bond length
(in green) and of the COH angle (in red). These are the internal coordinates that we expect
the most easily coupled with the OH stretching. As can be seen, no signature is observed at
$\approx 3680$ cm$^{-1}$ which means that none of these extra internal coordinates is involved in this
mode. We do not expect other internal coordinates to be involved in the mode at $\approx 3680$ cm$^{-1}$
(other than the internal coordinates investigated here), we will therefore consider this band as
$\approx 100\%$ composed by $\nu$(OH) stretch.

As illustrated here and in the next section for the fingerprint region, one should calculate
and plot all possible ICDOS spectra for all the (non redundant) internal coordinates of the
molecule, and see (as in figure 9.4) if there are overlaps between ICDOS peaks of different
internal coordinates. If not, as for $\nu$(OH), the peak is assigned to 100% of the OH stretching.
Next section shows what is done when there are overlaps between ICDOS signatures of
different internal coordinates.
9.1. ICDOS: STRENGTHS AND LIMITATIONS.

Figure 9.4: Phenol ICDOS signatures of the OH covalent bond length (in black), one CH covalent bond length (in purple), the CO covalent bond length (in green) and the COH angle (in red).

9.1.2 Fingerprint range

We will now study the fingerprint region for which it is known and expected that vibrational modes are combinations of multiple internal coordinates, i.e. COH angle, CCH angles, CC stretches and CO stretches for the specific case of phenol. To illustrate that, we plot in figure 9.5 the ICDOS signatures of one CCH angle (in orange), of the COH angle (in red), of the CO covalent bond length (in green) and one of the CC stretching covalent bond length (in cyan). As we expected, we observe for each internal coordinate ICDOS spectrum multiple signatures between 800 and 1400 cm$^{-1}$ and we also observe common signatures inbetween these ICDOS spectra. See the peaks at 989, 1013, 1066, 1153, 1222, 1316 and 1338 cm$^{-1}$. The underlying modes correspond to our definition of delocalised modes, i.e. composed of a large number of internal coordinates which induce a spatial delocalisation of the mode over the molecule. By plotting all possible ICDOS spectra, we will find all internal coordinates involved in each mode. Hence the mode at 1338 cm$^{-1}$ is for example composed of (at least) COH bend, CCH bend and CC stretch. The one at 1153 cm$^{-1}$ is composed (at least) by all four motions used for the plot. This provides a qualitative interpretation of the vibrational peaks, but not a quantitative one, as we do not know the respective % of participation of each motion in the mode. This is not entirely true however. Judging the relative intensities of each ICDOS spectrum from each internal coordinate into a given mode, one can somehow quantify the participation of each internal coordinate into each vibrational mode (if we ignore problematics related to the equipartition of energy in the vibrational modes). Hence, the ICDOS signatures at 1338 cm$^{-1}$ has strong participations from $\delta$(COH) and $\delta$(CCH), while the one at 1153 cm$^{-1}$ strongly involves $\delta$(COH) and $\nu$(CO). This illustrates that the ICDOS are not particularly easy to interpret for delocalised
modes but they are definitely enough for our investigations.

![Figure 9.5: Phenol ICDOS signatures of one CCH angle (in orange), one COH angle (in red), the CO covalent bond length (in green) and one CC covalent bond length (in cyan).](image)

We now use the harmonic normal modes decomposition to have ideas of the possible internal coordinates involved for example in the mode at 1154 cm\(^{-1}\) (in the harmonic spectrum, 1153 cm\(^{-1}\) in the ICDOS spectrum). The harmonic mode is composed by 19.1% of the COH angle as main contribution, and up to a total of 25% of other contributions are from C-C and C-O stretchings, and the remaining contributions are mainly arising from CCH bendings. The harmonic mode and the ICDOS mode decomposition thus have the same qualitative interpretation. Furthermore, one sees that just knowing the main players in the mode is enough to get a good interpretation. ICDOS spectra decomposition is therefore enough for our purpose.

Let us now take the problem on the opposite way and focus on one single internal coordinate. We choose here the COH angle and its ICDOS spectrum is plotted in figure 9.5. We see that almost all the contributions of the COH angle internal coordinate are in vibrational modes located within the range 1000-1400 cm\(^{-1}\). While the harmonic normal modes decomposition gives the biggest contributions of \(\delta(\text{OH})\) by up to 19.1% in the mode at 1154 cm\(^{-1}\) and 10.2 % for the mode at 1332 cm\(^{-1}\), the ICDOS spectrum of this COH angle presented in figure 9.5 (in red) provides a peak at \(\approx 1335\) cm\(^{-1}\) twice more intense than the peak at \(\approx 1150\) cm\(^{-1}\). The ICDOS thus provides the same qualitative view as the harmonic normal modes, i.e. COH angle bending involved in the vibrational modes in this region but the % of participation in the different modes appears to be different between the two representations if we compare directly the relative intensities of the peaks of the ICDOS spectrum with the percentages given by the harmonic normal modes. At this stage, it is however not possible to distinguish
whether this is due to 1) DFT-MD inducing a different coupling between internal coordinates (or other anharmonicities) with respect to the harmonic approximation, 2) whether this is due to equipartition of energy not fully respected in all vibrational modes and thus altering band intensities (a combination of these two explanations is the most probable option).

9.1.3 Waggings

Already in the mid infrared range discussed above, we have seen the possible difficulty in using and interpreting ICDOS spectra for the "fingerprint" range modes involving several internal coordinates. It might be the same situation in the far infrared where delocalised modes are expected. However we (nicely) found some local modes in the far infrared for which the ICDOS method will be hence well suited. The wagging modes for example are found local. Figure 9.6 presents the ICDOS spectrum of the dihedral angle C-C-O-H of phenol, probing the $\omega$(OH) wagging (out of plane) motion. Only one signature is found around 330 cm$^{-1}$ in the whole 0-4000 cm$^{-1}$ range (except the very tiny peak around 500 cm$^{-1}$). I have only presented in figure 9.6 the 0-800 cm$^{-1}$ far infrared range. The decomposition of the associated normal mode gives 74% of participation from the $\omega$(OH) wagging in the mode at 305 cm$^{-1}$ (25 cm$^{-1}$ red shifted in position with respect to the DFT-MD representation) and participations below 10% of the CCOH dihedral internal coordinate can be seen in other modes in the far infrared. The remaining 26% in the 305 cm$^{-1}$ peak are composed of $\omega$(CH) wagging motions of the hydrogen atoms of the ring and out of plane motions of the ring. The interesting information for us is that the mode is local for both representations, i.e. largely dominated by $\omega$(OH). Interestingly, we observe similar vibrational modes for harmonic and anharmonic DFT-MD spectra despite a 25 cm$^{-1}$ shift in frequency.

9.1.4 Delocalised modes

For the other far infrared modes (not dominated by wagging motions of hydrogen atoms), we find collective/delocalised motions, as expected in this spectral range. The modes are collective in the sense that multiple internal coordinates are involved (as in figure 9.5 plotted in the mid infrared domain) and they are delocalised in the sense that the internal coordinates involved are spread over several parts of the molecule, inducing a spatial delocalisation of the mode. For instance, still focusing on phenol, according to the normal modes decomposition, the mode at 217 cm$^{-1}$ (lowest frequency in the infrared spectrum of phenol) is composed by $\approx$60% of motions corresponding to out of plane motions of the ring, $\approx$25% of motions corresponding to wagging motions of hydrogen atoms of the ring and $\approx$18% coming from the C-O wagging. 10 among the 33 internal coordinates of the system are involved in this mode (all the dihedral angles of the system are involved).The same global view is given by ICDOS (without the exact % decomposition), but one can immediately see that a large number of ICDOS spectra has to be analysed to get that picture. This is quite tedious. This is very well illustrated for Z-Ala$_6$-NH$_2$ in figure 9.26 later in this chapter.

One last remark. For wagging modes and any of the torsional modes, ICDOS spectral signatures are extracted using dihedral angles. In our choice of internal coordinates, a dihedral angle is simultaneously probing the out of plane motions of two different atoms. The OH
wagging motion in phenol is extracted for example using the dihedral angle H-O-C-C that probes the out of plane motion of the hydrogen atom (that we want to characterise) but also probes the out of plane motion of the carbon atom (at the extremity of the quadruplet of atoms). In practice, this is not a problem as displayed in figure 9.6 since the $\omega$(OH) signature is ultra dominant in the ICDOS due to the lightness of the hydrogen atom in comparison to the mass of the carbon atom. Figure 9.7 displays in black the same signatures as in figure 9.6, with a zooming scale. In purple are plotted the ICDOS signatures of the dihedral angle H-O-C$_{av}$-C$_{av}$ for which the average positions of the carbon atoms over the trajectory are now used. By averaging these positions (purple spectrum), the dihedral angle is not probing the out of plane motion of the carbon atom anymore and in the spectrum the main peak at 330 cm$^{-1}$ can be seen not affected, it overlaps with the black spectrum while the out of plane contributions from the terminal carbon atom are removed. Peaks located at $\approx$500, 680 and 740 cm$^{-1}$ totally disappear in the averaged positions spectrum. On the other hand the intensity of the peak at $\approx$ 220 cm$^{-1}$ increases in the averaged position spectrum.

9.1.5 Conclusion

We have seen (briefly) that the ICDOS method is a good tool for analysing local modes but it is more tedious to use to unravel all motions into a given vibrational peak as soon as multiple internal coordinates are involved, typically for the collective/delocalised modes. In this case, the harmonic normal modes might somehow appear as better suited even if we can not fully trust the composition of these modes (as well as the band positions in some cases, see
9.1. **ICDOS: STRENGTHS AND LIMITATIONS.**

Figure 9.7: In black, ICDOS signatures of the C-C-O-H dihedral angle internal coordinate for the phenol molecule. In purple, ICDOS signatures of the dihedral angle $C_{av}-C_{av}-O-H$ for which the average positions of the carbon atoms over the trajectory are used.

the related discussion on this issue in chapter 10). The differences in the modes between a harmonic calculation and ICDOS/DFT-MD calculation can be due to the intrinsic different internal coordinate combinations when anharmonicities are taken into account. Differences can also arise from the bad equipartition of energy in the modes in the DFT-MD trajectory, which is tricky to achieve for the gas phase. Anyway, we can not provide the quantitative composition of vibrational modes (in terms of percentage composition of each internal coordinate involved in the mode) from ICDOS spectra, but as illustrated above it is not an issue. As soon as delocalised modes are found in the far infrared, they involve several internal coordinates motions, and knowing the exact (small) % of participation of each internal coordinate does not give any substantial information. The qualitative information given by ICDOS is enough for interpretation. For the vibrational modes mapping presented in sections 9.3 and 9.4, only ICDOS will be presented. Harmonic normal modes have been used sometimes as guidelines but the final interpretation of the vibrational peaks is done by ICDOS calculations only.
9.2 Isotopic substitution in experiments

I have spoken above only on the theoretical analysis tools that I can apply in order to assign infrared bands to vibrational motions. There is of course one purely experimental one well-known method to get that information, i.e. the isotopic substitution. Let me explain it briefly and let us see whether this is a good and reliable method to apply in the far infrared/THz domain. To illustrate how isotopic substitution works, let us take the example of a diatomic molecule. The frequency $\nu$ of the single vibrational mode of a diatomic molecule can be calculated by:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (9.2)$$

with $k$ the force constant, defined as:

$$k = \frac{\partial^2 E}{\partial^2 Q_i} \quad (9.3)$$

with $E$ the ground state electronic energy of the molecule and $Q_i$ the normal mode $i$ (here stretching between the two atoms of the diatomic), $\mu$ is the reduced mass of the associated normal mode, here for a diatomic stretching:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (9.4)$$

with $m_1$ and $m_2$ the respective mass of the two atoms in the diatomic.

One way to calculate the frequency of a mode in a system involving more than two atoms is presented in section 4.5 of chapter 4. However, this simple model above allows one to understand the idea of the isotopic substitution for the vibrational spectroscopy.

The frequency of a vibrational mode thus depends on two things: the potential energy surface $E$ and the reduced mass of the mode $\mu$. It is possible to modify the mass of one atom and therefore the reduced mass of the whole system without changing the potential energy surface by using isotopic substitution (e.g. H is replaced by D, $^{17}$O by $^{18}$O, ...).

As an illustration, in the paper by Stearns et al.$^{90}$ isotopic substitution has been used to spectroscopically characterise the Ac-Phe-(Ala)$_5$-Lys-H$^+$ peptidic system. Nitrogen atoms are substituted by their $^{15}$N isotope to assign, without ambiguity, the signature of the corresponding N-H stretching and therefore to have insight about their local environment. Figure 9.8 has been copied from that paper (“Spectroscopy and conformational preferences of gas-phase helices”$^{90}$) to illustrate the effect of the isotopic substitution on an infrared spectrum in the 3000-4000 cm$^{-1}$ stretching domain.

For the four isomers (A-D), the nitrogen atoms of the second, fourth and sixth amino acids over the chain have been substituted and three infrared spectra have been measured for each isotopologue. For each of them, one single peak is shifted in frequency which means that one peak (i.e. one vibrational mode) corresponds to one NH stretching only. The association be-
between one mode and one NH function is therefore possible and directly seen by comparing the infrared peaks of the spectrum of the deuterated and non deuterated species (here 3000-4000 cm\(^{-1}\) range).

![Figure 9.8: Infrared-ultraviolet double resonance spectra of isotopically substituted Ac-Phe-(Ala)\(_5\)-Lys-H\(^+\). For each conformer, the unsubstituted peptide spectrum is shown at the top, with the singly-substituted peptides shown underneath. The shifted peaks are circled, while a dashed line shows the position of those transitions in the unsubstituted peptide. This figure has been copied from the article: "Spectroscopy and conformational preferences of gas-phase helices"\(^{90}\).]

The \(^{15}\)N substitution in a peptidic system or the \(^{17}\)O substitution in systems as phenol derivatives is indeed a useful approach for the N-H or O-H stretching motions because of the local character of these modes. No other normal modes involving the \(^{15}\)N or \(^{17}\)O substituted atom can be found in the 3000-4000 cm\(^{-1}\) range, the stretching modes are generally only slightly coupled, and therefore, as seen above, each substitution induces one single shift of the related absorption peak in the spectrum of the isotopologue. For example, for the phenol molecule (structure presented in figure 9.2), the OH covalent bond length internal coordinate is involved in only one normal mode, \(i.e\). the OH stretching (see subsection 9.1.1). Any \(^{18}\)O substitution (also true for deuteration) of this OH would therefore provide a shift in the peak of the O-H/O-D stretching that would unambiguously allow its identification in the whole spectrum.

It is supposedly more complicated for the far infrared range, where delocalised modes are now expected. Several nitrogen or oxygen atoms could be involved in the same normal mode (the nitrogens of the backbone in a peptidic system for example). Therefore, the isotopic substitution is expected to change the reduced mass of more than one normal mode and therefore to shift multiple absorption peaks. The amount of information provided would drastically
decrease. For example, the normal mode decomposition mentioned in section 9.1 indicates that the internal coordinates involving the oxygen atom of phenol (OH bond length, CO bond length, two angles and four dihedral angles) have participation, among the 11 normal modes present in the range 0-800 cm\(^{-1}\), up to 23\%, 19\% and 10.5\% for 3 normal modes and between 5\% and 10\% for 5 other normal modes (note that the redundancies in the internal coordinates have been only roughly removed here). We thus expect major modifications in the isotopologue spectrum for three features and minor modifications for five features.

At the same time, the internal coordinates involving the hydrogen atom of the OH function of phenol (OH covalent bond length, the COH angle and the two redundant C-C-O-H dihedral angles), to keep this example, are mainly involved in one normal mode only in the far infrared (at 309 cm\(^{-1}\)), up to a total of 74\%. Four other modes have a participation between 5\% and 10\%. In comparison with a \(^{18}\)O substitution, the deuteration of this hydrogen atom will subsequently have much less impact on the deuterated far infrared spectrum using the phenol-OH far infrared spectrum as reference.

In our paper: "Anharmonic, dynamic and functional level in far-infrared spectroscopy: phenol derivatives"\(^{108}\), presented in chapter 7, we therefore chose the \(^2\)H substitution/deuteration of the hydrogen atom for the phenol molecule instead of the \(^{17}\)O substitution to extract information on vibrational modes in going from phenol-OH to its phenol-OD isotopologue.

Figure 9.9: Infrared-ultraviolet double resonance spectra of phenol and isotopically substituted (deuterated) phenol. For deuterated phenol, the hydrogen atom of the OH function has been replaced by OD. The position of each absorption peak and their vibrational assignment are labelled above each signature. The signatures arising from the \(\nu_0 \rightarrow \nu_1\) OH or OD wagging transition are highlighted in blue, the signature of the associated \(\nu_0 \rightarrow \nu_2\) overtones are highlighted in yellow. These experimental data are presented in our paper: "Anharmonic, dynamic and functional level effects in far-infrared spectroscopy: phenol derivatives"\(^{108}\)

Figure 9.9 presents a comparison between the experimental spectra of phenol and its deuterated isotopologue in the far infrared. For deuterated phenol, the hydrogen atom of the
OH function has been replaced by OD. Several differences can be directly noticed: the peaks located at 309 and 588 cm\(^{-1}\) in the phenol-OH spectrum vanish in the phenol-OD spectrum, while the 226 cm\(^{-1}\) Phenol-OH peak is replaced by two more intense peaks located at 213 and 247 cm\(^{-1}\) in the phenol-OD spectrum. Two small amplitude peaks at 419 and 446 cm\(^{-1}\) are also appearing in the spectrum of the deuterated molecule. Also, a small frequency shift can be observed for the peak at 400 cm\(^{-1}\) (Phenol-OH), now located at 380 cm\(^{-1}\) (Phenol-OD). As a last comment, the spectrum of phenol-OH is much more modified by the OD deuteration, here < 800 cm\(^{-1}\) than in the range 3000-4000 cm\(^{-1}\) for the experiment presented in figure 9.8.

To extract some information and to understand the modifications in the spectra arising from deuteration, theoretical analyses have been performed (harmonic, MD and VPT2 representations presented in chapter 4, all spectra are discussed in chapter 10). The assignment of the modes is reported in figure 9.9 next to each feature. It shows that the OD torsion couples with an out of plane deformation of the ring (located at 226 cm\(^{-1}\) in phenol-OH) to produce the two peaks at 213 and 247 cm\(^{-1}\) in the phenol-OD spectrum. We know that the hydrogen atom has participation in the transition at 588 cm\(^{-1}\) which is due to its position (in terms of frequency) in the phenol OH spectrum can be most probably assigned to the \(\nu_0 \rightarrow \nu_2\) overtone of the OH torsion (309 cm\(^{-1}\)). While the 588 cm\(^{-1}\) peak naturally disappears in the phenol-OD spectrum, at the same time two \(\nu_0 \rightarrow \nu_2\) overtones of the two modes combining OD torsion and out of plane ring deformations appear in the deuterated phenol spectrum at 419 and 446 cm\(^{-1}\).

The transition associated to the CCO in plane bending (400 cm\(^{-1}\) for the OH phenol) mode is also red shifted by 20 cm\(^{-1}\) after deuteration due to the increase in the mass of the OH function.

As a conclusion, I have illustrated that even for this small system, the amount of information that one can extract from an isotopic substitution experiment in the far infrared/THz is less clear than in the 3000-4000 cm\(^{-1}\). We said earlier that we expected less information from isotopologues spectra in the far infrared range because of the collective character of the far infrared modes. This is not directly true here since we have seen that the OH wagging mode remains rather local. The spectral differences between phenol-OH and phenol-OD spectra can be explained in a somewhat different way. After deuteration, the larger weight of the deuterium atom makes the coupling with other internal coordinates much easier and we thus observe different normal modes for the deuterated species (which does not simplify our explanation/understanding of the spectroscopy of phenol). One has therefore to be cautious with deuteration once applied to far-infrared spectroscopy.
CHAPTER 9. VIBRATIONAL MODE MAPPING

9.3 Mapping local wagging modes

We now apply our ICDOS toolkit to unravel vibrational modes in the far infrared/THz spectral domain. In this section, we talk about the local wagging modes active in the far infrared/THz, which can be easily assigned through the ICDOS method discussed above. A mapping of the delocalised vibrational modes will be presented in section 9.4. The local modes are out of plane motions (or waggings) of hydrogen atoms and their rather localised nature probably arises from the lightness of these atoms. ICDOS spectra of dihedral angles are well suited to unravel these motions, as seen in section 9.2, because one internal coordinate has one major contribution into the wagging. The assignments of local wagging motions, i.e. $\omega$(N-H), $\omega$(C-H), $\omega$(O-H) or the hindered rotational motions of CH$_3$ groups, are presented below using all systems investigated in this thesis. We refer to chapter 2 for the presentation of these systems. These modes are located below 900 cm$^{-1}$. The rather large amplitude of these motions induce a large variation of the dipole moment. Hence these modes have the strongest infrared intensities in the far infrared and therefore dominate the spectral range. We will see in section 9.4 that delocalised modes may overlap with wagging modes, but they are less infrared intense.

9.3.1 C-H phenyl ring waggings (690-760 cm$^{-1}$)

CH groups can be present in two locations in natural peptidic systems, i.e. the backbone C$_\alpha$-H$_\alpha$ group (see figures 2.1 and 2.2 in chapter 2) and in the side chain residue. CH groups can be also found in the terminal cap for non natural peptides, for example in Z-Ala$_6$-NH$_2$, see figure 2.11 in chapter 2. The C$_\alpha$-H$_\alpha$ motions systematically mix with other internal coordinates to produce delocalised/collective modes as described in section 9.4. This is also true for the CH groups in the side chain at the exception of CH in two specific chemical groups, i.e. CH$_3$ functions and phenyl rings. CH$_3$ motions are discussed in subsection 9.3.5 while we will here describe $\omega$CH motions of the phenyl ring.

Due to the need of a chromophore in our systems (for the absorption of UV photons in the IR-UV ion dip scheme, see section 3.2 in chapter 3), most of the systems studied here carry a phenyl ring:
- i) it is the phenylalanine for the Ac-Phe-'AA'-NH$_2$ dipeptide series and in the Ac-Phe-OMe system, see sections 2.1.1 and 2.1.2 of chapter 2,
- ii) it is the Z-cap for the Z-(Ala)$_6$-NH$_2$ peptide, see section 2.1.3 of chapter 2,
- iii) of course, the phenyl ring is the principle constituent of the phenol derivatives, see section 2.2 of chapter 2.

We found that the $\omega$CH motions of the hydrogen atoms of these phenyl rings couple to produce two kinds of wagging modes, i.e. in phase modes for which the variation of the dipole moment of each CH group add up, providing intense modes in the far infrared spectra, and out of phase modes for which variations of the dipole moment of the CH groups compensate providing modes weakly intense, found active only if coupled with other internal coordinates. The following assignment is supported by the ICDOS analyses presented in figures 9.10, 9.11 and 9.12, where ICDOS spectra of the dihedral angles labelled as $\Phi_{123} = $ CCCH$_{123}$, are plotted. See figures 2.6, 2.10 and 2.16 in chapter 2 for their definitions, the same figure is not pre-
Let us summarise here the main results before going into more details. We observe the same pattern of signatures for all the monosubstituted phenyl rings (see figures 9.10, 9.11 and 9.12) with in phase motions (intense infrared modes) at $\approx 700$, $\approx 730$ and $\approx 750$ cm$^{-1}$ and out of phase modes at $\approx 400$, $\approx 840$ and $\approx 900$ cm$^{-1}$ (signatures with no or only small infrared intensities). Therefore, the signatures of the $\omega$(CH) wagging motions of the hydrogen atoms that belong to the phenyl ring remain similar from one system to another and therefore are not really sensitive to the environment. In other words, these modes can not be used for conformational assignments.

We have two cases for which the phenyl ring is not mono-substituted anymore but carries two groups, i.e. catechol and saligenin (figure 9.12). Strong spectral differences are actually expected because of the different substitution of the aromatic ring. According to the ICDOS spectra, the $\omega$(CH) signatures are indeed strongly modified in terms of frequencies with respect to the mono substituted aromatic rings. For catechol, we find two $\omega$(CH) in the experimental spectrum at 736 and 442 cm$^{-1}$. Both correspond to in phase motion and for the band at 442 cm$^{-1}$, the $\omega$(CH) motion is coupled with $\omega$(OH) motion of the bonded OH function. The peak at 736 cm$^{-1}$ is very intense while the peak at 442 cm$^{-1}$ is barely visible in the experimental spectrum. The experimental spectrum has not been measured above 800 cm$^{-1}$ but we observe in the ICDOS spectrum two contributions that correspond to out of phase $\omega$(CH) motions at $\approx 820$ and $900$ cm$^{-1}$ with small counterpart in the infrared DFT-MD spectrum. We find three $\omega$(CH) features in the experimental spectrum of saligenin, two intense features at 727 and 756 cm$^{-1}$ that correspond to in phase motions and one peak barely visible at 540 cm$^{-1}$ that corresponds to out of the phase motions.

Let us now go into more details for every single system starting with the dipeptides series.

In the Ac-Phe-AA-NH$_2$ series, figure 9.10, three absorption bands for $\omega$(CH) are systematically found in the 700-800 cm$^{-1}$ spectral domain in both experimental and DFT-MD infrared spectra, located around 700, 730 and 750 cm$^{-1}$ (A similar figure can be found in the paper: "Mapping gas phase dipeptides motions in the far-infrared and terahertz domain."$^{65}$ presented in chapter 6). These bands arise from the in phase waggings of the hydrogen atoms that belong to the phenylalanine ring, as shown with the decomposition into ICDOS.

Other features involving $\omega$(CH) are systematically present in the ICDOS spectra around 400, 840 and 900 cm$^{-1}$ but with only small counterparts (small activity) in the DFT-MD infrared spectrum or in the experimental spectrum. These features correspond to out of the phase $\omega$(CH) waggings that have small or no infrared activities since there are small or no dipole moment variations induced by these motions.

All the dipeptides systems have an "a" orientation for the phenyl ring. The "a" orientation corresponds to the dihedral angle $\chi_1=180^\circ$ (see figures 2.6 and 2.7 in chapter 2 for illustrations). Despite the same spectral signatures for the aromatic C-H wagging motions, the phenylalanine aromatic ring can be found in three different environments within the dipeptides of interest here:
Figure 9.10: Unraveling $\omega$(CH) wagging modes for the dipeptide series: from top to bottom, infrared experimental spectra (in color), theoretical DFT-MD infrared spectra (in black) and ICDOS signatures of the dihedral angles $\Phi_{123} = \text{CCCH}_{123}$ of the phenyl ring. See chapter 2 for detailed descriptions of all these systems.
• interacting with the amide N-H group of the 'AA' amino acid in the \( \gamma \)-turn geometries of Ac-Phe-Gly-NH\(_2\), Ac-Phe-Ala-NH\(_2\), Ac-Phe-Cys-NH\(_2\) (\( \gamma \)-turn), Ac-Phe-Ser-NH\(_2\), Ac-Phe-Val-NH\(_2\), (fig. 2.7-a of chapter 2);

• in the \( \beta \)-turn conformation of Ac-Phe-Cys-NH\(_2\), this interaction is weaker (fig. 2.7-c of chapter 2);

• it interacts with the proline ring via a \( \pi \)-interaction in the \( \gamma \)- and \( \beta \)-turn conformations of Ac-Phe-Pro-NH\(_2\) peptide (fig. 2.7-b and d of chapter 2).

This shows that \( \omega \)(CH) modes are non sensitive to the environment details, and therefore are not conformer selective.

In figure 9.11 are now presented the ICDOS signatures of the dihedral angles \( \Phi=CCCH_{123} \) for the Ac-Phe-OMe monomer, the (Ac-Phe-OMe)\(_2\) dimer and Z(Ala)\(_6\)-NH\(_2\). See chapter 2 for more details on these systems. The definitions of the dihedral angles remain consistent with the Ac-Phe-AA-NH\(_2\) dipeptide series. The phenyl ring also has different environments in these systems:

• In the Ac-Phe-OMe monomer, the ring adopts a "g⁺" orientation that corresponds to the dihedral angle \( \chi_1=60^\circ \) (see figures 2.9 and 2.10 in chapter 2 for illustrations) and is totally free of interaction, see figures 2.9 and 2.10 in chapter 2, and we observe small differences in the infrared experimental spectrum in comparison with the dipeptide series for example. Instead of the three peaks located at \( \approx 700, 730 \) and \( 750 \) cm\(^{-1}\) with intensity ratios of 3:1:2 in the dipeptide series, here, we only have two intense experimental signatures at \( \approx 700 \) and \( 740 \) cm\(^{-1}\). A new feature also appears at \( \approx 790 \) cm\(^{-1}\) in the ICDOS spectrum but it is not active in the experimental infrared spectrum. This small displacement is the biggest deviation from the general rule described above.

What may be the origin of the 790 cm\(^{-1}\) band, blueshifted from the 750 cm\(^{-1}\) in the dipeptides? We have two hypotheses here:

- \( i \) The aromatic ring is totally free of interaction in Ac-Phe-OMe, which is never the case for the dipeptide series.

- \( ii \) All the dipeptides systems have an "a" orientation for the phenyl ring. The "a" orientation corresponds to the dihedral angle \( \chi_1=180^\circ \) (see figures 2.6 and 2.7 in chapter 2 for illustrations). In the Ac-Phe-OMe system, the phenyl ring adopts a "g⁺" orientation that corresponds to the dihedral angle \( \chi_1=60^\circ \) (see figures 2.9 and 2.10 in chapter 2 for illustrations).

Both these hypotheses are invalidated by the example of the (Ac-Phe-OMe)\(_2\) dimer discussed below.

• In the (Ac-Phe-OMe)\(_2\) dimer, the phenyl ring of one strand adopts a "g⁺" orientation that corresponds to the dihedral angle \( \chi_1=60^\circ \) while the phenyl ring of the second strand adopts a "g⁻" orientation that corresponds to the dihedral angle \( \chi_1=-60^\circ \) (see figures 2.9 and 2.10 in chapter 2 for illustrations). The ring of the \( \beta_L(g^-) \) strand is free of interaction as in Ac-Phe-OMe while the ring of the \( \beta_L(g^+) \) strand (same orientation as in the monomer) is in interaction with the CH\(_2\) function belonging to the other strand.
Figure 9.11: Unraveling $\omega$(CH) wagging modes for the Ac-Phe-OMe monomer and dimer, and the Z-Ala$_6$-NH$_2$ peptide: from top to bottom, infrared experimental spectra (in color), theoretical DFT-MD infrared spectra (in black) and ICDOS signatures of the dihedral angles $\Phi_{123}$ = CCCH$_{123}$ of the phenyl ring. See chapter 2 for detailed descriptions of all these systems.
According to the ICDOS decomposition, the free ring in the $\beta_L(g^-)$ strand provides the same $\omega(CH)$ signatures as the rings in the dipeptide series that are engaged in $\pi$ interactions ($\approx 400, \approx 700, \approx 730, \approx 750, \approx 840$ and $\approx 900 \text{ cm}^{-1}$), while the interacting ring (with the CH$_2$ function) in the $\beta_L(g^+)$ strand provides the same signatures as in the monomer in which the ring is free ($\approx 400, \approx 700, \approx 740, \approx 790, \approx 830$ and $\approx 900 \text{ cm}^{-1}$) as described above.

We see that the "$g^+$" orientation for phenyl rings does not induce change of vibrational signatures and we see that a ring totally free of interaction can provide identical signatures to ring engaged in $\pi$ interactions.

- In Z(Ala)$_6$-NH$_2$, the aromatic ring is fixed at the extremity of the peptidic chain and stays in interaction with a CH$_3$ residue. Since the hydrogen and carbon atoms have equivalent electronegativities, the CH$_3$ moiety is not expected to engage in strong interactions with other moieties. Nevertheless the CH$_3$-$\pi$ interaction is never broken over the whole trajectory. The $\omega(CH)$ signatures adopt the same patterns as the ones observed for the dipeptide series, i.e. $\approx 400, \approx 700, \approx 730, \approx 750, \approx 840$ and $\approx 900 \text{ cm}^{-1}$, with intense infrared modes (in phase motions) at $\approx 700, \approx 730$ and $\approx 750 \text{ cm}^{-1}$.

The observations on the Ac-Phe-OMe systems and on Z-Ala$_6$-NH$_2$ confirm our first conclusions from the dipeptide series: the $\omega(CH)$ modes are not sensitive to the environment and are not conformer selective. Minor changes of the spectral features are more probably related to the contributions from other internal coordinates involved in the modes.

Figure 9.12 now presents the ICDOS signatures of the dihedral angles $\Phi=CCCH_{23456}$ for four phenol derivatives. The structures are presented in figure 2.15 in chapter 2 and the definitions of the dihedral angles are presented in figure 2.16.

For phenol, the intense infrared modes including $\omega(CH)$ contributions are located at $\approx 500, 685$ and $755 \text{ cm}^{-1}$ in the experimental spectrum. The experimental spectrum has not been measured above $800 \text{ cm}^{-1}$ but we find less intense $\omega(CH)$ modes at $\approx 800$ and $870 \text{ cm}^{-1}$ (according to the ICDOS and DFT-MD spectra). We find the same patterns as for all other systems even though some differences are notable, i.e. a signature is located at $500 \text{ cm}^{-1}$ instead of the $400 \text{ cm}^{-1}$ observed in the peptides described above, there is no peak at $\approx 730 \text{ cm}^{-1}$ and the peaks at $840$ and $900 \text{ cm}^{-1}$ in the dipeptide series are redshifted by $40$ and $30 \text{ cm}^{-1}$ respectively and are now located at $800$ and $870 \text{ cm}^{-1}$ in the phenol experimental spectrum.

For the phenol-water cluster, the signatures from $\omega(CH)$ are almost identical to phenol, suggesting that the addition of the water molecule has no influence on these modes.

We observe really different signatures for catechol and saligenin, as described above, mostly probably because of the different substitution of the phenyl ring.
Figure 9.12: Unraveling $\omega$(CH) wagging modes for the phenol derivatives: from top to bottom, infrared experimental spectra (in red), theoretical DFT-MD infrared spectra (in black) and ICDOS signatures of the dihedral angles $\Phi_{123} = \text{CCCH}_{123}$ of the phenyl ring. See chapter 2 for detailed descriptions of all these systems.
9.3.2 N-H waggings (400-900 cm\(^{-1}\))

\(\omega(NH)\) waggings are found with infrared signatures in the range 400-900 cm\(^{-1}\) for the systems investigated here.

Figure 9.13 presents \(\omega(NH)\) signatures for the dipeptide series. The two dihedral angles used to extract these wagging signatures, \(H_{Phe}\)\(-\)\(N_{Phe}\)\(-C_{\alpha}\)\(-C\) and \(H_{AA}\)\(-\)\(N_{AA}\)\(-C_{\alpha}\)\(-C\), are presented in figure 2.6 in chapter 2. Two separate Amide V wagging signatures are obtained for the two backbone N-H amide groups. They are labelled \(\omega(N - H)_{Phe}\) and \(\omega(N - H)_{AA}\) in the following, related to the phenylalanine amide and 'AA' residue amide respectively.

\(\omega(N - H)_{Phe}\) is located between 470 and 510 cm\(^{-1}\), as highlighted by the green box in Figure 9.13. This specific amide group has a common signature within a 40 cm\(^{-1}\) interval for all Ac-Phe-AA-NH\(_2\) dipeptides, which is not surprising as this N-H\(_{Phe}\) is always involved in a weak C5 interaction with C=O\(_{Phe}\) (see illustrations of \(\gamma\)- and \(\beta\)-turn structures in figure 2.7 in chapter 2). The 470-510 cm\(^{-1}\) region can therefore be used as a reference for \(\omega(N - H)\) wagging signatures of N-H amide groups engaged in a C5 backbone interaction.

When this N-H\(_{Phe}\) is free of interaction as is found in the \(\beta\)-turn of the Ac-Phe-Cys-NH\(_2\) (only exception in the series, see Figure 2.7), \(\omega(N - H)_{Phe}\) is found around 410-490 cm\(^{-1}\). This certainly provides an absolute reference for \(\omega(N - H)\) of backbone free N-H groups.

What we observe here is that NH engaged into interactions has vibrational signatures blue-shifted from the \(\omega(NH)\) wagging mode of the free NH. We observe the opposite in the 3000-4000 cm\(^{-1}\) domain where the NH stretching modes are red shifted once engaged into interactions.

\(\omega(N - H)_{AA}\) displays various signatures between 400 and 640 cm\(^{-1}\) depending on the chemical composition of the residue and interactions of this backbone N-H with its surrounding. For Ac-Phe-Gly-NH\(_2\), Ac-Phe-Ala-NH\(_2\), and the two Ac-Phe-Val-NH\(_2\) \(\gamma\)-turn structures, there is a \(\pi\)-interaction between N-H\(_{AA}\) and the Phe aromatic ring, see figure 2.7-a in chapter 2. The associated wagging signatures are therefore located between 530 and 550 cm\(^{-1}\) for these dipeptides, as illustrated by the orange box in figure 9.13. This represents up to a maximum of \(\approx 90\) cm\(^{-1}\) blue-shift from the free \(\omega(NH)_{Phe}\) wagging signature characterised above. For the Ac-Phe-Ser-NH\(_2\) and Ac-Phe-Cys-NH\(_2\) \(\gamma\)-turns, the same weak Phenylalanine/N-H\(_{AA}\) interaction is present. However, due to the different chemical nature of the side chain containing OH or SH groups respectively, a supplementary blue shift of \(\omega(N - H)\) is observable, \(\omega(NH)\) being now located at 555 and 620 cm\(^{-1}\) for Ac-Phe-Ser-NH\(_2\) and Ac-Phe-Cys-NH\(_2\) respectively. See the corresponding yellow and violet boxes in figure 9.13. In the \(\beta\)-turn conformation of Ac-Phe-Cys-NH\(_2\), N-H\(_{AA}\) is only weakly interacting with the aromatic ring (Figure 2.7), and the corresponding wagging \(\omega(N - H)_{AA}\) is found around 565 cm\(^{-1}\). In Ac-Phe-Pro-NH\(_2\), there is no N-H\(_{AA}\), since the hydrogen atom is replaced by the proline ring.

As seen above, the Amide V mode is found to be diagnosis for the local environment around the N-H backbone group, and can be used as a valuable tool for conformational assignment of gas phase peptides providing information equivalent to the 3000-4000 cm\(^{-1}\) NH stretching
Figure 9.13: Unraveling $\omega$(NH) wagging modes of the dipeptide series. From top to bottom, infrared experimental spectrum (in color), theoretical DFT-MD infrared spectrum (in black) and ICDOS signatures of the $\omega$(NH). For structures and labels, we refer the reader to figures 2.6 and 2.7 in chapter 2.
9.3. MAPPING LOCAL WAGGING MODES

range. Furthermore, we have just shown that each backbone amide N-H group gives rise to one distinct wagging signature in the 400-700 cm\(^{-1}\) domain, hence reflecting backbone interactions with the surrounding, just as seen in the 3000-4000 cm\(^{-1}\) domain.\(^{61,90}\)

In figure 9.14 are displayed the ICDOS spectra of the NH waggings for Z-Ala\(_6\)-NH\(_2\). Six different NH functions are present in the peptide for the six amino acids, see figure 2.12 in chapter 2. The first striking point is that the ICDOS spectra signatures are much broader than in the dipeptide series spectra (although same temperature in the dynamics). It seems also that some NH waggings are coupled since some features are shared by multiple NH waggings. For example, the feature at \(\simeq 620\) cm\(^{-1}\) is shared by the waggings of the NH functions number 2, 5 and 6 and the feature at \(\simeq 660\) cm\(^{-1}\) is shared by the waggings of the NH functions number 2 and 5. These modes remain highly dominated by wagging motions.

One of the aims of the study of the peptidic systems presented here is to find specific signatures for each motif commonly observed in proteins, i.e. C5 in \(\beta\) strands, C7 in \(\gamma\)-turns, C10 in \(\beta\)-turns and 3\(_{10}\), C13 in \(\alpha\)-helices and \(\beta\)-sheets (see chapter 2 for more insight about proteins organisation). For the dipeptide series presented above, one NH (Phe) is systematically engaged into a C5 interaction, providing a common signature between all peptides in the series, while the second NH (AA) function provides diverse signatures related to diverse external interactions and internal chemical environments. We might obtain various other signatures by investigating larger systems and in this context Z-Ala\(_6\)-NH\(_2\) system would seem particularly relevant. But this system adopts a gas phase globular conformation, not typical of the structures found for larger proteins and peptides. This system however adopts motifs of the type C7, C10, C17 and C20, which signatures can be analysed. One surprising result is that the hydrogen bond motifs in Z-Ala\(_6\)-NH\(_2\) are more complex than the `classic' motifs described above.

Let us see that in details:

- The NH functions of the first and second amino acids in Z-Ala\(_6\)-NH\(_2\) are linked to the same C=O function (of the last amino acid, see figure 2.12) forming respectively a C20 and a C17 hydrogen bond interaction (corresponding to a 20 and 17 hydrogen bonded membered ring). Signatures of \(\omega(NH)_{1}\) are found between 530 and 600 cm\(^{-1}\) and signatures of \(\omega(NH)_{2}\) are found between 580 and 670 cm\(^{-1}\). These C17 and C20 patterns can not however be fully compared to a C17 or C20 interaction where only one single NH function would be involved in the hydrogen bond structure, and therefore the associated signatures found here might not be predictive for generic C17 and C20 interactions.

- The third NH function in Z-Ala\(_6\)-NH\(_2\) is engaged in a \(\beta\)-turn interaction, i.e. C10, a 10 membered ring, with the C=O function that links the ring with the first amino-acid, see figure 2.11). The structure presented in figure 2.11 was obtained from a geometry optimisation. DFT-MD on this geometry, at the very low temperature of 50K shows some structural reorganisation. From the molecular dynamics, we see that the \((NH)_{3}\) function is now more or less simultaneously bonded with two C=O functions at the same time (the C=O group of the first amino acid and the C=O that binds the Z-ring with the first amino acid). Signatures can
Figure 9.14: Unraveling $\omega$(NH) wagging modes for Z-Ala$_6$-NH$_2$ peptide. From top to bottom, infrared experimental spectrum (in purple), theoretical DFT-MD infrared spectrum (in black) and ICDOS signatures of the $\omega$(NH) (in cyan). For the structure, we refer the reader to figure 2.11 in chapter 2. For NH involved into a hydrogen bond, the length of the hydrogen bond is written in the figure, for free NH functions, the label "Free" is written. Motifs formed by hydrogen bonds are also displayed, i.e. C7, C10, C17, C20, ... (N-membered ring interaction).
be found between 500 and 670 cm\(^{-1}\), but again, the resulting β and γ-turns (C10 and C7
interactions, 10 or 7 hydrogen bonded membered rings) hence formed here might not
have signatures fully comparable to generic β and γ-turns, for which one NH function is
linked with one C=O function. We observe several γ and β-turns conformations in our
dipeptide series, but these motifs are formed by NH\(_2\) functions and we can not compare
the signatures.

- The (NH)\(_4\) function, free of hydrogen bond, displays signatures between 380 and 510
  cm\(^{-1}\) that are directly comparable with the 410-490 cm\(^{-1}\) reference established in the
dipeptide series.

- The last two NH functions (NH)\(_5\) and (NH)\(_6\) in Z-Ala\(_6\)-NH\(_2\) are involved in β-turns of type
  I and IV respectively\(^{13}\) (in the Ramachandran notation\(^{110}\) type I: \(\phi_{x+2}=-60^\circ\), \(\psi_{x+2}=-30^\circ\),
  \(\phi_{x+3}=-90^\circ\) and \(\psi_{x+3}=0^\circ\), with x the amino acid that carries the CO function; type IV is
  a kind of miscellaneous category, which includes two of the dihedral angles more than
  40° away from the ideal values for any other type of β-turn). In Z-Ala\(_6\)-NH\(_2\), we have
  \(\phi_4=-60^\circ\), \(\psi_4=-21^\circ\), \(\phi_5=-76^\circ\), \(\psi_5=-8^\circ\), \(\phi_6=74^\circ\), \(\psi_6=-62^\circ\).

We can observe for both NH functions signatures perfectly similar to each other in the
range 600-680 cm\(^{-1}\).

While some motifs expected in large systems appear to be difficult to access in gas phase
studies (C10 in 3\(_{10}\) and C13 in α-helices, even though they have been characterised in the
gas phase\(^{66–68}\)), signatures of ω(NH) of NH functions involved into β-sheets are more easily
accessible. To that end, the system Ac-Phe-OMe and its dimer (Ac-Phe-OMe)\(_2\) organised in a
β-sheet structure (despite its small size), see section 2.1.2 and figure 2.9, are excellent model
systems.

For the monomer, the NH function is free of interaction. The ICDOS spectrum of the ω(NH)
displays a signature at \(\sim 470\) cm\(^{-1}\) (see figure 9.15) that corresponds to the signatures of the
free NH functions in Ac-Phe-Cys-NH\(_2\) dipeptide (410-490 cm\(^{-1}\)) and Z-Ala\(_6\)-NH\(_2\) (380-510
cm\(^{-1}\)) (see figures 9.13 and 9.14).

For the (Ac-Phe-OMe)\(_2\) β-sheet dimer, both hydrogen bonds have really similar bond lengths:
the NH function of the β\(_L\)(g\(^+\)) and β\(_L\)(g\(^-\)) strands are engaged into hydrogen bonds of 2.02±0.09
Å and 1.96±0.06 Å lengths respectively in the 50K DFT-MD trajectory. The proximity of the two
NH functions and their comparable environments induce a strong coupling between the two
ω(NH) motions. As can be seen in figure 9.15, it results in signatures in the ICDOS spectrum of ω(NH) in the range 520-750 cm\(^{-1}\), with two main signatures at \(\sim 520\) cm\(^{-1}\) and \(\sim 640\)
cm\(^{-1}\), thus blue shifted from the 470 cm\(^{-1}\) of the free ω(NH) and red shifted with respect to the
β-turn signature observed for Z-Ala\(_6\)-NH\(_2\). Also note that the ICDOS of the two ω(NH) are
nearly identical which is also an evidence of the coupling between the two amides NH.

We have also studied non peptidic systems that carry NH functions, \textit{i.e.} DNA base pairs and
their analogues presented in section 2.3 of chapter 2. As can be seen in figures 2.20, 2.21,
and 2.22, some NH functions in the nucleobases are free of interaction and others are involved
in strong NH···N hydrogen bonds. Each base is linked to the other by 3 hydrogen bonds: the
NH···N central hydrogen bond, NH\(_2\)···O and NH\(_2\)···N hydrogen bonds. In this context the cen-
Figure 9.15: Unraveling ω(NH) wagging modes for Ac-Phe-OMe and its dimeric form (Ac-Phe-OMe)_2. From top to bottom, infrared experimental spectrum (in dark green), theoretical DFT-MD infrared spectrum (in black) and ICDOS signatures of the ω(NH) (in light green). For structures and labels, we refer the reader to figures 2.9 and 2.10 in chapter 2. For NH involved into a hydrogen bond, the length of the hydrogen bond is written, for free NH functions, the label "Free" is written.
Figure 9.16: Unraveling $\omega$(NH) wagging modes for the DNA base pairs. From top to bottom, infrared experimental spectrum (in blue), theoretical DFT-MD infrared spectrum (in black) and ICDOS signatures of the $\omega$(NH) (in light blue). For structures and labels, we refer the reader to figures 2.20, 2.21 and 2.22 in chapter 2. For NH involved into a hydrogen bond, the length of the hydrogen bond is written, for free NH functions the label "Free" is written.
tral hydrogen bond is shorter (1.86±0.04, 1.96±0.05 and 1.90±0.03 Å for guanine-guanine, guanine-cytosine and ethylated-guanine-methylated-cytosine respectively) than the "optimal" NH····N length that can be observed for example in a β-sheet organisation (2.02±0.09 and 1.96±0.06 Å for the two hydrogen bonds of the (Ac-Phe-OMe)2).

In the ICDOS spectra presented in figure 9.16 for the base pairs, ω(NH) signatures from the central NH····N hydrogen bond can be observed at ≃890, ≃850 and ≃895 cm⁻¹ (for the three systems respectively), i.e. at much higher frequency than the NH waggings for the (Ac-Phe-OMe)2 dimer. This directly reflects the strength of hydrogen bonding in the DNA base pairs.

In the guanine-cytosine base pair, the free NH function has an ICDOS signature at 460 cm⁻¹ comparable to the signatures of the free NH in Ac-Phe-Cys-NH2 (β-turn) dipeptide, Z-Ala₆-NH₂ and Ac-Phe-OMe. But for some reasons, the free NH functions in the guanine-guanine base pair display signatures at 500 cm⁻¹, blue shifted. Maybe a reason for such blue shift is the fact that the NH function now belongs to an aromatic ring, thus with a specific chemical environment. For the specific case of the G1-NH free NH function of the guanine-guanine base pair (see figure 2.20 in chapter 2) the ICDOS peak is even located at ≃570 cm⁻¹. This blue shift might be explained by the NH being surrounded by C=O and NH₂ functions in the ring.

9.3.3 NH₂ waggings (230-860 cm⁻¹)

Figure 9.17 presents the ICDOS decompositions of the NH₂ wagging motions for the Ac-Phe-Gly-NH₂ dipeptide and for the Z-Ala₆-NH₂ peptidic system and illustrates the combination between the two ω(NH) wagging motions of the NH₂ functions, i.e. ω(NH_Bond) and ω(NH_Free), respectively for the hydrogen bonded and free N-H of the NH₂ group (extracted using the values of the H_Bond-N-C-C_α and H_Free-N-C-C_α dihedral angles, see figure 2.6). The results presented here for one dipeptide illustrate results for all dipeptide series.

For Ac-Phe-Gly-NH₂, the NH₂ function is involved in a weak hydrogen bond (2.14±0.10 Å from the dynamics) and we observe common signatures for the two ω(NH_Bond) and ω(NH_Free), both participating to the two series of peaks located between 320 and 450 cm⁻¹ and between 615 and 635 cm⁻¹. This is typical of the coupling between the two motions. Once the two waggings are combined into symmetric and asymmetric motions extracted from the trajectories as the sum (sym) and the difference (asym) between the two H_Bond-N-C-C_α and H_Free-N-C-C_α internal coordinates, we see that the symmetric motion displays its signatures between 320 and 450 cm⁻¹ while the signatures of the asymmetric motion are at ≃620 cm⁻¹. ω(NH_Bond) and ω(NH_Free) are not independent and it makes no sense to observe the signatures of the internal coordinates independently.

For Z-Ala₆-NH₂, we see that the ω(NH_Bond) and ω(NH_Free) contributions are now decoupled. The reason is the strong hydrogen bond that the NH₂ function engages in (1.85±0.07 Å from the dynamics). For strong hydrogen bonds, the recomposition into symmetric and asymmetric motions is not observed and we thus find identical signatures for the symmetric and asymmetric contributions (inbetween 490 and 520 cm⁻¹ and at 720 cm⁻¹, see figure 9.17). What we find is the signature of ω(NH_Bond) being located at ≃720 cm⁻¹ while the ω(NH_Free) signature is found inbetween 490 and 520 cm⁻¹.

We have no NH₂ function with the two NH groups free simultaneously of hydrogen bonds.
9.3. MAPPING LOCAL WAGGING MODES

Figure 9.17: Unraveling $\omega$(NH) wagging modes of NH$_2$ functions in Ac-Phe-Gly-NH$_2$ and Z-Ala$_6$-NH$_2$ peptides. From top to bottom, the infrared experimental spectrum (in color), the theoretical DFT-MD IR spectrum (in black) and the ICDOS signatures of the $\omega$(NH). For the structure and labels of Ac-Phe-Gly-NH$_2$, we refer the reader to figures 2.6 and 2.7 in chapter 2 and the structure of Z-Ala$_6$-NH$_2$ can be found in figure 2.11 in chapter 2. $\omega$(NH$_{Bond}$) refers to the hydrogen atom engaged in a hydrogen bond while $\omega$(NH$_{Free}$) refers to the free NH group. The symmetric and asymmetric motions are extracted by adding and subtracting the values of the two internal coordinates H$_{Bond}$-N-X-X and H$_{Free}$-N-X-X with X=N,C. The length of the hydrogen bond is written (values are extracted from molecular dynamics).
in the systems investigated in the context of this thesis and we can not discuss this precise case, but we found as general rule that NH$_2$ strongly hydrogen bonded induces $\omega$(NH$_{Bond}$) and $\omega$(NH$_{Free}$) contributions not being coupled (and therefore two separate $\omega$(NH) signatures) and NH$_2$ moderately hydrogen bonded induces a coupling between the $\omega$(NH$_{Bond}$) and $\omega$(NH$_{Free}$) contributions. In the following, both approaches (symmetric/asymmetric contributions and NH$_{Bond}$/NH$_{Free}$ contributions) will be used function of the strength of the hydrogen bond.

All the systems in the dipeptides series are capped with one NH$_2$ function that is engaged into a hydrogen bond to form either a $\gamma$-turn structure or a $\beta$-turn structure with the C=O$_{AA}$ neighbour group or with the C=O$_{Phe}$ group respectively, see figures 2.7 and 2.6. Figure 9.18 presents the ICDOS spectra for these NH$_2$ motions. As highlighted in blue, a strong correlation is observed between the NH$_2$···O=C hydrogen bond length and the frequency of the asymmetric out-of-plane $\omega$(NH) wagging motion. The stronger the hydrogen bond, the more blue-shifted the asymmetric wagging signature. Note that such correlation also exists for the N-H backbone Amide V mode dissected above, see subsection 9.3.2, and similar blue-shifts have also been reported for OH···OH hydrogen bonds, see subsection 9.3.4 below. We thus find asymmetric $\omega$(NH$_2$) wagging inbetween $\sim$620 cm$^{-1}$ for the weakest hydrogen bond formed in the $\gamma$-turn conformer A1 of Ac-Phe-Val-NH$_2$ (2.21±0.11 Å from the dynamics) and $\sim$695 cm$^{-1}$ for the stronger hydrogen bonds formed in the $\gamma$-turn and $\beta$-turn structures of Ac-Phe-Pro-NH$_2$ (1.95±0.09 Å from the dynamics). Such correlation between $\omega$(NH$_2$) and NH$_2$···O=C H-Bond strength is not observed for the symmetric wagging motion, as highlighted in red in figure 9.18. Signatures of this motion are systematically found in the 400-510 cm$^{-1}$ interval, somehow similar to $\omega$(NH$_{Free}$) Amide V for the backbone. $\omega_{sym}$(NH$_2$) is however found offset at 595 cm$^{-1}$ for the $\beta$-turn conformation of Ac-Phe-Pro-NH$_2$.

In the Z-Ala$_6$-NH$_2$ system, one NH of the NH$_2$ group is engaged into a 1.85±0.07 Å hydrogen bond and the $\omega$(NH$_{Free}$) and $\omega$(NH$_{Bond}$) signatures fit in the boxes presented in figure 9.18. Signatures of the $\omega$(NH$_{Free}$) are found inbetween 490 and 520 cm$^{-1}$and signature of the $\omega$(NH$_{Bond}$) is found at 720 cm$^{-1}$. As already said, This means that for decoupled motions, $\omega$(NH$_{Bond}$) signatures can fit in the box where the asymmetric motions are displayed and $\omega$(NH$_{Free}$) signatures can fit in the box where the symmetric motions are displayed.

Figure 9.19 presents the ICDOS decomposition of the motions of the NH$_2$ function for all the base pairs investigated here. There are two NH$_2$ functions per base pair, both engaged into strong hydrogen bonds, (from the dynamics: 1.75±0.05 Å and 2.07±0.05 Å for guanine-guanine, 2.07±0.06 Å and 1.79±0.07 Å for guanine-cytosine and 1.90±0.06 Å and 1.74±0.04Å for ethylguanine-methylcytosine). We found that $\omega$(NH$_{Free}$) and $\omega$(NH$_{Bond}$) internal coordinates are not coupled for the strongly hydrogen bonded NH$_2$ functions, and $\omega$(NH$_{Free}$) and $\omega$(NH$_{Bond}$) internal coordinates are coupled for less strongly hydrogen bonded NH$_2$ functions, as already emphasised above for the other molecules.

Starting with the weak hydrogen bond in the guanine-guanine base pair (2.07±0.05 Å in the dynamics), one can find the signatures at $\sim$600 cm$^{-1}$ for the asymmetric motion and at $\sim$370 cm$^{-1}$ for the symmetric motion. Both signatures would fit in the red and blue boxes
Figure 9.18: Unraveling $\omega$(NH) wagging modes of the NH$_2$ group for the dipeptide series. From top to bottom, the infrared experimental spectrum (in color), the theoretical DFT-MD infrared spectrum (in black) and the ICDOS signatures of the $\omega$(NH). For the structure of the dipeptides, we refer the reader to figure 2.7 in chapter 2. The symmetric and asymmetric motions are extracted by adding and subtracting the values of the two internal coordinates H$_{Bond}$-N-X-X and H$_{Free}$-N-X-X, with X=N,C, see figure 2.6 in chapter 2. The length of the hydrogen bond is written (values are extracted from molecular dynamics).
Figure 9.19: Unraveling $\omega$(NH) wagging modes of the NH$_2$ group of DNA base pairs. From top to bottom, the infrared experimental spectrum (in blue), the theoretical DFT-MD infrared spectrum (in black) and the ICDO signatures of the $\omega$(NH) (in cyan). For the structure and labels of the DNA base pairs, see figures 2.20, 2.21 and 2.22 in chapter 2. $\omega$(NH$_{Bond}$) refers to the hydrogen atom engaged in a hydrogen bond while $\omega$(NH$_{Free}$) refers to the free NH. The symmetric and asymmetric motions are extracted by adding and subtracting the values of the two internal coordinates H$_{Bond}$-N-X-X and H$_{Free}$-N-X-X, with X=N,C. The length of the hydrogen bond is written (values are extracted from molecular dynamics).
9.3. MAPPING LOCAL WAGGING MODES

presented in figure 9.17 indicating that we might extrapolate universal relationship (not only for peptidic systems) linking the frequency of the symmetric motions/\(\omega(NH_{Bond})\) on one hand and the asymmetric motion/\(\omega(NH_{Bond})\) signatures on the other hand with the hydrogen bond length.

All the NH\(_2\) groups in this series are engaged in hydrogen bonds shorter than 1.95 Å (which is the value for the hydrogen bond in Ac-Phe-Pro-NH\(_2\) in the \(\beta\)-turn conformation) at the exception of two 2.07 Å hydrogen bonds in the guanine-guanine and guanine-cytosine that will be described below.

The \(\omega(NH_{Bond})\) signatures of these strong hydrogen bonded NH functions are blue shifted in comparison with the dipeptides signatures. We found two massifs inbetween 697 and 720 cm\(^{-1}\) and inbetween 807 and 831 cm\(^{-1}\) for the 1.79±0.07 Å hydrogen bond of guanine-cytosine, a massif inbetween 631 and 642 cm\(^{-1}\) (+ICDOS signature at 715 cm\(^{-1}\) that is not active in the infrared spectrum) for the 1.90±0.06 Å hydrogen bond of the 9-ethylated guanine···1-methylated cytosine base pair, and one ICDOS signature at 847 cm\(^{-1}\), not active in the infrared spectra (DFT-MD and experimental) for the 1.74±0.04 Å hydrogen bond of the 9-ethylated guanine···1-methylated cytosine base pair.

Most of the \(\omega(NH_{Free})\)/symmetric signatures are located in the range 390-520 cm\(^{-1}\) exactly as in the dipeptides series: peaks at 407 and 451 cm\(^{-1}\) for the 1.79±0.07 Å hydrogen bond of guanine-cytosine, massif inbetween 322 and 400 cm\(^{-1}\) for the 1.90±0.06 Å hydrogen bond of the 9-ethylated guanine···1-methylated cytosine base pair and inbetween 430 and 447 cm\(^{-1}\) for the 1.74±0.04 Å hydrogen bond of the 9-ethylated guanine···1-methylated cytosine base pair. The hydrogen bond lengths reported here are for the second NH function of the NH\(_2\) group (the one engaged into a hydrogen bond).

We find three exceptions out of the general rule that we can not explain (for now):

- For the 1.75±0.05 Å hydrogen bond in guanine-guanine, one can find the \(\omega(G1_{Bond})\) signature at \(\approx\)810 cm\(^{-1}\) which is more or less what we could expect based on the previous observations, but the \(\omega(G1_{Free})\) signatures are found between \(\approx\)240-290 cm\(^{-1}\) which correspond to very low frequencies in comparison with the range 390-520 cm\(^{-1}\) observed up to now.

- For the 2.07±0.06 Å hydrogen bond in guanine-cytosine, the asymmetric signatures are found between \(\approx\)470 and 510 cm\(^{-1}\) in the ICDOS spectrum but this mode is not active in infrared (since no counterpart in the DFT-MD theoretical spectrum can be found). The symmetric signatures are found between 250 and 330 cm\(^{-1}\). Both signatures are red shifted in comparison to the frequencies expected and would be found out of the boxes presented figure 9.18 (\(\approx\)400-510 cm\(^{-1}\) might be expected for the symmetric signatures and \(\approx\) 590-740 cm\(^{-1}\) for the asymmetric signatures for this hydrogen bond length).

Among the systems investigated in this thesis, none of them have a NH\(_2\) group that is planar with respect to its environment because all the NH\(_2\) functions investigated are engaged in hydrogen bonds. It has been shown by Yatsina et al.\(^{204}\) that planar NH\(_2\) functions have very specific signatures in the far infrared due to tunnelling inversion processes. This type of planar function would be interesting to study to complete the discussion presented here.
9.3.4 OH waggings (200-700 cm\(^{-1}\))

The signatures of the \(\omega(\text{OH})\) OH waggings have been extensively described in our paper "Anharmonic, dynamic and functional level effects in far-infrared spectroscopy: phenol derivatives"\(^{108}\) presented in chapter 7. While the paper discusses the systems phenol, catechol, saligenin, nitrophenol and two conformations of the salicylic acid, we chose here to restrict our presentation to phenol, catechol and saligenin and to add the phenol-water complex for which the far infrared spectrum has not been published yet. These four systems are presented in section 2.2 and their structures in figure 2.15.

This selected series has been chosen for the progressive enhancement in the strength of the hydrogen bond: in phenol, OH is free, there is a weak hydrogen bond in catechol (hydrogen bond length: 2.20\(\pm\)0.07 Å), the saligenin has a strong hydrogen bond (hydrogen bond length: 1.98\(\pm\)0.06 Å) and the phenol-water cluster has a very strong hydrogen bond (hydrogen bond length: 1.93\(\pm\)0.10 Å). Hydrogen bond distances given here are extracted from molecular dynamics.

Figure 9.20 presents the ICDOS decomposition of the \(\omega(\text{OH})\) OH wagging. The free OH function of phenol displays a signature located at 309 cm\(^{-1}\) in the experimental spectrum. This value is probably the absolute reference for the free OH wagging, lower in frequency than the signature of the free NH located at \(\simeq\)470 cm\(^{-1}\).

Once engaged in a weak hydrogen bond in catechol (2.20\(\pm\)0.07 Å from the dynamics), the signature of the OH wagging is blue shifted and we observe a band located at 415 cm\(^{-1}\) in the experimental spectrum (we find multiple peaks in the DFT-MD spectrum at 396, 409 and 423 cm\(^{-1}\)). For the free OH function in this molecule, we observe one peak at 221 cm\(^{-1}\) in the experimental spectrum (red shifted by 88 cm\(^{-1}\) in comparison with the reference in phenol). The DFT-MD spectrum displays signatures of the free OH wagging at 233 and 282 cm\(^{-1}\). This second feature has no counterpart in the experimental spectrum. This red shift might probably be explained by the increase in the electronic density on the free OH function in catechol.

When the OH is engaged into a strong hydrogen bond as in saligenin (1.98\(\pm\)0.06 Å from the dynamics), the \(\omega(OH)\) ICDOS signatures are strongly blue shifted, now located at \(\simeq\)665 and 680 cm\(^{-1}\) with only one experimental counterpart at 690 cm\(^{-1}\). The free OH wagging ICDOS shows participations into the experimental peaks located at 340 and 385 cm\(^{-1}\), blueshifted with respect to the 309 cm\(^{-1}\) reference in phenol.

For the phenol-water cluster that provides the shorter hydrogen bond length in this series (1.93\(\pm\)0.10 Å from the dynamics), \(\omega(OH)\) is located at 695 cm\(^{-1}\) in the experimental spectrum. For the two free OH functions of water, we observe the symmetric contribution between 220 and 260 cm\(^{-1}\) while the asymmetric motion displays signatures below 100 cm\(^{-1}\).
Figure 9.20: Unraveling $\omega$(OH) wagging modes. In each plot: from top to bottom, the infrared experimental spectrum in red, the theoretical DFT-MD IR spectrum in black and the ICDOS signatures of the $\omega$(OH) in brown. The four systems selected, i.e. phenol, catechol, saligenin and phenol water are presented in section 2.2 and their structures in figure 2.15 in chapter 2. The lengths of the hydrogen bonds are written. When the OH function is not engaged into a hydrogen bond, i.e. for phenol, the label "Free" is used. The index 1 and 2 refer to the position of the OH function on the aromatic ring, see figure 2.16.
9.3.5 CH$_3$ rotational modes ($<300$ cm$^{-1}$)

In the dipeptide series, one can find two spectral features for CH$_3$ hindered rotational motions related to the two methyl populations in these systems. We will use the example of the Ac-Phe-Ala-NH$_2$ system that carries two CH$_3$ functions that correspond to these two populations (but all CH$_3$ functions follow the same results). The assigned structure of Ac-Phe-Ala-NH$_2$ is displayed in panel a of figure 9.21 with two circles highlighting the CH$_3$ functions of the residue (green) and of the terminal (purple). The CH$_3$ terminal function is carried by a carbon atom in a sp$^2$ hybridisation while the CH$_3$ function of the residue is carried by a carbon atom in a sp$^3$ hybridisation. The CH$_3$ rotational motions are extracted by ICDOS of dihedral angle coordinates, i.e. H$_{CH_3}$-C-C-N$_{Phe}$ and H$_{CH_3}$-C-C$_\alpha$-N$_{AA}$, respectively for the backbone CH$_3$ and for CH$_3$ in the alanine residue, see figure 2.6 for the definitions. For one given methyl group, we look at the simultaneous signatures of the three hydrogen atoms, which indeed give us the hindered rotational motion of CH$_3$ as a 'solid body'. The evolution with time of the two dihedral angles values are plotted over the length of the trajectory in panel b of figure 9.21. We observe that the standard deviation is much bigger for the CH$_3$ terminal function (sp$^2$ hybridised). This difference has a direct influence on the ICDOS signatures as can be seen in panel c, figure 9.21. Indeed, signatures of the hindered rotation of the terminal CH$_3$ group of the dipeptides are found below 100 cm$^{-1}$, and are systematically involved in delocalised modes through couplings to backbone dihedral motions (modes described in section 9.4), while for the (sp$^3$ hybridised) alanine residue in Ac-Phe-Ala-NH$_2$ dipeptides, signatures of the hindered rotation of the methyl group are found at 220 and 242 cm$^{-1}$.

Distinct rotational signatures for residue methyl groups ('hydrophobic') and backbone ('hydrophilic') methyl groups have already been observed for peptides in the condensed phase within the same vibrational range as in the gas phase. For our systems, we clearly observe that the terminal methyl group has more amplitude in its hindered rotational motion because of its sp$^2$ carbon hybridisation. When the methyl is part of the peptide residues, its carbon being now sp$^3$ hybridised, a different environment is provided to the hydrogen atoms when the methyl group is rotating, the methyl is thus more constrained in its rotation, leading to a hindered rotational motion at higher frequency. This observation remains true for all the CH$_3$ terminal functions systematically present in the dipeptides series or for all the CH$_3$ functions of the six alanine amino acids of the Z-Ala$_6$-NH$_2$ system. These ICDOS spectra are not presented here.

To illustrate further the difference between sp$^2$ and sp$^3$ hybridisations, we plotted the potential energy surface for the CH$_3$ rotational motion in figure 9.22 and it appears that the barrier that blocks the free rotation of the CH$_3$ group is $\simeq 15$ times higher for the sp$^3$ environment than for the sp$^2$ one.
Figure 9.21: Panel a presents the assigned structure for the Ac-Phe-Ala-NH$_2$ dipeptide with purple and green circles highlighting the CH$_3$ function of the residue and the CH$_3$ terminal function respectively. Panel b presents the evolution with time of the dihedral angles H-C-C-N in the trajectory, probing the rotation of the CH$_3$ group. Purple for the CH$_3$ of the residue and green for the CH$_3$ terminal. The average value and standard deviations are reported. Panel c displays the theoretical DFT-MD infrared spectrum in black and the ICDOS spectra of the dihedral angles H-C-C-N (purple: CH$_3$ of the residue; green: CH$_3$ terminal).
Figure 9.22: Potential energy surface for the H-C-C-N dihedral angle in Ac-Phe-Ala-NH₂. This angle drives the rotation of the CH₃ group. In purple, the potential energy surface for the CH₃ function of the residue (sp³ environment) and, in green, the potential energy surface for the CH₃ terminal function (sp² environment), see figure 9.21. For all the dihedral angle values, the same CH lengths and C-C-H angles are fixed for the three hydrogen atoms.
9.4 Mapping delocalised modes

As discussed in section 9.1, while the ICDOS spectra are well suited for unraveling local wagging motions, delocalised modes are more tedious to be unraveled by ICDOS spectra and no percentages of composition of the modes can be provided, but we can still provide relevant interpretation in the infrared spectra. We found that collective vibrational motions delocalised over the entire structure are present in the entire far infrared range (<800 cm$^{-1}$), but these modes can be hidden by intense local wagging modes. Indeed delocalised/collective modes are complex combinations of motions, some of them inducing strong variation of the dipole moment while others induce no or few variations of the dipole moment, and there can be compensations into the dipole variations from certain motions. This results in modes with of more or less the same intensity, in any case less intense than the wagging modes. For the dipeptide series and the larger Z-Ala$_6$-NH$_2$ peptide, we found that the local $\omega$(NH) and $\omega$(CH) modes dominate in intensity the range $\approx$400-800 cm$^{-1}$ while delocalised and collective modes dominate the range 0-400 cm$^{-1}$.

9.4.1 Peptide backbone and lateral chain motions (<400 cm$^{-1}$)

Although being less useful for a quantitative interpretation of delocalised modes, the ICDOS spectra can still provide good information on the weight of one chosen internal coordinate into the more delocalised vibrational modes and therefore provide a good idea of the composition of the collective modes. For illustration, the ICDOS spectra of the angles and dihedral angles of the backbone of Ac-Phe-Ser-NH$_2$ have been plotted in figure 9.23, top panel for the angles and bottom panel for the dihedral angles. One can see that ICDOS spectra of the angles of the backbone display signatures in the whole range 0-800 cm$^{-1}$ but the most intense signatures arise below 400 cm$^{-1}$. In the same way, we find the most intense signatures for the dihedral angles below 100 cm$^{-1}$. Therefore, even if the angles and dihedral angles motions might be coupled with other internal motions (at least for the peptidic systems) the range 100-400 cm$^{-1}$ is dominated by modes with bendings of the backbone as main contributions and the range <100 cm$^{-1}$ by modes with torsions of the backbone as main contributions.

Looking a bit more closely at the ICDOS spectra in figure 9.23, one can see that these bending and dihedral motions are highly coupled: we observe multiple signatures of each internal coordinate in the spectra, and several dihedral angles simultaneously providing the same spectral signature (i.e. same frequency). The resulting modes are thus collective and highly delocalised over the whole peptide backbone. The fact that we mentioned many times that we are not able to provide a % of participation of each internal motion into the final mode does not matter so much for such collective modes. We see clearly how many angles/dihedral angles are involved into a peak and their respective weight is of no great importance in the end for the final interpretation.

At the time of writing, the FELIX Free Electron Lasers laboratory facility can be tuned down to 80-90 cm$^{-1}$ and is limited for lower frequencies. Obviously, our theoretical calculations do not have such limitations, and the theoretical dynamical infrared spectra presented here provide the expected spectral signatures below 100 cm$^{-1}$, to be validated by experiments when they will become available.
Figure 9.23: For each panel, (from top to bottom) the experimental spectrum of Ac-Phe-Ser-NH$_2$ in light blue, the theoretical DFT-MD infrared spectrum in black and the ICDOS signatures of the internal coordinates of the backbone in dark blue. ICDOS of the backbone angles are displayed in the top panel and ICDOS of the backbone dihedral angles are displayed in bottom panel, see figure 2.6 for the definitions. For clarity reasons, the intensities of the ICDOS signatures of the dihedral angles have been multiplied by four above 100 cm$^{-1}$. For the dihedral angles, the Ramachandran notation, introduced in figure 2.6, is used$^{110}$. Note that we use the dihedral angles $\omega'$ (C-C-N$\text{Phe}/\text{AA}$-C$_{\alpha}$) instead of the traditional dihedral angle $\omega$ (O-C-N$\text{Phe}/\text{AA}$-H). Indeed $\omega$ (O-C-N$\text{Phe}/\text{AA}$-H) probes the out of plane motions of the oxygen and hydrogen atoms instead of the backbone deformations.
9.4.2 Hydrogen bond signatures (<below 250 cm$^{-1}$>)

One of the most sought after signatures are those of the hydrogen bonds. In figure 9.24, top panel, are plotted the ICDOS spectra of the internal coordinates involving the hydrogen bond in the Ac-Phe-Ser-NH$_2$ dipeptide. The $\nu$(H-Bond) corresponds to the ICDOS of the hydrogen bond-length (between the NH$_2$ and the C=O functions), $\delta$(H-Bond) and $\omega$(H-Bond) correspond respectively to the angle N-H···O and dihedral angle N-H···O-C of the hydrogen bond.

All three ICDOS spectra display a large amount of peaks in the range 0-800 cm$^{-1}$ with identical signatures in terms of frequency, i.e. all three internal coordinates are involved in the same vibrational modes. It shows how much the hydrogen bond signatures are coupled to other motions and how delocalised in the whole spectrum it is. While the ICDOS intensities of the angle N-H···O and dihedral angle N-H···(N/O)-C are really similar, we observe different relative intensities for the $\nu$(H-Bond) signatures.

$\delta$(H-bond) and $\omega$(HBond) ICDOS give a main signature corresponding to the $\omega$(NH) mode (at $\approx$450 cm$^{-1}$). This is not surprising because all three internal coordinates are redundant (on other words, because of the intramolecular character of the hydrogen bond, the three internal coordinates chosen are not canonical) which means that these internal coordinates are just other probes of the motions investigated with canonical internal coordinates. Here we see that $\delta$(H-bond) and $\omega$(HBond) will efficiently probe the $\omega$(NH) dihedral angle and we observe a similar issue for $\nu$(H-Bond). Indeed, the most intense signature in the $\nu$(H-Bond) ICDOS is located at $\approx$135 cm$^{-1}$ but the ICDOS of the $\omega$(C=O$_{AA}$) (not presented here, out of plane motion of the O$_{AA}$ atom, the carbon atom of the C=O$_{AA}$ function of the serine amino acid) also provides an intense signature at this frequency, and this is expected as the C=O$_{AA}$ is the one involved in the hydrogen bond. In Ac-Phe-Ser-NH$_2$, the $\nu$(H-Bond) internal coordinate is another probe of the $\omega$(C=O$_{AA}$) out of plane motion of the O$_{AA}$ atom.

The issue here is that using only the ICDOS tool, it is difficult to definitely say that the hydrogen bond stretching is the driving force of this mode.

In the following, we discuss only the $\nu$(H$_{Bond}$) ICDOS spectrum that corresponds to hydrogen bond stretchings.

For a system such as the guanine-guanine base pair (smaller than Ac-Phe-Ser-NH$_2$, 32 atoms instead of 40) with three intermolecular hydrogen bonds instead of one for Ac-Phe-Ser-NH$_2$, the percentage of participation of hydrogen bonds into the vibrational modes is mechanically higher. The bottom panel of figure 9.24 presents the ICDOS spectra of the three hydrogen bond stretchings of the system. We observe couplings between the stretchings (i.e. the same peaks) and intense signatures at $\approx$80, 100 and 130 cm$^{-1}$. There are few supplementary signatures of the hydrogen bond stretchings above 130 cm$^{-1}$, but they systematically correspond to the $\omega$(NH) modes (of course mechanically related to the hydrogen bond stretch). With the present information, we can not be sure that the three modes at $\approx$80, 100 and 130 cm$^{-1}$ can indeed be assigned to (pure) hydrogen bond modes.
Figure 9.24: For each panel, (from top to bottom) the experimental spectrum in blue, the theoretical DFT-MD infrared spectrum in black and the ICDOS signatures of the internal coordinates involving the hydrogen bonds of the systems in cyan. The top panel presents the ICDOS signatures of the length, angle and dihedral angle of the single hydrogen bond present in the Ac-Phe-Ser-NH$_2$ dipeptide system (hydrogen bond between the NH$_2$ and the C=O functions, see figure 2.6). The bottom panel presents the guanine-guanine base pair (system with three hydrogen bonds, see figure 2.20) for which only hydrogen bond stretching motions are displayed.
In our paper "Fingerprints of Inter- and Intramolecular Hydrogen Bond in saligenin-Water Clusters Revealed by Mid- and Far-Infrared Spectroscopy"\textsuperscript{111}, we have furthermore found a correlation between the strength of the hydrogen bond and the frequency of the signatures. This paper can be found in chapter 8 and presents a study of the saligenin system in interaction with one to three water molecules, see figure 2.14 and section 2.2 for the structures.

This correlation is not easy to unravel because the hydrogen bond stretching is not a local mode, as seen above (at least in saligenin-water clusters). In our paper\textsuperscript{111}, we therefore used the frequency of the mode that includes the largest contribution of the hydrogen bond length internal coordinate (based on ICDOS spectra and harmonic normal modes decomposition). The results are reported in figure 9.25. Two different measures for the hydrogen bond strength are evaluated in this work, namely the hydrogen donor OH covalent bond length $L_{OH}$ and the length of the hydrogen bond $L_{HB}$. It was previously established that $L_{OH}$ is a good measure for the strength of intramolecular hydrogen bonds\textsuperscript{107}. Note that all the hydrogen bond lengths used for this figure come from optimised structures at the MP2/6-311++G(2p,2d) level of theory (as used in the paper).

The assigned experimental bands are represented by crossed circles with a pink line and the calculated frequencies are displayed by squares. The color of the symbols indicates the interaction that the specific OH group is involved in: the OH groups involved in a OH···π interaction in red, the hydrogen bonded OH groups of the saligenin molecule are in blue and the hydrogen bonded OH groups of waters are in black.

The frequencies of the modes are displayed in the top panel as a function of the increase in $L_{OH}$ with respect to the minimum OH bond length $L_{OH,\min} (0.9629 \text{ Å})$ found in the saligenin-water complex. The pink and grey lines fit the experimental and calculated results respectively, using linear functions. In the bottom panel, the frequencies of the hydrogen bond related modes are displayed as a function of the hydrogen bond length ($L_{HB}$). The hydrogen bond length varies from 1.71 Å as found in the hydrogen bond donated by the phenolic OH group of saligenin in SLG-3w-b, see figure 2.14, to 3.731 Å for the extremely weak OH···π interaction in the bare saligenin, see figure 2.15. The pink and grey lines fit the experimental and calculated results respectively, using exponential functions.

We therefore observed that the hydrogen bond stretching is sensitive and correlated to the strength of the hydrogen bond with a blue shift for the strongest (more electronic density on the hydrogen bond). But this is not a probe as direct as the stretchings or waggings of the covalent bonds since the hydrogen bond stretching is usually involved into several vibrational modes that have therefore a high spatial delocalisation. One can note the exception of the base pair systems for which the higher number of hydrogen bonds as well as the rigidity of these hydrogen bonds seem to provide vibrational modes dominated by hydrogen bond stretchings.
Figure 9.25: For saligenin(H$_2$O)$_{1-3}$ complexes, frequencies of the hydrogen bond stretching vibrations as a function of (top panel) the length of the OH moiety $L_{OH}$ corrected for the minimum OH length ($L_{OH,\,min}=0.9629$ Å ) and (bottom panel) as function of the hydrogen bond length. The OH covalent bond length is a probe of the strength of the involved hydrogen bond, where a longer OH bond corresponds to a stronger hydrogen bond$^{107}$. Calculated results are shown as squares, experimental results as crossed circles. For the $L_{OH}$ length, a linear relation is observed between the frequency shift and the OH length, with linear fits to the experimental frequencies presented as pink lines and fits to the calculated data as grey lines. For the hydrogen bond length, a non linear relation is observed between the frequency shift and the OH length, fits to the experimental frequencies presented as pink lines and fits to the calculated data as grey lines. This figure has been extracted from our paper: "Fingerprints of Inter- and Intamolecular Hydrogen Bond in saligenin-Water Clusters Revealed by Mid- and Far-Infrared Spectroscopy"$^{111}$ presented in chapter 8.
9.4. MAPPING DELOCALISED MODES

9.4.3 Alternative analyses

Figure 9.26 presents the experimental and theoretical infrared spectra of the Z-Ala$_6$-NH$_2$ peptide and the ICDOS spectra of the twenty dihedral angles of the backbone (3 per amino acid, $\omega$, $\phi$ and $\psi$ plus Dih$_{-1}$ and Dih$_{-2}$ that correspond to the junction between the peptidic backbone and the chromophore, see figure 2.12). With this figure, I want to illustrate and discuss one more time the spatial delocalisation of the modes. The signatures arising from dihedral angles above 600 cm$^{-1}$ are relatively localised in space over the backbone, for instance the signatures at $\approx$660 and $\approx$725 cm$^{-1}$ are spread over 6 and 5 dihedral angles respectively, thus representing approximately only 2 amino acids over the backbone. The modes below 400 cm$^{-1}$ are much more delocalised spatially with signatures for instance over 13 dihedral angles for the mode at $\approx$390 cm$^{-1}$ (i.e. 4 amino acids) or 10 dihedral angles for the mode at $\approx$330 cm$^{-1}$ (3 amino acids). We observe a clear correlation between the frequency and the spreading of the mode over the backbone: the lower the frequency, the larger the spatial delocalisation over the length of the backbone.

One other thing that can be learned from figure 9.26 is that it seems really difficult for a normal mode to spread through a hydrogen bond. If we observe, for example, the mode around 310 cm$^{-1}$, we see signatures from seven dihedral angles between $\psi_4$ and $\psi_6$, showing a mode spreading over the backbone (localised around the amino acids 4 to 6). If we take a look at the structure figure 2.12, we see that the NH functions of the 5$^{th}$ and 6$^{th}$ amino acids are engaged into hydrogen bonds with the C=O functions of the 2$^{nd}$ and 3$^{rd}$ amino acids, and at the same time the C=O functions of the 6$^{th}$ amino acid is hydrogen bonded to the NH function of the 1$^{st}$ and 2$^{nd}$ amino acids. Despite these hydrogen bonds the same mode at 310 cm$^{-1}$ displays no signature in the spatial range of the three first amino acids. Therefore, the delocalisation of the modes (for peptides at least) occurs along the backbone and not through the hydrogen bonds.

Another approach for studying and unraveling delocalised modes is presented in figure 9.27, for the dimer (Ac-Phe-OMe)$_2$ (see figures 2.9 and 2.10 for the structures). We have plotted here the mass weighted percentage of the displacement of the atoms over each strand according to a harmonic normal mode decomposition for each normal mode. If the two values are both at 50%, it means that the mode is delocalised over the two strands; if the two values are at 0 and 100% it means that the mode is localised over one strand only. Beyond 600 cm$^{-1}$, we can see that some modes are localised over one strand only while others are delocalised over the two strands. Interestingly the modes delocalised over the two strands are the (local) wagging modes that we analysed before. The two $\omega$(NH) motions couple to produce a mode located over both strands. Despite this remark, only these two internal coordinates are strongly coupled in this mode and the two NH functions are spatially close, therefore the local adjective remains correct for the NH wagging modes. It is the same for the $\omega$(CH) motions of the two aromatic rings that couple together and the associated modes for example around 690 cm$^{-1}$ are well delocalised over the two strands. Below 600 cm$^{-1}$, where delocalised modes are expected (as seen before), there is a really good correlation between the frequency of the mode and its spatial localisation (into one or two strands): the modes are almost perfectly localised over one strand between 400 and 600 cm$^{-1}$ and almost perfectly delocalised over the two strands below 150 cm$^{-1}$. The domain 150-400 cm$^{-1}$ is found intermediate.
Figure 9.26: From top to bottom, the experimental spectrum of Z-Ala$_6$-NH$_2$ peptide in purple, the theoretical DFT-MD infrared spectrum in black and the ICDOS signatures of the twenty dihedral angles of the backbone, see figure 2.11. For the dihedral angles, the Ramachandran notation, introduced in figure 2.6, is used$^{110}$. Note that we use the dihedral angles $\omega'$ (C$_{\beta}$-C$_{\alpha}-$N$_{\text{Phe/AA}}$) instead of the traditional dihedral angle $\omega$ (O-C$_{\text{N$_{\text{Phe/AA}}$}-H}$). Indeed $\omega$ (O-C$_{\text{N$_{\text{Phe/AA}}$}-H}$) probes the out of plane motions of the oxygen and hydrogen atoms instead of the backbone deformations. The index numbers correspond to the label of each aminoacid in the peptidic chain. Dih – 1 and Dih – 2 refer to the dihedral angles that can be found at the junction between the peptidic chain and the chromophore, see figure 2.11.
9.5. OUTLOOK

Figure 9.27: For each normal mode of (Ac-Phe-OMe)$_2$ β-sheet dimer, plot of the mass weighted percentage of the displacement of the atoms over each strand according to a harmonic normal modes decomposition. When the two values are at 50%, the mode is perfectly delocalised over the two strands of the dimer. When the couple of values is 0 and 100%, the mode is perfectly localised over one strand only.

9.5 Outlook

In this chapter we have analysed the far infrared vibrational modes for all the systems investigated in this thesis, and from this we have given a final view of the modes to be expected. This has been achieved by the ICDOS decomposition, which principles have been presented in section 9.1. Two kinds of modes have be found in the far infrared/THz domain, i.e. 'local' modes for which one single internal coordinate (wagging motion of hydrogen atoms) dominates the motions in the vibrational mode and 'delocalised/collective' modes for which several internal coordinates dominate the motions in the mode. In this case, motions are therefore much more spatially delocalised over the molecule and this has been nicely illustrated with the larger Z-Ala$_6$-NH$_2$ peptide.

The local modes, i.e. $\omega$(CH), $\omega$(NH), $\omega$(OH) found here in the range 300-900 cm$^{-1}$ are mainly composed of "large amplitude" wagging motions that induce large variations of the gas phase molecule dipole moment, and thus provide rather large intensities of these modes in the infrared spectrum. Delocalised modes are active in the whole range 0-800 cm$^{-1}$ even though their infrared intensities are weaker than for the waggings and therefore hidden when they overlap with the wagging modes. We find that the collective backbone delocalised modes are infrared active mainly below 400 cm$^{-1}$.

A mapping of the vibrational modes in the far infrared/THz spectral domain has been pre-
presented in this chapter, and an overview is presented in figure 9.28. This mapping has been achieved more specifically for peptides, base pairs and phenol derivatives gas phase molecules. We believe our results are transferable to other systems, however incomplete it might be. This chapter is currently transformed into an article to be submitted in November 2017 to J. Phys. Chem. A. (invitation for a Festschrift).

![Figure 9.28: Mapping of the far infrared/Tera Hertz gas phase vibrational modes achieved in this work. Analyses are based on peptides, base pairs and phenol derivatives molecules.](image-url)
Chapter 10

Which method could be "qualified as better" in order to calculate far infrared/THz spectra?

In this chapter, different theoretical methods are tested for the calculation of vibrational infrared spectra, the spectra are compared with experiments. This is done for most of the systems introduced in chapter 2. It provides the demonstration of the quality of the DFT-MD/BLYP-D3 representation (used in all other chapters of this thesis) that takes into account anharmonicities of the potential energy surface and of the dipole surface as well as mode couplings. This demonstration is essential to ensure the quality of the vibrational mode assignments presented in chapter 9 and of the conformational assignments of (Ac-Phe-OMe)$_2$ or Saligenin-water clusters respectively presented in chapters 8 and 11 (and more generally in chapters 5-7). Comparisons between anharmonic DFT-MD, anharmonic VPT2 and harmonic spectra are presented in this chapter and we identify here in which context (i.e. for which systems, chemical functions, vibrational modes) the anharmonicities need to be taken into account. DFT-MD/B3LYP-D3 spectra are also introduced and discussed, to see if going beyond a GGA DFT functional is necessary.

Section 10.1 presents a comparison between experimental and DFT-MD/BLYP-D3 spectra. The molecular dynamics and the methodology to extract an infrared spectrum from the trajectory are presented respectively in sections 4.2 and 4.3 of chapter 4.

Section 10.2 presents a comparison between DFT-MD/BLYP-D3 spectra of Ac-Phe-Ser-NH$_2$ calculated with two different gaussian basis sets and discusses the influence of the basis set on the infrared spectrum. This comparison is necessary to introduce sections 10.3 and 10.4 that present comparisons between DFT-MD infrared spectra and alternative theoretical infrared spectroscopy methods (harmonic and anharmonic VPT2) that are unfortunately not calculated with the same basis set as the DFT-MD. This is probably the biggest weakness of this chapter, but I believe that the comparisons become relevant as soon as section 10.2 has demonstrated that the basis set effects are negligible.
The DFT-MD (anharmonic) spectra are much more expensive to calculate than the harmonic spectra (method presented in chapter 4) and in this respect it is important to understand what are the vibrational modes that are indeed anharmonic (and "deserve" or require an anharmonic treatment) and which modes are intrinsically harmonic and for which one can use the cheaper harmonic approximation for vibrational spectroscopy (DFT-MD is roughly 200 times more expensive than the harmonic approximation, time given here for a 20 ps trajectory). A MD/harmonic comparison at the BLYP-D3 electronic level is presented in section 10.3.

Section 10.4 presents a comparison between anharmonic VPT2 (as implemented in the Gaussian package by Barone\textsuperscript{134–136}), anharmonic MD and harmonic methods for infrared spectra. The VPT2 spectroscopy is presented in section 4.6 of chapter 4. It is an approximation that takes into account anharmonicities in a different way than the molecular dynamics for a price slightly cheaper (DFT-MD is roughly twice more expensive as VPT2 calculation, for a 20 ps trajectory). One weakness of the VPT2 approach is that it keeps the vibrational modes from the harmonic approximation calculation, it is therefore a (perturbative) method being a correction to the harmonic approximation.

Finally, section 10.5 presents a comparison between DFT-MD/BLYP-D3 and DFT-MD/B3LYP-D3 spectra for phenol derivatives. This comparison has not been performed for larger systems due to a lack of time but that would be interesting to do. B3LYP-D3 is a hybrid GGA functional that is supposedly "known" to be a better electronic representation than the GGA functional, such as BLYP-D3 (as well as more expensive). Both density functionals are described in section 4.1 in chapter 4. We will discuss what improvements can be expected in the infrared spectra from a non GGA density functional.

All discussions in this chapter are for the far infrared/Tera Hertz spectral domain.
10.1 Anharmonic infrared spectra calculated from DFT-MD/BLYP-D3

This section presents a comparison between experimental and DFT-MD/BLYP-D3 spectra (this is the only theoretical tool used in the other chapters of this thesis). The molecular dynamics and the methodology to extract an infrared spectrum are presented respectively in sections 4.2 and 4.3 of chapter 4. The vibrational mode assignments used here are presented mainly in chapter 9 (as well as the advantages/drawbacks of the ICDOS method).

10.1.1 Ac-Phe-AA-NH₂ series

The far-infrared spectra measured experimentally and extracted from the DFT-MD/BLYP-D3 theoretical calculations for the dipeptide series are reported in Figure 10.1. The dipeptides series has been introduced in section 2.1.1 of chapter 2 and all detailed analyses related to this series are presented in chapters 5, 6 and 9.

Spectra are organised from top to bottom, with the order of the γ-turn A₁ conformation of Ac-Phe-Val-NH₂, the β-turn of Ac-Phe-Cys-NH₂, γ-turns of Ac-Phe-Gly-NH₂, Ac-Phe-Ala-NH₂, (A2) Ac-Phe-Val-NH₂, Ac-Phe-Ser-NH₂, Ac-Phe-Cys-NH₂, Ac-Phe-Pro-NH₂ and the β-turn of Ac-Phe-Pro-NH₂. We report the 0-800 cm⁻¹ far infrared spectra of these peptides in figure 10.1, intensities have been doubled below 400 cm⁻¹ for clarity reasons. Experimental spectra are plotted in color while the theoretical anharmonic dynamical DFT-MD spectra are plotted in black, just below each experiment.

We observe a good match between theory and experiment with in particular almost all peaks being reproduced by the theoretical representation (only exceptions are few minor experimental contributions without theoretical counterparts like the peaks at 652 or 716 cm⁻¹ for Ac-Phe-Ser-NH₂). A peak to peak detailed comparison shows an average 6 cm⁻¹ difference between experimental and theoretical peaks positions, calculated over all 9 dipeptides and over all peaks in the 90-800 cm⁻¹ range. This is for systems of this size (≃ 40 atoms) an excellent (maybe even impressive) result. A maximum deviation from experiment of about 20 cm⁻¹ is systematically obtained (in all systems) for the peak located at 740 cm⁻¹ experimentally and around 720 cm⁻¹ theoretically. This peak is part of a triplet related to ω(CH) wagging motions of the phenylalanine CH groups (see chapter 9) that is not conformational selective, as it is present at the same position for all systems investigated, see section 9.3.1 in chapter 9. All the vibrational assignments are supported by ICDOS analyses presented in chapter 9.

Theoretical bands are furthermore found as well resolved as the experimental ones, i.e. showing the same number of absorption peaks, the same narrow band-widths and the same band-shapes, in general also including subtle shoulders in some broader peaks. There are some exceptions in the 400-500 cm⁻¹ with less well resolved peaks in the calculations. Although experimental band-intensities are globally rather well reproduced in our dynamical spectra (without any adjustable parameter and/or model), some peaks however display either too low or too high intensities in comparison to the experiments. See for instance the peaks located around 700 cm⁻¹ for the A₁ conformation of the Ac-Phe-Val-NH₂ dipeptide or the γ-turn
CHAPTER 10. METHOD COMPARISONS

Figure 10.1: Experimental IR-UV ion dip spectra of the Ac-Phe-AA-NH₂ dipeptide series (in color) compared with the calculated dynamical DFT-MD/BLYP-D3 infrared spectra (in black). All the spectra intensities have been multiplied by a factor of 2 below 400 cm⁻¹ for the sake of clarity.

For this dipeptide series the DFT-MD/BLYP-D3 representation does an excellent job at reproducing the far infrared/THz infrared experimental spectra.

10.1.2 Phenol derivatives: Phenol, Catechol, Saligenin, Phenol-Water

Experimental and theoretical anharmonic dynamical DFT-MD/BLYP-D3 spectra for phenol derivatives are plotted in figure 10.2 in red and in black respectively. These systems are presented in section 2.2 in chapter 2.

Theoretical spectra appear, at first sight, to be in less good agreement with the experimental spectra than the spectra obtained for the dipeptide series presented in subsection 10.1.1. Three main reasons support this comment:

- One reason is that only one trajectory has been calculated for each of the phenol derivatives, phenol, catechol, saligenin and phenol-water (instead of an average over three
10.1. ANHARMONIC INFRARED SPECTRA CALCULATED FROM DFT-MD/BLYP-D3

Figure 10.2: Experimental IR-UV ion dip spectra of phenol derivatives (in red) compared with the calculated dynamical DFT-MD/BLYP-D3 spectra (in black).

trajectories for the dipeptides). Therefore the relative intensities of the DFT-MD infrared spectra are not expected to be as good as for the spectra of the dipeptides, due to a less good equipartition of energy within all vibrational modes. One can see, for example, for phenol that the $\omega$(OH) mode at 330 cm\(^{-1}\) in the DFT-MD infrared spectrum (at 309 cm\(^{-1}\) in the experimental spectrum) is found 9 times more intense than any other calculated peak in the theoretical spectrum. This would probably be improved by averaging more trajectories (it is our experience that averaging reduces such discrepancies).

- Secondly, some peaks are clearly missing in the DFT-MD/BLYP-D3 spectra. One can cite the 588 cm\(^{-1}\) in the experimental spectrum of phenol. Another example can be found in the saligenin experimental spectrum for which the peak at 586 cm\(^{-1}\) and the mas-sif between 607 and 643 cm\(^{-1}\) have no clear counterparts in the theoretical spectrum. The last example we can cite is the peak at 620 cm\(^{-1}\) and its shoulder at 635 cm\(^{-1}\) for the phenol-water complex experimental spectrum without counterpart in the DFT-MD spectrum.

One could have in mind to perform supplementary trajectories to see whether the peaks will show up and blame again a bad equipartition of energy in the modes whenever they do not show up in the dynamical spectra, but we do not believe this to be the correct explanation. A detailed isotopic substitution analysis, presented in figure 9.9 in chapter 9 for phenol, shows that the 588 cm\(^{-1}\) peak most probably arises from an overtone of the OH wagging motion or a combination band involving the same motion. Within the time length of the DFT-MD trajectory and with the amount of energy in the system, the DFT-MD is not able to reproduce these features. We believe that the other features missing in
the spectra of saligenin and phenol-water arise from the same phenomena, i.e. overtone or combination bands.

• Finally, some peaks displayed in the DFT-MD infrared spectra have no counterpart in the experiment. For example, the peak at 660 cm$^{-1}$ in the theoretical spectrum of phenol or the peaks at 538 and 577 cm$^{-1}$ in the theoretical spectrum of catechol. By decreasing the ratio signal/noise of the experimental spectra, one might find tiny contributions in front of the theoretical peaks. Again here, maybe these peaks are too intense in the DFT-MD spectra because of too much energy in these specific modes, i.e. deviating from equipartition (that might be also fixed by increasing the statistics over several trajectories).

Once the extra and missing peaks have been explained (or at least mentioned), we find a good agreement between DFT-MD/BLYP-D3 and experimental spectra with peak-to-peak mean frequency deviations of 8 cm$^{-1}$ for phenol, 10 cm$^{-1}$ for catechol, 8 cm$^{-1}$ for saligenin and 4 cm$^{-1}$ for phenol-water complex which is comparable to the 6 cm$^{-1}$ deviation obtained for the dipeptide series.

Using vibrational mode assignments presented in this thesis, we observe that the experimental/theoretical differences are really modes dependent. We further refer the reader to chapters 7 and 9 for vibrational mode assignments for phenol derivatives. We found notable deviations for the $\omega$(OH) and $\omega$(CH) local wagging modes: 20 cm$^{-1}$ for $\omega$(OH) in phenol, 12 and 4 cm$^{-1}$ for the two modes involving $\omega$(OH) motions in catechol, and for the $\omega$(CH) contributions, deviations of 8 and 17 cm$^{-1}$ for the $\omega$(CH) modes in catechol. The delocalised modes are almost perfectly well reproduced by the theory with 4 cm$^{-1}$ mean frequency deviation.

If we anticipate a little over section 10.3.3, which presents comparisons between anharmonic DFT-MD and harmonic spectra, we can already say that the most anharmonic modes, i.e. $\omega$(OH), are the ones sometimes suffering the largest deviations from experiments (max 20 cm$^{-1}$ shift!), while the harmonic delocalised modes are perfectly reproduced.

We can also say that non fundamental transitions (overtones and combinations band) are not reproduced by DFT-MD infrared spectra presented here and at the time of writing, we still do not know why this is the case. All fundamental transitions are very well reproduced.

### 10.1.3 DNA base pairs

We propose in figure 10.3 a comparison between the experimental spectra of guanine-guanine, guanine-cytosine and ethylated guanine - methylated cytosine base pairs (see their description in section 2.3 in chapter 2), plotted in blue, and the theoretical DFT-MD infrared spectra at the BLYP-D3 electronic level (plotted in black). To summarise the results we observe more missing peaks in the DFT-MD spectra with respect to the experimental spectra than for phenol derivatives that we explain by the presence of more overtones and/or combination bands. But once these missing peaks are excluded from the analysis, the agreement between theory and experiment remains fair.
10.1. ANHARMONIC INFRARED SPECTRA CALCULATED FROM DFT-MD/BLYP-D3

Figure 10.3: Experimental IR-UV ion dip spectra of three DNA base pairs (in blue) compared with the calculated dynamical DFT-MD/BLYP-D3 infrared spectra (in black).

We want our first discussion on the DNA base pairs to be purely restricted to the experimental spectra. The signatures of the experimental spectra are much broader than the signatures of the experimental spectra of the dipeptide series. All these spectra have been measured using the same experimental setup described in chapter 3 but despite that we believe that the temperature of the DNA base pairs for these spectra is higher than the temperature of the dipeptides for the spectra presented above. We have no explanation for this, especially because these base pairs are smaller than the dipeptide systems (large systems are more difficult to cool down). Another point that supports this hypothesis is that the 50K DFT-MD provide spectra as broad as the experimental spectra for the dipeptide series but dynamics at the same temperature provide features much thinner than the experimental counterparts for the DNA base pairs.

When comparing theoretical and experimental spectra, we observe three general cases for the experimental bands, that we will illustrate using the example of the guanine-guanine base pair, that corresponds to the worst scenario in terms of DFT-MD/experimental agreement/disagreement. In the following, the cases of the guanine-cytosine and ethylated guanine - methylated cytosine base pairs will not be discussed in details but the illustration on the guanine-guanine case can be transposed to the two other base pairs.

- Worst case is that an experimental peak or massif has no counterpart at all in the DFT-MD infrared spectrum. This is for example the case for the massif between 707 and 749 cm$^{-1}$ that has no match in the DFT-MD/BLYP-D3 spectrum.

- At the other end of the comparison scale, we have fortunately experimental peaks well reproduced by the theory. We can take the example of the peaks at 506 and 568 cm$^{-1}$ in
the experimental spectrum. These bands both arise from $\omega$(NH) wagging modes (see figure 9.16 in chapter 9 for vibrational mode assignments) and we know from the dipeptide series described above that the $\omega$(NH) modes are systematically well reproduced by the DFT-MD/BLYP-D3 spectra. We are not saying here that the reproduction of these features by DFT-MD is perfect. The DFT-MD peaks are redshifted with respect to the experimental spectrum by 5 and 13 cm$^{-1}$ respectively and the DFT-MD counterpart for the experimental peak at 568 cm$^{-1}$ is too intense. But we believe that the relative intensities between the two peaks would be improved by increasing the statistics over a higher number of trajectories and the frequency deviations observed here correspond to the average deviation observed for the dipeptide series (6 cm$^{-1}$) and we do not expect a better result for the DFT-MD spectra.

- Intermediate in the discussion is the case where a theoretical peak is either strongly shifted in frequency with respect to the experiment or too small in intensity. It is thus difficult to definitely assign a theoretical band to an experimental feature in these conditions. We can illustrate this case with the experimental intense double peak at 535 and 543 cm$^{-1}$. The only theoretical counterpart is a very small peak at 533 cm$^{-1}$. This is typically the case where extra dynamics are needed to increase statistics and improve the relative intensity between the bands in the DFT-MD spectrum and see if this feature will increase in intensity. This argument is related to the equipartition of energy not respected in all modes. Note again that only one trajectory of 20 ps has been calculated for DNA base pairs instead of three for the peptides.

As conclusion here the DFT-MD method shows some severe drawbacks for the base pairs spectroscopy with several experimental peaks without any counterpart in the theoretical spectrum. Once these peaks are excluded from analysis, a fair agreement remains between the experimental and the DFT-MD/BLYP-D3 spectra.

10.2 Basis set effects

Sections 10.3 and 10.4 will present comparisons between anharmonic DFT-MD, harmonic and anharmonic VPT2 infrared spectra. Harmonic and anharmonic VPT2 spectra are calculated with the gaussian package$^{109}$ employing a gaussian basis set only while the DFT-MD spectra are calculated using the CP2K package$^{122}$ using a dual basis set representation (both plane waves and gaussian basis sets) and pseudopotentials. Therefore, differences between DFT-MD and harmonic spectra (at the same functional level) may arise because of basis set effects, even though both basis sets employed for the two methods are chosen to be as equivalent as possible. To ensure the quality of the following comparisons, we first need to exclude basis set effects.

Figure 10.4 presents a comparison between two DFT-MD/BLYP-D3 spectra of Ac-Phe-Ser-NH$_2$ calculated with two different gaussian basis sets (and same plane wave basis set). The top spectrum is the same as presented in figure 10.1, averaged over three trajectories of 20 ps using the gaussian basis set aug-TZV2P (see chapter 4 for description of the method). For the
bottom spectrum, the larger basis set m-TZV2PX has been used. Note that only one trajectory of 20 ps has been used and therefore the repartition of energy in all vibrational modes is presumably less good for this second spectrum and may affect the quality of the intensities of the bands as already mentioned before. Relative intensities of the peaks between the two spectra should not be compared too much, we only discuss the positions of the bands.

![DFT-MD/BLYP-D3 infrared spectra of Ac-Phe-Ser-NH$_2$](image)

Figure 10.4: DFT-MD/BLYP-D3 infrared spectra of Ac-Phe-Ser-NH$_2$ calculated with two different gaussian basis sets (same plane waves basis set). The top spectrum has been calculated with the aug-TZV2P gaussian basis set and the bottom spectrum has been calculated with the m-TZV2PX gaussian basis set. The aug-TZV2P spectrum has been averaged over 3 trajectories of 20 ps each while the m-TZV2PX spectrum has been calculated from one single trajectory of 20 ps only. In all calculations, the plane wave basis set is set to 450 Ry.

In terms of band frequencies, both spectra are almost identical, only the peak related to the NH$_2$ symmetric motion (assignment supported by ICDOS spectra presented in chapter 9) is displaced between the two spectra, with a small blue shift when the size of the basis set is increased (peak located at 456 cm$^{-1}$ when the aug-TZV2P basis set is used and 462 cm$^{-1}$ when the m-TZV2PX basis set is used). By comparing both these theoretical spectra with the experimental spectrum that can be found in figure 10.1, one can observe that the change in the gaussian basis set has almost no impact on the final agreement with experiment.

As shown here, overall the basis set effect is small, as long as large enough basis sets are employed in the calculation. We always employ large basis sets for all the DFT-MD simulations/spectra calculations presented in this thesis. Therefore this bias can be excluded from the discussion of the following harmonic/DFT-MD and VPT2/DFT-MD comparisons, presented below in sections 10.3 and 10.4 respectively.
10.3 Harmonic BLYP-D3 infrared spectra

We have seen in section 10.1 that DFT-MD/BLYP-D3 spectra provide good results with respect to experimental data but this method is computationally expensive: DFT-MD is roughly 200 times more expensive than the harmonic approximation, when a 20 ps trajectory is accumulated. One therefore needs to identify what are the modes in need to take the anharmonicities into account and when it can be avoided. We present here a comparison between DFT-MD/BLYP-D3 and BLYP-D3 harmonic spectra. Section 10.2 has shown that the basis set effects are only small and therefore strong modifications between DFT-MD and harmonic spectra will arise most probably from anharmonic effects only (of course, using the same DFT functional in both calculations). We analyse our three preferred systems below, dipeptides, phenol derivatives, base pairs.

10.3.1 Dipeptide series

Figure 10.5 presents a comparison between DFT-MD/BLYP-D3 infrared anharmonic spectra in color and BLYP-D3 harmonic spectra in black for the Ac-Phe-'AA'-NH$_2$ dipeptide series. At first sight, both spectra appear to be rather similar in the whole 0-800 cm$^{-1}$ range for the whole series (same band positions, similar band shapes, rather similar intensities). While the 0-400 and 700-800 cm$^{-1}$ ranges (related to delocalised modes over the backbone and local $\omega$(CH) respectively, see chapter 9) present no/or few anharmonicities, we however observe some notable differences between the two calculations in the range 400-700 cm$^{-1}$.

Figure 10.5: Dipeptide Ac-Phe-AA-NH$_2$ series: calculated anharmonic DFT-MD/BLYP-D3 infrared spectra (in color) and BLYP-D3 harmonic infrared spectra (in black). All the spectra infrared intensities have been multiplied by a factor of 2 below 400 cm$^{-1}$ for the sake of clarity.
To allow a more precise comparison, three spectra have been plotted in figure 10.6 for three chosen systems, i.e. Ac-Phe-Val-NH$_2$ (A1 conformation), with the weakest hydrogen bond of this series (2.16 Å), Ac-Phe-Ser-NH$_2$ (hydrogen bond: 2.06 Å) and Ac-Phe-Pro-NH$_2$ (β-turn conformation) with the strongest hydrogen bond of this series (1.92 Å). Values given here for the hydrogen bond lengths are from the optimised geometries. See figure 9.18 in chapter 9 for the related average distances in the DFT-MD trajectories.

Figure 10.6: Ac-Phe-Val-NH$_2$ (A1 conformer), Ac-Phe-Ser-NH$_2$ and Ac-Phe-Pro-NH$_2$ (β-turn conformation): calculated DFT-MD/BLYP-D3 infrared spectra (in color) and BLYP-D3 harmonic infrared spectra (in black). All the spectra infrared intensities have been multiplied by a factor of 2 below 400 cm$^{-1}$ for the sake of clarity. Modes involving strong contributions of $\omega$(CH) are highlighted in grey, modes involving strong contributions of $\omega$(NH) are highlighted in cyan, modes involving strong contributions of $\omega$(NH$_2$) are circled in blue and modes involving strong contributions of $\omega$(OH) (of the OH group of the serine amino acid) are circled in red.

The three peaks between 700-800 cm$^{-1}$ highlighted in grey in figure 10.6 arise from $\omega$(CH) vibrational modes and their frequencies are the same between harmonic and anharmonic DFT-MD calculations. This is also true for the delocalised modes below 400 cm$^{-1}$, see chapter 9 for a description of these modes.

On the other hand, we observe some differences for the bands between 400 and 700 cm$^{-1}$. For both representations, $\omega$(NH) modes are highlighted in light blue, $\omega$(NH) modes of NH$_2$ are circled in dark blue boxes and $\omega$(OH) modes (for serine) are circled in red boxes.

The first thing that one can say before going further in the analysis is that the spectra differences between harmonic and anharmonic DFT-MD remain small in comparison for example with $\omega$(OH) wagging that can be shifted up to 100 cm$^{-1}$ in catechol between the anharmonic
DFT-MD and harmonic spectra (we further refer the reader to figure 10.8 in subsection 10.3.3 below where the catechol theoretical spectra are presented). The most important spectral difference for the three selected dipeptides here can be observed for Ac-Phe-Val-NH₂ (A1 conformer). The counterpart of the narrow intense peak at 465 cm⁻¹ in the harmonic spectrum is the broader massif inbetween 438 and 468 cm⁻¹ in the DFT-MD/BLYP-D3 spectrum. These features correspond to ω(NH) modes of the NH₂ terminal function. All other differences between harmonic and anharmonic DFT-MD spectra remain overall small with differences in band shapes, and frequency deviations inferior to 10 cm⁻¹. Nevertheless even though a deviation of 10 cm⁻¹ is very small, this result has to be put in perspective with the even fewer cm⁻¹ of deviation for the modes located below 400 cm⁻¹.

Despite equivalent harmonic and anharmonic DFT-MD signatures, we observe anharmonicities in these systems in terms of composition of the wagging modes in the range 400-700 cm⁻¹. In the three spectra, the NH waggings in NH₂ have the most anharmonic character, see the blue boxes that correspond to these motions in figure 10.6. This will be here illustrated with the case of the Ac-Phe-Pro-NH₂ (β-turn conformation) system for which we observe the biggest differences between the harmonic and anharmonic DFT-MD ω(NH) modes in NH₂ functions. The signatures of the NH₂ wagging are involved in the band at 464 cm⁻¹ and in all the modes between 494 and 648 cm⁻¹ in the harmonic spectrum, while these signatures are involved in the modes at 465, 488, 597 cm⁻¹ and in the massif between 615 and 632 cm⁻¹ in the BLYP-D3-MD spectrum. Note also here that there are infrared peaks in the DFT-MD spectrum overlapping with the peaks in the harmonic spectrum assigned to ω(NH) in NH₂: these DFT-MD peaks do not show substantial participation of the NH₂ wagging motions (see figure 9.18 in chapter 9 for ICDOS analyses). The NH₂ waggings are participating in a lower number of active peaks in the anharmonic DFT-MD spectrum than in the harmonic one. In other words, ω(NH) waggings in the NH₂ function are somehow more spatially localised in the dynamical representation, or can be seen as being more mixed with other motions in the harmonic representation.

We observe here that most of the anharmonicities come from the wagging motions. This is directly seen in figure 10.6 by the differences in frequency position for the ω(NH) waggings in NH and NH₂ functions (follow the light blue highlights and the dark blue boxes).

In subsection 10.3.2 below we study the Ac-Phe-OMe system that does not carry a NH₂ function to observe where the anharmonicities arise from.

### 10.3.2 Ac-Phe-OMe

Figure 10.7 presents the DFT-MD/BLYP-D3 spectrum of the Ac-Phe-OMe peptide in green and the harmonic one in black. This system has no NH₂ moiety and one can see immediately that the two spectra are almost identical. This "identical" term is not exactly true. Let me be more clear. In terms of position of the peaks and number of peaks; both infrared spectra look identical. In terms of band shapes and band intensities, this is not true anymore. One can see in particular that the DFT-MD infrared spectrum is more complex than the harmonic one with in particular several broader features.
10.3. HARMONIC BLYP-D3 INFRARED SPECTRA

Figure 10.7: Ac-Phe-OMe: the calculated DFT-MD/BLYP-D3 infrared spectrum in green and the BLYP-D3 harmonic spectrum (in black). The mode involving strong contributions of $\omega$(NH) is highlighted in blue.

In terms of band positions, we observe a very small deviation of 6 cm$^{-1}$ between the two methods for the $\omega$(NH) mode highlighted in blue (458 cm$^{-1}$ for the harmonic spectrum and 464 cm$^{-1}$ for the DFT-MD spectrum). We observe another difference between the two spectra for the mode at 274 cm$^{-1}$ in the harmonic spectrum. This mode corresponds to motions delocalised all over the backbone (harmonic and anharmonic alike). This feature is almost not shifted in the DFT-MD spectrum (located at 277 cm$^{-1}$) but is much broader, showing a more complex phenomenon than the harmonic view can provide. The broader peak in the DFT-MD spectrum improves the match with the experiment (see figure 10.12 for a comparison with the experimental spectrum).

10.3.3 Phenol derivatives

We present in this section further dynamical anharmonic/harmonic spectra comparisons for the phenol derivatives molecules. These systems do not carry the NH or NH$_2$ functions that we have seen above to be possibly the most anharmonic, but these systems are particularly interesting because perfectly suited to study OH functions in different environments (free, hydrogen bonded). Figure 10.8 presents the harmonic/BLYP-D3 (in black) and anharmonic DFT-MD/BLYP-D3 (in red) spectra for four phenol derivatives. Modes with $\omega$(OH) as main contributions are highlighted in red, modes with $\omega$(CH) as main contributions are highlighted in grey. The remaining modes correspond to delocalised modes over the phenol ring (in plane, out of plane deformations of the ring).
The first thing we can observe is that important spectral differences between harmonic and DFT-MD anharmonic spectra now show up, especially in terms of band positions. This is a first difference with respect to the peptides presented in sections 10.3.1 and 10.3.2 of this chapter. There is another difference. For peptides we had different compositions of the motions in the wagging modes between the two harmonic/anharmonic representations, while this is not the case here anymore for the OH wagging motion. The wagging motions are involved in the same features in the two methods but they are shifted in position. We are not saying here that the composition is exactly the same for all these modes, but looking at the main component of these wagging modes they are indeed identical. Other subtle differences might be seen in the couplings to other motions, but they are of no importance to our interpretation. We further refer the reader to the discussion in section 9.1 in chapter 9 about some limitations of the ICDOS tool (especially, we do not have a full quantitative view on the vibrational modes).

We will now go further in the discussion by analysing in details the spectral shifts between harmonic and DFT-MD anharmonic spectra:

- The delocalised modes (neither highlighted in red nor in grey in figure 10.8) are almost perfectly harmonic at the BLYP-D3 level. This is consistent with our observations in sections 10.3.1 and 10.3.2. We indeed have no significant shift in the position of these peaks between harmonic and anharmonic spectra. Shapes of these bands are also roughly identical in the two spectra.

- We find that OH waggings (red in figure 10.8) for the phenol derivative series have
stronger anharmonic behaviours than the NH waggings in the peptidic systems, analysed in sections 10.3.1 and 10.3.2 of this chapter. We hence observe a 25 cm\(^{-1}\) blueshift of \(\omega(\text{OH})\) in phenol in the anharmonic DFT-MD spectrum with respect to the harmonic one. In this case the harmonic feature (305 cm\(^{-1}\)) matches better the experimental peak (309 cm\(^{-1}\)) than does the DFT-MD one (330 cm\(^{-1}\)). For catechol the two harmonic peaks at 131 and 182 cm\(^{-1}\) are strongly (by \(\approx 100\) cm\(^{-1}\)) blue shifted in the anharmonic spectrum and are now located at 234 and 283 cm\(^{-1}\) in the DFT-MD spectrum. This shift makes the DFT-MD spectrum in much better agreement with the experimental one. The harmonic peak at 399 cm\(^{-1}\) splits into multiple peaks to form a massif inbetween 418 and 448 cm\(^{-1}\) in the DFT-MD spectrum. For saligenin the counterpart for the harmonic peak at 352 cm\(^{-1}\) is located at 349 cm\(^{-1}\) in the DFT-MD spectrum. The harmonic peak at 416 cm\(^{-1}\) is splitted into multiple peaks in the DFT-MD spectrum (at 396, 409 and 423 cm\(^{-1}\)). The harmonic mode at 692 cm\(^{-1}\) is red shifted and splitted into two peaks in the DFT-MD spectrum at 664 and 679 cm\(^{-1}\). For the saligenin case the harmonic spectrum does a better job than DFT-MD for reproducing the experimental spectrum in particular for the experimental peak at 690 cm\(^{-1}\). In any case, the dynamical representation shows the associated motion to be anharmonic (maybe too much anharmonic, therefore failing to match the experiment). For the phenol-water complex, the counterpart for the thin harmonic peak at 256 cm\(^{-1}\) is the broad massif inbetween 208 and 243 cm\(^{-1}\). We observe also a red shift for the \(\omega(\text{OH})\) mode located at 756 cm\(^{-1}\) that is red shifted to 695 cm\(^{-1}\) (in the shoulder) in the DFT-MD spectrum.

- Conversely, deviations for \(\omega(\text{CH})\) are inferior to 10 cm\(^{-1}\) between the harmonic and anharmonic DFT-MD spectra as already observed for the other systems in sections 10.3.1 and 10.3.2. The features are systematically blue shifted in the DFT-MD spectrum with respect to the harmonic one.

In conclusion, only the \(\omega(\text{OH})\) modes have a strong anharmonic character in the phenol derivatives while \(\omega(\text{CH})\) and the ring angles/dihedrals motions forming the more delocalised modes are identical (or almost) in both harmonic and anharmonic DFT-MD spectra. Even though a direct comparison between experimental and harmonic spectra are not presented in the figures above, we observed that the use of DFT-MD does not systematically improve the theory/experiment match for \(\omega(\text{OH})\) wagging modes in these phenol derivatives, sometimes over-/under-estimating the anharmonicity in these modes. There is however one case where the DFT-MD greatly improves the theory/experiment match in comparison with the harmonic spectrum: the strongly anharmonic features in catechol are blue shifted by \(\approx 100\) cm\(^{-1}\) between harmonic and anharmonic DFT-MD spectra. As already observed for \(\omega(\text{NH})\) waggings of the NH\(_2\) groups in the previous sections, such large anharmonicities are related to the more flexible hydrogen bonds in the molecules (to the rather weaker H-Bonds formed), i.e. therefore related to rather large amplitude motions. A harmonic representation is unable to provide a good modeling of such motions.
10.3.4 DNA base pairs

We have seen in subsection 10.1.3 that there is a poor agreement for the DNA base pairs analogues between the experimental and the DFT-MD/BLYP-D3 spectra. Several peaks are missing in the DFT-MD spectrum but the peaks reproduced by theory are in fair agreement with the experimental features, both in terms of band positions and band shapes.

In these systems, the two bases are linked through three hydrogen bonds. These three hydrogen bonds are based on:
- one NH function (N-H\cdots N=1.86 Å for guanine-guanine, =1.93 Å for guanine-cytosine and =1.89 Å for ethylated guanine - methylated cytosine, values obtained for optimised geometries).
- two NH₂ functions, one strongly hydrogen bonded (N-H\cdots O=1.75 Å for guanine-guanine, =1.79 Å for guanine-cytosine and =1.75 Å for ethylated guanine - methylated cytosine) and one a little less strongly bonded (N-H\cdots N=2.04 Å for guanine-guanine, N-H\cdots O=2.04 Å for guanine-cytosine and N-H\cdots O=1.89 Å for ethylated guanine - methylated cytosine). The given distances are from optimised geometries, see figures 2.20, 2.21 and 2.22 in chapter 2 for the 3D structures.

A comparison between MD and harmonic spectra at the BLYP-D3 level for the three base pairs (introduced in section 2.3 of chapter 2) is presented in figure 10.9. \(\omega(NH)\) modes involved in a hydrogen bond are highlighted in light green (the length of the hydrogen bond from geometry optimisation is written in light green as well) and \(\omega(NH)\) modes of the free NH functions are highlighted in dark green. \(\omega(NH)\) modes of NH₂ functions involved in the strong hydrogen bonds are circled in cyan (the length of the hydrogen bond from geometry optimisation is written in cyan) and \(\omega(NH)\) modes of NH₂ functions involved in the weaker hydrogen bond are circled in blue (the length of the hydrogen bond from geometry optimisation is written in blue). The vibrational mode assignments for DFT-MD spectra are supported by ICDOS analyses presented in figure 9.16 and 9.19 in chapter 9.

We observe only small anharmonicities for the NH waggings (highlighted in light and dark green in figure 10.9), with an average 8 cm\(^{-1}\) red shift for the DFT-MD signatures of the hydrogen bonded NH functions and a larger average 17 cm\(^{-1}\) blue shift for the DFT-MD signatures of the free NH functions (with respect to the harmonic spectrum). The \(\omega(NH)\) motions are involved in the same modes for both harmonic and anharmonic DFT-MD methods (same composition of the \(\omega(NH)\) modes).

As for the peptides systems presented in sections 10.3.1 and 10.3.2 of this chapter, the main source of anharmonicities in the spectra of the DNA base pairs are the NH₂ moities. For the peptides we mainly observed anharmonicities in terms of composition of the harmonic and anharmonic vibrational modes but almost no (or small) band shifts between the two methods. Here we will use the guanine-guanine base pair as example to show that we observe both spectral shifts and different mode compositions between the two methods, conclusions given for this base pair remain for the others. The vibrational mode assignments for the NH waggings of the two NH₂ functions can be found in figure 9.19 in chapter 9 and are labeled as \(\omega(G1_{Bond})\),
10.3. HARMONIC BLYP-D3 INFRARED SPECTRA

Figure 10.9: DNA base pairs and derivatives: calculated BLYP-D3-MD infrared spectra (in blue) and BLYP-D3 harmonic spectra (in black).

Modes involving strong contributions of \( \omega(\text{NH}) \) of NH functions involved in a hydrogen bond are highlighted in light green (the length of the hydrogen bond is written in light green), modes involving strong contributions of \( \omega(\text{NH}) \) of free NH functions are highlighted in dark green, modes involving strong contributions of \( \omega(\text{NH}) \) of NH\(_2\) functions involved in the strong hydrogen bonds are highlighted in cyan (the length of the hydrogen bond is written in cyan) and modes involving strong contributions of \( \omega(\text{NH}) \) of NH\(_2\) functions involved in the weaker hydrogen bonds are highlighted in blue (the length of the hydrogen bond is written in blue).

\( \omega(G_1^{\text{Free}}) \), \( \omega(G_2^{\text{Bond}}) \) and \( \omega(G_2^{\text{Bond}}) \). We further refer the reader to figure 2.20 in chapter 2 to find a structure of the guanine-guanine base pair as well as a definition of these labels.

Roughly, we see the same general conclusions appearing as for the previously discussed systems in the previous sections. \( \omega(\text{NH}) \) motions for the N-H groups are systematically found less harmonic than the \( \omega(\text{NH}) \) motions for the NH\(_2\) groups. One can observe from a few cm\(^{-1}\) to a maximum ~ 20 cm\(^{-1}\) shifts in the positions of the \( \omega(\text{NH}) \) bands inbetween the two representations. As seen for the other molecular systems presented above, \( \omega(\text{NH}) \) of free NH groups are the ones with the maximum deviations from harmonicity, because of possible large amplitude motion in the out-of-plane wagging, same for NH groups forming rather weak or intermediate H-Bonds, while the NH groups involved in the stronger H-Bonds are not as much flexible anymore, and therefore their signatures are found more harmonic. \( \omega(\text{NH}) \) motions for the NH\(_2\) groups are found strongly anharmonic, which is seen both for band positions, band shapes and composition of the modes. Here the two N-H of the NH\(_2\) group have coupled motions, and both motions are affected by the H-Bonds formed by each N-H as well as by the H-Bonds formed by the other. For the guanine-guanine base pair, we observe deviations...
from harmonicity from \( \sim 10 \) to \( 60 \) cm\(^{-1}\) in the position of the \( \text{NH}_2 \) wagging motions once anharmonicities are taken into account, which is representative of the other base-pairs as well.

In more details for the guanine-guanine base pair, we see that the \( \omega(G_{1\text{Bond}}) \) signatures (cyan circles) are located at 807 and 833 cm\(^{-1}\) in the harmonic spectrum and at 809 and 816 cm\(^{-1}\) in the DFT-MD spectrum. The \( \omega(G_{2\text{Bond}}) \) signatures (blue circles) are found in the peak at 612 cm\(^{-1}\) and in its shoulder at 618 cm\(^{-1}\) in the harmonic spectrum. \( \omega(G_{2\text{Bond}}) \) is only present in one feature in the DFT-MD spectrum at 606 cm\(^{-1}\). The \( \omega(G_{2\text{Free}}) \) signatures (blue circle) are found in only one feature at 389 cm\(^{-1}\) in the harmonic spectrum while we find two features at 372 and 381 cm\(^{-1}\) in the DFT-MD spectrum. The \( \omega(G_{1\text{Free}}) \) signatures (cyan circles) can be found in the double peak at 297 and 302 cm\(^{-1}\) in the harmonic spectrum and they are red shifted in the DFT-MD spectrum in a massif inbetween 245 and 275 cm\(^{-1}\).

All the modifications observed in the DFT-MD/BLYP-D3 with respect to the harmonic spectrum improve the match between theory and experiment, in particular for the massif between 245 and 275 cm\(^{-1}\) (DFT-MD spectrum) in the guanine-guanine base pair that corresponds to the less broad massif inbetween 237 and 251 cm\(^{-1}\) in the experimental spectrum, while no counterpart was found in the harmonic spectrum.

10.3.5 Summary of what we have learned up to now

This part of the chapter has allowed us to understand what vibrational modes have an harmonic character in their motions and what vibrational modes have an anharmonic character. To that end, we have compared harmonic and dynamical anharmonic DFT-MD spectra (same DFT functional used and basis set effects have been shown in this chapter to be non important considering the size used in the two calculations done here). Based on our knowledge (preceding chapters in this thesis and also first part of this chapter) that the DFT-MD calculated spectra are in rather good agreement with experiments for peptides, phenol derivatives molecules, and base pairs (although peaks are unfortunately missing in the calculated spectra of these base pairs with respect to experiments), we trust the vibrational anharmonicities given by the DFT-MD representation, and we hence discuss deviations from harmonicities by direct comparison of DFT-MD spectra to the harmonic spectra.

Among all the systems investigated here, only the wagging motions of hydrogen atoms provide remarkable vibrational anharmonicities, in particular the \( \omega(\text{OH}) \) modes and \( \omega(\text{NH}) \) in \( \text{NH}_2 \) groups. We hence observe band shifts for these modes between harmonic and DFT-MD anharmonic spectra. We also observe, especially for the \( \text{NH}_2 \) wagging motions, different compositions of these modes between harmonic and anharmonic spectra. In many cases, we have observed red shifts of these waggings in the DFT-MD spectra with respect to the harmonic ones. Also band-shapes of these waggings are sometimes seen modified in the anharmonic dynamical spectra, and their more complex and more broader shapes are generally in much better agreement with the experimental features.

To go deeper into the results unearthed, we have shown that the \( \omega(CH) \) wagging motions (700-900 cm\(^{-1}\)) are intrinsically harmonic in nature. Same is true for the backbone (in pep-
tides) and ring (phenol derivatives molecules, base pairs) deformations made of angular and
dihedral angular motions, below 400 cm\(^{-1}\). These motions are in no need of a anharmonic
vibrational representation, the harmonic representation provides a clear enough and non am-
biguous view.

This is not the case anymore for the wagging motions of N-H and O-H groups. \(\omega(OH)\) are
always found pretty anharmonic, one case shown here even leads to a 100 cm\(^{-1}\) blue-shift
of the position of the band once anharmonicities are taken into account. \(\omega(OH)\) waggings
are maybe the most anharmonic motions investigated in the systems presented in this thesis
work. \(\omega(NH)\) wagging motions arising from NH\(_2\) groups are competitors to the \(\omega(OH)\) wag-
ging motions for anharmonicities. \(\omega(NH)\) of N-H groups are found with far less anharmonic
motions.

Whether we are looking at \(\omega(NH)\) or \(\omega(OH)\) wagging motions, one general conclusion is
clear, only the larger amplitude motions of these groups are giving rise to large anharmonic-
ities. The large amplitude motions of the N-H or O-H groups are obtained (at least in our inves-
tigated systems, but I believe this is a generality) whenever these groups are either free of any
hydrogen bonding interactions or are involved in the most flexible hydrogen bonds, i.e. weak
and long H-Bonds. Once these groups are involved in strong hydrogen bonds, their motions
are by construction less flexible (more constrained) so that they loose anharmonicities. The
resulting wagging signatures are therefore more harmonic. In that latter case, band-positions
calculated at the vibrational harmonic level of theory would hence be enough to interpret the
experimental signatures. However harmonic the band-positions might be for the strongly H-
Bonded N-H or O-H groups, some subtle anharmonicities might however arise in the related
band-shapes of the modes, reflecting more complex motions and couplings than the harmonic
view might tell us. This has been seen in a few cases here.

In terms of band-positions of the N-H and O-H anharmonic wagging motions, we have
roughly seen deviations of the order 20-60 cm\(^{-1}\) from the harmonic behaviour. No real sys-
tematic red-shift or blue-shift has been seen. Depending on the community a spectroscopist
belongs to, this might be seen as a huge deviation (already the 20 cm\(^{-1}\)) or not so large (al-
though obviously a shift of e.g. 60 cm\(^{-1}\) in band-position is a lot, but maybe not 20 cm\(^{-1}\) !).
Maybe we can live with the smaller 20 cm\(^{-1}\) deviations for the interpretation of the experimen-
tal spectra, but it is hard to believe we can live with larger deviations. Even if we think we
can live with the smaller 20 cm\(^{-1}\) deviations for the interpretation of the experimental spectra,
it is not clear how much deviation one should expect from scratch. We have given here some
views related to the free/weakly H-Bonded/strongly H-Bonded N-H and O-H groups that can
obviously be used as guidelines in order to roughly pre-determine shifts from harmonicities.

If one calculates harmonic spectra only, using our guidelines for deviations for the NH and
OH wagging anharmonic motions, and tries to make a match with experimental data, one
might however ends-up with quite some troubles as soon as many of these NH and/or OH
oscillators are present in the molecular system. As we have not seen systematic red/blue-shifts
from harmonic band-positions, there will be quite a possible mess in the changes in positions
of the wagging modes resulting from anharmonicities as soon as many of these oscillators are
involved. My advice is thus to use DFT-MD for anharmonic spectra calculations, as we have seen that these are reliable for the wagging anharmonic motions.

10.4 Anharmonic infrared spectra calculated from VPT2/BLYP-D3

This section presents a comparison between anharmonic DFT-MD, harmonic and anharmonic VPT2 infrared spectra. The VPT2 spectroscopy presented in section 4.6 of chapter 4 is a method that calculates vibrational modes based on corrections to the harmonic spectrum, using the same definition of motions involved in the vibrational modes as in the harmonic approximation.

We have made VPT2/BLYP-D3 calculations for the Ac-Phe-Ser-NH$_2$ dipeptide, the Ac-Phe-OMe capped amino acid, four phenol derivatives and three DNA base pairs. For each series of systems, a figure presents a comparison between the experimental, the MD, the VPT2 and the harmonic spectra, all at the BLYP-D3 level and a second figure presents a comparison between the experimental spectrum, the VPT2 spectrum and a decomposition of this last spectrum into its fundamental ($\nu_0 \rightarrow \nu_1$), overtone ($\nu_0 \rightarrow \nu_2$) and combination transitions ($\nu_0 \rightarrow \nu_1$, $\nu'_0 \rightarrow \nu'_1$).

10.4.1 Case of Ac-Phe-Ser-NH$_2$ dipeptide and Ac-Phe-OMe peptide model

We know from section 10.3 that vibrational modes of the studied peptides are fairly harmonic in the far infrared, at the exception of the wagging motions of the NH$_2$ groups for which the harmonic/anharmonic differences are somehow small for the band shifts, and subtle modifications in the composition of the modes were observed. Both harmonic and anharmonic DFT-MD spectra were in a way in excellent agreement with the infrared experimental spectrum. We add here VPT2 spectra to our investigations but we restrict here the analysis to the Ac-Phe-Ser-NH$_2$ and Ac-Phe-OMe systems.

The first thing that we can observe in figure 10.10 is that the Ac-Phe-Ser-NH$_2$ fundamental VPT2 spectrum (only $\nu_0 \rightarrow \nu_1$ fundamental transitions) provides a less good agreement than the harmonic spectrum in terms of band positions, once compared to the experiment. One can see for example that all peaks in the VPT2 spectrum inbetween 400 and 800 cm$^{-1}$ are red shifted with respect to the experiment (up to $\simeq$15 cm$^{-1}$), while these peaks were perfectly reproduced by both harmonic and anharmonic DFT-MD spectra. On the other hand, frequencies of the delocalised modes below 400 cm$^{-1}$ are as well reproduced as in the two other theoretical methods. We found that the main issue in this VPT2 spectrum is the overestimation of band intensities. In particular, the peaks inbetween 540 and 560 cm$^{-1}$ (corresponding to coupled OH and NH waggings of the serine amino acid) are 200 times more intense than all the other far infrared bands (which can not be correct).

The strength of this methodology is however the calculation of both overtones and combination bands, as presented in figure 10.11. Only one combination band is intense enough to be visible in the total spectrum at 626 cm$^{-1}$. It is composed by the combination of the fundamen-
tal transitions located at 569 and 79 cm\(^{-1}\) in the harmonic spectrum. The mode at 569 cm\(^{-1}\) in the harmonic spectrum corresponds to a coupling between the OH and NH waggings (the NH function of the serine amino acid) and the mode at 79 cm\(^{-1}\) involves internal coordinates over the whole system (backbone deformations). The harmonic mode at 569 cm\(^{-1}\) is shifted at 550 cm\(^{-1}\) in the VPT2 spectrum and corresponds to the super intense peak discussed above. Therefore it is difficult to trust the intensity of this peak and the intensities of all bands in general and, it is still difficult to understand if the few missing peaks in the experimental spectrum (located at 652 or 716 cm\(^{-1}\)) are explained by combination bands.

Figure 10.10: Ac-Phe-Ser-NH\(_2\): the experimental spectrum (in blue) and the DFT-MD, total VPT2 and harmonic spectra (in black). For theoretical spectra, intensities have been multiplied by two below 400 cm\(^{-1}\) for clarity reasons.
Figure 10.11: Ac-Phe-Ser-NH$_2$: the experimental spectrum (in blue) and the total anharmonic VPT2 spectrum (in black) and decomposition of the total VPT2 spectrum into three spectra displaying respectively fundamental transitions (in purple), overtones (in green) and combination bands (in cyan). The total anharmonic VPT2 spectrum is the sum of these three spectra below (and divided by 2). Intensities of the theoretical spectra have been multiplied by two below 400 cm$^{-1}$ for clarity reasons.

For the Ac-Phe-OMe system, as can be observed in figure 10.12, no vibrational mode presents anharmonic features (already seen for band-positions by the comparison between DFT-MD and harmonic calculations in section 10.3.2, figure 10.7) and both harmonic/BLYP-D3 and DFT-MD/BLYP-D3 spectra are in excellent agreement with the experimental spectrum. The anharmonic VPT2 spectrum is red shifted with respect to the experimental spectrum (up to $\simeq$15 cm$^{-1}$, the same value as for the dipeptide series) in the range 400-800 cm$^{-1}$, and is as good as the two other methods in the range 0-400 cm$^{-1}$.

Figure 10.13 shows that there is no important contributions coming from overtones nor combination bands (VPT2 spectrum). In summary, if we use figure 10.12 to analyse the experiment/VPT2 spectra comparison, we find results similar to the ones obtained for Ac-Phe-Ser-NH$_2$, i.e. delocalised modes below 400 cm$^{-1}$ as well reproduced as in the other representations while the modes between 400 and 800 cm$^{-1}$ are red shifted with respect to the other theoretical spectra and also with respect to the experimental spectrum.
10.4. ANHARMONIC INFRARED SPECTRA CALCULATED FROM VPT2/BLYP-D3

Figure 10.12: Ac-Phe-OMe: the experimental spectrum (in green) and the MD, total VPT2 and harmonic spectra (in black). For theoretical spectra, intensities have been multiplied by two below 400 cm\(^{-1}\) for clarity reasons.

Figure 10.13: Ac-Phe-OMe: the experimental spectrum (in green) and the total VPT2 spectrum (in black) and decomposition of the total VPT2 spectrum into three spectra displaying respectively fundamental transitions (in purple), overtones (in green) and combination bands (in cyan). For theoretical spectra, intensities have been multiplied by two below 400 cm\(^{-1}\) for clarity reasons.
10.4.2 Case of phenol derivatives

Figure 10.14 presents a comparison between experiment, anharmonic DFT-MD, anharmonic VPT2 and harmonic spectra for four phenol derivatives. We will discuss the quality of the anharmonic VPT2 spectra for these systems hereafter. $\omega$(OH) modes are highlighted in red and $\omega$(CH) modes in grey in figure 10.14.

Let us start with the delocalised modes (neither highlighted in red nor in grey in figure 10.14). These modes are almost identical in the harmonic, anharmonic DFT-MD and anharmonic VPT2 spectra.

For the VPT2 spectra of peptides, see section 10.4.1, we observed a small systematic red shift for the $\omega$(CH) modes ($\simeq 10$ cm$^{-1}$) in the VPT2 spectrum with respect to the experimental one. We do not observe this systematic shift for phenol derivatives. As illustration we can use the case of phenol and catechol. For phenol, the harmonic signature at 492 cm$^{-1}$ is red shifted at 488 cm$^{-1}$ in the VPT2 spectrum and the two harmonic signatures at 661 and 727 cm$^{-1}$ are blue shifted to 698 and 731 cm$^{-1}$ in the VPT2 spectrum. In catechol, the harmonic band at 436 cm$^{-1}$ is blue shifted to 445 cm$^{-1}$ in the VPT2 spectrum and the two harmonic signatures at 717 and 747 cm$^{-1}$ are red shifted to 710 and 735 cm$^{-1}$ in the VPT2 spectrum. The $\omega$(CH) modes in the harmonic spectra are systematically red shifted with respect to the experiment. As we find sometimes red shifts or blue shifts of these modes in the VPT2 from the harmonic signatures, there is therefore no systematic improvement for matching the experiment.

The biggest modification between harmonic and VPT2 spectra arise for $\omega$(OH) modes and we observe that the VPT2 method overestimates the anharmonicities of these modes.

For phenol, subsection 10.3.3 showed that we had a blue shift of 25 cm$^{-1}$ in the DFT-MD spectrum with respect to the harmonic one (peak at 305 cm$^{-1}$ in the harmonic and peak at 330 cm$^{-1}$ in the DFT-MD spectrum). The band in the experimental spectrum is located at 309 cm$^{-1}$ which means that the band in the DFT-MD spectrum is already slightly over blue shifted. In the VPT2 spectrum the band is located at 347 cm$^{-1}$, 42 cm$^{-1}$ blue shifted with respect to the experiment.

For saligenin if we focus on the experimental $\omega$(OH) mode at 690 cm$^{-1}$ of the OH bonded group, we observe a red shift between the harmonic band at 692 cm$^{-1}$ and the two DFT-MD bands at 664 and 679 cm$^{-1}$. As for phenol the band is "over corrected" in frequency in DFT-MD and it is even worse for the VPT2 spectrum in which this $\omega$(OH) mode is located at 599 cm$^{-1}$, thus $\simeq 91$ cm$^{-1}$ too low in frequency from experiment.

To conclude the discussion about $\omega$(OH) modes, we will highlight the mode at 220 cm$^{-1}$ in the catechol experimental spectrum that has shown a huge anharmonic character: we indeed observe a blue shift of 102 cm$^{-1}$ in the DFT-MD spectrum with respect to the harmonic one (position of this peak at 131 cm$^{-1}$ for the harmonic and 233 cm$^{-1}$ for the MD spectrum). The same mode in the VPT2 spectrum is located at 1006 cm$^{-1}$, 786 cm$^{-1}$ blue shifted with respect to the experiment! Note that the band at 225 cm$^{-1}$ in the VPT2 spectrum with $\omega$(OH) contributions corresponds to the harmonic peak at 182 cm$^{-1}$.

The overestimation of anharmonicities in the VPT2 method is also true for the other $\omega$(OH)
Figure 10.14: Four phenol derivatives: the experimental spectrum (in red) and the DFT-MD, total VPT2 and harmonic spectra (in black). Modes involving strong contributions of $\omega(\text{CH})$ are highlighted in grey and modes involving strong contributions of $\omega(\text{OH})$ are highlighted in red.
motions of these four phenol derivatives and can be used as a general rule.

In subsection 10.1.2 we discussed the peak at 588 cm\(^{-1}\) in the experimental spectrum of phenol that is missing in the DFT-MD spectrum. An isotopic substitution experiment has been performed and is presented in figure 9.9 of chapter 9. This figure shows that the 588 cm\(^{-1}\) signature comes from one mode involving the hydrogen atom of the OH function. In this frequency range, the \(\omega(\text{OH})\) motion is the only one expected to involve this hydrogen atom. Therefore the overtone or a combination band involving \(\omega(\text{OH})\) is the most probable explanation for the presence of this peak.

Figure 10.15 presents now the decomposition of the VPT2 total spectra for the four phenol derivatives that we investigate here. For phenol, the only non fundamental transition intense enough to have a visible contribution on the total VPT2 spectrum is the \(\omega(\text{OH})\) overtone peak at 721 cm\(^{-1}\). If this assignment is correct, the theoretical band is 133 cm\(^{-1}\) blue shifted with respect to the experiment. However, the overtone is predicted by the VPT2 method while it is not in the DFT-MD one.

As conclusion, if we focus on fundamental transitions, DFT-MD/BLYP-D3 spectra for these four phenol derivatives are more reliable than the VPT2/BLYP-D3 spectra for a price marginally more expensive (DFT-MD is roughly twice more expensive as VPT2 calculations for a 20 ps trajectory). The only advantage of the VPT2 spectroscopy is the calculation of overtones and combination bands that are missing in our DFT-MD spectra. However we have seen that there is a huge frequency shift for phenol between VPT2 and experimental spectra for the overtone we assigned.

### 10.4.3 Case of DNA base pairs

We present in this subsection the VPT2 spectroscopy of the three DNA base pairs analogues introduced in section 2.3 of chapter 2 and already discussed in this chapter in sections 10.1.3 and 10.3.4. Figure 10.16 presents a comparison between experimental spectra in blue and anharmonic DFT-MD, anharmonic VPT2 and harmonic infrared spectra in black.

The first striking thing is that relative intensities between the peaks in the three VPT2 spectra are in poor agreement with respect to the experiment in the far infrared, with few peaks much more intense than all the others. These VPT2 theoretical peaks are analysed below for the three systems:

- **For guanine-guanine** we have two peaks at 217 and 246 cm\(^{-1}\) respectively 180 and 780 times more intense than any other peak in the far infrared. These modes are respectively located at 303 and 298 cm\(^{-1}\) in the harmonic spectrum and both assigned to the wagging motion of the free hydrogen atom of the strongly hydrogen bonded NH\(_2\) function.

- **For guanine-cytosine** we have four peaks at 392, 409, 472 and 481 cm\(^{-1}\) more than 60 times more intense than any other peaks in the far infrared. These modes are respectively located at 378, 384, 525 and 529 cm\(^{-1}\) in the harmonic spectrum. The modes at 378 and 384 cm\(^{-1}\) are both mainly composed by wagging of the free hydrogen atom of the
Figure 10.15: Four phenol derivatives: the experimental spectrum (in red) and the total VPT2 spectrum (in black) and decomposition of the total VPT2 spectrum into three spectra displaying respectively fundamental transitions (in purple), overtones (in green) and combination bands (in cyan). The total VPT2 spectrum is the sum of the three spectra (divided by 2).
Figure 10.16: Three DNA base pairs analogues: the experimental spectrum (in blue) and the DFT-MD, total VPT2 and harmonic spectra (in black).
strongly hydrogen bonded NH$_2$ function and the modes at 525 and 529 cm$^{-1}$ are both mainly composed of $\omega$(OH) motions of the OH function of cytosine coupled with NH$_2$ wagging motions of the NH$_2$ of guanine.

- For ethylated guanine - methylated cytosine we have two peaks at 256 and 361 cm$^{-1}$ respectively 15 and 48 times more intense than any other peak in the far infrared. These modes are respectively located at 387 and 381 cm$^{-1}$ in the harmonic spectrum and once again both assigned to the wagging motion of the free hydrogen atom of the strongly hydrogen bonded NH$_2$ group.

The overestimation of the intensities are systematically found for NH waggings of NH$_2$ groups, i.e. the most anharmonic modes (see section 10.3.4). We observed in VPT2 spectra of phenol derivatives an overestimation of the intensities for the $\omega$(OH) modes, i.e. the most anharmonic modes in phenol derivatives. As a rule, it appears that VPT2 overestimates the intensities of anharmonic bands.

Overall the match between VPT2 and experiment is less good than the match between DFT-MD and experiment in terms of intensity and frequency. Despite that, some missing peaks in the DFT-MD/BLYP-D3 spectrum are present in the VPT2 spectrum.

- The double peak located at 778 and 785 cm$^{-1}$ in the guanine-guanine experimental spectrum and missing in the DFT-MD/BLYP-D3 spectrum is present in the VPT2 spectrum, and located at 777 and 787 cm$^{-1}$. These two peaks correspond to the peaks located at 801 and 805 cm$^{-1}$ in the harmonic spectrum. The vibrational modes are both related to NH waggings of the hydrogen bonded hydrogen atom of the NH$_2$ function. In the DFT-MD spectrum these modes are located around 800 cm$^{-1}$.

If the double peak at 777 and 787 cm$^{-1}$ in the experimental spectrum indeed corresponds to $\omega$(NH) modes, they are better reproduced by anharmonic VPT2 spectroscopy than by the anharmonic DFT-MD and harmonic methods.

- The peak at 741 cm$^{-1}$ in the experimental spectrum of guanine-cytosine can be assigned to the peak at 756 cm$^{-1}$ in the VPT2 spectrum (768 cm$^{-1}$ in the harmonic spectrum). Counterparts (wagging from NH$_2$ functions) in the DFT-MD spectrum are located at 691 and 808 cm$^{-1}$. One issue here is these two DFT-MD peaks are already used to explain other experimental features (the peak at 675 cm$^{-1}$ and the large peak inbetween 776 and 791 cm$^{-1}$).

We also find the opposite phenomenon for the peak at 568 cm$^{-1}$ in the experimental spectrum of ethylated guanine - methylated cytosine perfectly matched by the DFT-MD and harmonic spectra (566 cm$^{-1}$ in the DFT-MD/BLYP-D3 and 565 cm$^{-1}$ in the harmonic spectrum) and however without any peak in the VPT2 spectrum. The peak is in fact present in the VPT2 spectrum, slightly shifted at 558 cm$^{-1}$ but it is invisible due to its low intensity in the VPT2 spectrum. The mode corresponds to in plane deformations of the cytosine ring.
Figure 10.17: Three DNA base pairs analogues: the experimental spectrum (in blue) and the total VPT2 spectrum (in black) and decomposition of the total VPT2 spectrum into three spectra displaying respectively fundamental transitions (in purple), overtones (in green) and combination bands (in cyan). The total VPT2 spectrum is the sum of the three spectra below divided by 2. The experimental peaks finding no counterparts in the DFT-MD spectra are highlighted in orange.
10.4. ANHARMONIC DFT-MD/B3LYP-D3 SPECTRA

Figure 10.17 presents the decomposition of the VPT2 spectra into the fundamental, overtone and combination transitions for the three base pairs investigated here. There is no overtone or combination bands intense enough to match the intensity of the fundamental transitions. The experimental peaks that have no counterparts in the DFT-MD spectrum are highlighted in orange and we see that none of the peaks can be explained by overtones or combination bands if we trust the intensities in the VPT2 spectra. VPT2 spectra do not in the end help us to give an interpretation of the mysterious peaks present in the experiments that are missing from the harmonic and anharmonic dynamical calculated spectra.

10.4.4 Summary

For all the systems investigated, VPT2 spectra provide theoretical spectra in less good agreement with respect to experiments than the DFT-MD ones, for a prize yet equivalent (DFT-MD is roughly twice more expensive as a VPT2 calculation for a 20 ps trajectory). The overtones and combination bands in the VPT2 spectra are found not intense enough with respect to the experimental features to provide unambiguous assignments. This is especially troubling for the experiment/theory comparison for the base pairs, where broad experimental peaks are still missing assignments.

10.5 Anharmonic DFT-MD/B3LYP-D3: test case for four phenol derivatives

Despite the quality of the DFT-MD/BLYP-D3 spectra with respect to the experiment as presented in section 10.1 in this chapter, one can always wonder how to improve even better our theoretical tool. Figure 10.18 now presents the experimental spectra of four phenol derivatives comparing DFT-MD/B3LYP-D3 and DFT-MD/BLYP-D3 anharmonic spectra. We test here whether a hybrid functional could perform better than a GGA functional when used in molecular dynamics. The B3LYP-D3 functional is described in subsection 4.1.6 in chapter 4. The same comparison could have been done at the harmonic level of course, but we wanted to stick to DFT-MD.

Experimental spectra are plotted in red while theoretical B3LYP-D3-MD and BLYP-D3-MD spectra are plotted in black. $\omega$(CH) modes are highlighted in light grey and $\omega$(OH) modes are highlighted in red. The vibrational modes assignments are supported by ICDOS analyses presented in figures 9.12 and 9.20 in chapter 9 for BLYP-D3 and figures 10.19 and 10.20 for B3LYP-D3. The remaining bands (neither highlighted in red nor in grey) correspond to delocalised modes (in plane, out plane deformations of the ring).

The first thing one can observe for the four systems is that the amount of peaks is the same for the two theoretical spectra which means that the peaks missing (and discussed in section 10.1.2 of this chapter) in the DFT-MD/BLYP-D3 spectra are not reproduced either by using the B3LYP-D3 functional. It thus does not seem to be a functional issue.
Figure 10.18: Four phenol derivatives: experimental spectra (in red) and calculated DFT-MD/B3LYP-D3 and DFT-MD/BLYP-D3 infrared spectra (in black). Modes involving strong contributions of $\omega(CH)$ are highlighted in grey and modes involving strong contributions of $\omega(OH)$ are highlighted in red. Vibrational mode assignments for the B3LYP-D3-MD and BLYP-D3-MD spectra are supported by ICDOS analyses presented in figure 10.19 and in figure 9.12 in chapter 9 for the $\omega(CH)$ motions and in figure 10.20 and in figure 9.20 in chapter 9 for the $\omega(OH)$ motions.
The modes involving $\omega$(OH) are almost not shifted in position between the two functionals in all four phenol derivatives. Only the $\omega$(OH) signatures of the hydrogen bonded OH group of saligenin present small red shifts in the B3LYP-D3-MD spectrum with respect to the BLYP-D3-MD spectrum. Note that experimentally, only one mode involving $\omega$(OH) signature of the hydrogen bonded OH function has been located at 690 cm$^{-1}$ (using an isotopic substitution experiment, not presented in this thesis). The two $\omega$(OH) peaks are found at 664 and 679 cm$^{-1}$ in the DFT-MD/BLYP-D3 spectrum and at 658 cm$^{-1}$ (+ the shoulder at 653 cm$^{-1}$) in the DFT-MD/B3LYP-D3 spectrum. It shows that in terms of frequency the BLYP-D3 representation provides a better match for this peak. We also observe a shift for the $\omega$(OH) signatures of the H$_2$O molecule in the phenol-water complex (176-234 cm$^{-1}$ for B3LYP-D3 and 205-245 cm$^{-1}$ for BLYP-D3) but the experimental spectrum has not been measured below 220 cm$^{-1}$ (it is thus complicated to learn things from this comparison). We found that the addition of Hartree-Fock exchange on the BLYP-D3 functional does not improve much the $\omega$(OH) signatures.

The $\omega$(CH) signatures in the DFT-MD/B3LYP-D3 are systematically blue shifted with respect to the DFT-MD/BLYP-D3 spectrum. The $\omega$(CH) signatures are slightly red shifted with respect to the experiment in the DFT-MD/BLYP-D3 spectra (average mean red shift of 8 cm$^{-1}$) and the same signatures are now blue shifted in the DFT-MD/B3LYP-D3 spectra (average blue shift of 22.5 cm$^{-1}$).

Finally, in the DFT-MD/B3LYP-D3 spectra we observe a small blue shift for the delocalised modes (neither highlighted in red nor in grey in figure 10.18) with respect to the DFT-MD/BLYP-D3 spectrum. The $\omega$(CH) signatures is almost perfect for reproducing these delocalised modes when compared to experiments (average mean red shift of 1 cm$^{-1}$) and the same modes are blue shifted in the DFT-MD/B3LYP-D3 spectrum (average blue shift of 5 cm$^{-1}$ in comparison with the experiments). Both DFT-MD/BLYP-D3 and DFT-MD/BLYP-D3 are however excellent for these modes.

In conclusion, the BLYP-D3 functional used in combination with molecular dynamics provide better results than DFT-MD/B3LYP-D3 spectra, when the theoretical spectra are compared to experiments. Put in different words, using the hybrid B3LYP functional in the DFT-MD does not seem to provide any substantial improvements in the vibrational features from the ones obtained with the GGA BLYP functional, sometimes even the opposite (deteriorates the agreement).

The main difference between these two functionals is the inclusion of Hartree-Fock exchange in the B3LYP functional and for optimal results, it might be interesting to use a different amount of Hartree-Fock exchange than the one used here. There is also an inclusion of the VWN correlation function and of the LDA exchange functional (19 and 8 % respectively in B3LYP) that might be re-parametrised for more optimal results for infrared spectra. Of course one has to be careful when changing the parameters of a functional because even though we could observe an improvement in comparison with the phenol derivatives experimental spectra, it might also decrease the quality of the theoretical spectra for other systems. A first interesting step would be to use the DFT-MD/B3LYP-D3 method on other systems such as the
peptides investigated in this thesis. It has not be done here due to lack of time. It would also be interesting to try other functionals that have not been tested in the work such as PBE\textsuperscript{155} or PBE0\textsuperscript{153,154} its hybrid counterpart.

10.6 Conclusions and outlook

We have seen in this chapter that the DFT-MD/BLYP-D3 spectra are in rather good agreement with the experimental spectra for the systems presented here. In particular for peptidic systems for which all the major experimental bands are reproduced with an average deviation of 6 cm\textsuperscript{-1} by the dynamical theoretical spectra. The match is maybe less good for the four phenol derivatives and DNA base pairs at first sight, for which major contributions that we assume to correspond to overtones or combination bands are missing in the DFT-MD spectra. Once one forgets about these overtones/combination bands, the DFT-MD theoretical spectral features are within 10-20 cm\textsuperscript{-1} of the experiments, which is rather excellent (at least for our spectroscopic community). We have still more difficulties with the spectra of the base-pairs, where mysterious missing bands are not predicted by neither the DFT-MD, nor the harmonic, nor the VPT2 calculations (we believe the 3D structures are the right ones).

Comparisons between DFT-MD/BLYP-D3 and BLYP-D3 harmonic spectra have been presented and the OH groups (here present in the phenol derivatives molecules) have been identified as the most anharmonic moieties (with different degrees of anharmonicities for the different OH functions). The \(\omega(\text{NH})\) wagging motions of the \(\text{NH}_2\) functions in peptides and base pairs are the second source of large anharmonicities. This is especially crucial for the DNA base pairs where many of these H-Bonds are present. Anharmonicities have been shown here for band-positions and band-shapes, and also for the composition of the modes. All other \(\omega(\text{CH})\) wagging modes and backbone/ring angular/dihedral angular motions have been found to behave harmonically.

We have provided some guidelines for the anharmonicities of the \(\omega(\text{OH})\) and \(\omega(\text{NH})\) wagging motions in relation to flexible and non-flexible OH and NH groups. Whenever these groups are free of hydrogen bonds or involved into weak hydrogen bonds, these groups are flexible in their motions and large amplitude motions of these groups are allowed. These large amplitude motions are anharmonic and require an anharmonic vibrational treatment in order to predict their band-positions and band-shapes. Red/blue-shifts from the harmonic behaviour can not be predicted in advance. Whenever these OH/NH groups are involved in strong hydrogen bonds, their motions are somehow constrained by these H-Bonds and therefore reduced in flexibility, so that their vibrational signatures will be harmonic. A vibrational harmonic treatment of these groups will be enough to get their vibrational signatures.

We have shown that VPT2 spectra provided theoretical anharmonic spectra of less good quality than the DFT-MD spectra when compared to experiments for a prize yet equivalent (DFT-MD is roughly twice more expensive as VPT2 calculation for a 20 ps trajectory). The match with experiment is less good than the one provided by MD for all the systems investigated here. The overtones and combination bands provided by the VPT2 method (and absent in
Figure 10.19: Four phenol derivatives: (from top to bottom) the infrared experimental spectrum (in red), theoretical DFT-MD IR spectrum (in black) and ICDOS signatures of spectra of the dihedral angles $\Phi = \text{CCCH}$ (in brown). Theoretical spectrum and ICDOS spectra are calculated at the B3LYP-D3 level. The same decomposition is presented for the BLYP-D3 functional in figure 9.12 in chapter 9.
Figure 10.20: Four phenol derivatives: (from top to bottom) the infrared experimental spectrum (in red), the theoretical DFT-MD IR spectrum (in black) and the ICDOS signatures of the $\omega$(OH) (in brown). The four systems selected, i.e. phenol, catechol, saligenin and phenol water are presented in section 2.2 and their structures are presented in figure 2.15. The lengths of the hydrogen bonds is written and extracted from molecular dynamics. When the OH function is not engaged into a hydrogen bond (for the phenol) the label "Free" is written. The indices 1 and 2 refer to the position of the OH function on the aromatic ring, see figure 2.16. Theoretical spectrum and ICDOS spectra are calculated at the B3LYP-D3 level. The same decomposition is presented for the BLYP-D3 functional in figure 9.20 in chapter 9.
10.6. CONCLUSIONS AND OUTLOOK

all the DFT-MD spectra calculated here) are however very useful to explain experimental peaks missing in our theoretical DFT-MD spectra, but we found many times that these transitions are not intense enough in the VPT2 spectra (in comparison with the fundamental transitions) and assignments of these bands to experimental features still remain difficult and ambiguous.

We have also tested if a hybrid functional like B3LYP could provide any improvement with respect to the DFT-MD/BLYP-D3 spectra. We have tested here DFT-MD/B3LYP-D3 infrared spectra for four phenol derivatives. The hybrid electronic representation was not found to provide better or worse spectra than the GGA one.

I would like to finish this part of methods comparisons by saying that I believe that the DFT-MD dynamical anharmonic spectra calculated in the far infrared, as presented here, are robust, provide spectra in good to rather excellent agreements with the experimental far infrared spectra, and allow us to make a one-to-one comparison with the experiments without thinking twice. We can hence make the link between spectroscopic signatures and 3D structures without too many ambiguities.

If one calculates harmonic spectra instead, using our guidelines for deviations for the NH and OH wagging anharmonic motions, and tries to make a match with experimental data, one might think that this will work fine, but I believe one will definitely end-up with quite some troubles as soon as many of these NH and/or OH oscillators are present in the molecular system. As we have not seen systematic red/blue-shifts from harmonic band-positions, there will be quite a possible mess in the changes in positions of the wagging modes resulting from anharmonicities as soon as many of these oscillators are involved. This will lead to many ambiguities in comparing the harmonic spectra of different isomers and the experimental spectrum, not allowing a definitive and unambiguous assignment of the 3D conformation. My advice is thus to use DFT-MD for anharmonic spectra calculations, as we have seen that these are reliable for the wagging anharmonic motions. VPT2 anharmonic spectra (as calculated with Barone’s method) have not been found here to be of substantial help in making the one-to-one match to experiments in the far-IR, for a computational price equivalent to DFT-MD spectroscopy. Better to perform DFT-MD spectroscopy. The only downcast right now in the whole comparisons shown here is the clear missing of overtones/combination bands from our DFT-MD spectra, although in principle these modes anharmonicities should be present. The VPT2 method is slightly better in this respect, although still missing some of these bands and/or underestimating so much their IR intensities that it is hard to use these calculations to interpret these specific bands.

I believe this chapter would be worth transforming into a paper. It has not been done yet.
Chapter 11

A final illustration of the importance of far infrared/tera-hertz spectroscopy: conformational assignment of the \((\text{Ac-Phe-OMe})_2\) dimer.

As already said several times in this manuscript, characterising the three dimensional conformational arrangement of molecules is one of the main goals in chemical-physics and analytical chemistry, and in that respect vibrational spectroscopy is one of the key player in the domain. We believe that the far infrared spectroscopy is opening new opportunities in conformational assignment for systems for which the exact structure was still not accessible in the gas phase\(^{86,87}\) using other spectral domains. I believe I have demonstrated this point already in my whole thesis work, and here I take one more example, the Ac-Phe-OMe capped amino acid and its dimeric form \((\text{Ac-Phe-OMe})_2\).

This example is important. Indeed, for most of the systems investigated in this thesis, the three dimensional structure was already well known in literature. This was the case for the dipeptides series and DNA base pairs, investigated in the range 3000-4000 cm\(^{-1}\) or for phenol derivatives for which the conformational assignment is not especially challenging due to the small size of the systems. The investigation of these already known systems was important for building the vibrational mode mapping and for the method comparisons (all presented in chapters 9 and 10 respectively) so as to exclude any deviation between theoretical and experimental spectra to come from wrong conformational assignments. In this chapter we want to go back to the purpose of the infrared vibrational spectroscopy, \textit{i.e.} unravel the structure of molecular systems, using the far infrared/THz spectral domain.

Such conformational assignments of saligenin-water cluster\(^{111}\) has been presented in chapter 8 of this thesis. In this chapter we choose the Ac-Phe-OMe capped amino acid and its dimeric form \((\text{Ac-Phe-OMe})_2\) for two reasons:
The conformational assignment of the dimeric form appears to be particularly challenging by using the classical ranges 1000-2000 and 3000-4000 cm$^{-1}$ leaving one with only the information about the global family structure, i.e. a $\beta$-Sheet structure as shown by Gerhards's group in the last decade$^{69-71}$. A generic scheme of the $\beta$-sheet structure is presented in figure 11.1. We show in this chapter that using the far infrared spectroscopy, we manage to assign unambiguously the exact conformation of the dimer.

(Ac-Phe-OMe)$_2$ is probably the smallest peptidic system that adopts a $\beta$-sheet structure, a common structure for large size proteins and peptides. One of our goals is to provide specific signatures of protein motifs in the far infrared-THz domain. Characterising (Ac-Phe-OMe)$_2$ is a first step in this direction. Spectroscopists put a special emphasis on $\beta$-sheets structures because this generic structure is known to be involved into neurodegenerative diseases like Alzheimer, Parkinson or Huntington, having a common molecular basis related to the misfolding of proteins$^{206,207}$. 

![Figure 11.1: Generic scheme for an antiparallel $\beta$-sheet structure.](image-url)
11.1 Conformational assignment of the monomer: Ac-Phe-OMe

Before investigating the (Ac-Phe-OMe)$_2$ system, we will start with the monomer Ac-Phe-OMe. Our REMPI spectrum of the gas phase Ac-Phe-OMe molecule is presented in figure 11.2. The spectrum is nearly identical to the one published by Gerhards and Unterberg$^{69}$, only the peaks relative intensities are different in our two works. Our IR-UV ion dip infrared spectrum (fig. 11.6) has been measured via the $S_0\rightarrow S_1$ transition at 37579 cm$^{-1}$, corresponding to the dominant conformation in the UV spectrum. This is the only conformer investigated in this work and in ref.$^{69}$.

![Figure 11.2: REMPI spectra of the Ac-Phe-OMe monomer (in purple) and its dimeric form (Ac-Phe-OMe)$_2$ (in green).](image)

A conformational search for the Ac-Phe-OMe monomer has been performed through a 100 ps molecular dynamics using the semi empirical PM6 representation$^{142}$ at 1000K. The time step chosen is 0.4 fs. Hundred conformations have been extracted and optimised at the BLYP-D3/6-311+G(d,p) level using the Gaussian package$^{109}$. We can of course never be sure that the whole potential energy surface has been probed in this way and that we provide the exhaustive sampling of conformations. Among the 100 initial structures extracted from the dynamics, only 21 different ones remain after geometry optimisations. The electronic energies ($\Delta E$) and Gibbs free energies ($\Delta G$) of the five lower energy conformers of Ac-Phe-OMe have been calculated with the BLYP-D3/6-311+G(d,p) and B3LYP-D3/6-311+G(d,p) DFT electronic representations using the (all electron) Gaussian package$^{109}$. Values are gathered in table 11.1. A cut-off in terms of free energy has been arbitrary fixed around 2kcal.mol$^{-1}$ to select the conformations for further investigations. Note that the next conformer in terms of free energy lies at 3.07
kcal.mol\(^{-1}\) at the BLYP-D3/6-311+G(d,p) level.

<table>
<thead>
<tr>
<th>Ac-Phe-OMe</th>
<th>BLYP-D3/6-311+G(d,p)</th>
<th>B3LYP-D3/6-311+G(d,p)</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>∆E 0.99</td>
<td>∆G 50K 1.17</td>
<td>∆E 0.93</td>
<td>∆G 50K 1.09</td>
</tr>
<tr>
<td>1.13</td>
<td>1.31</td>
<td>1.10</td>
<td>1.25</td>
</tr>
<tr>
<td>1.66</td>
<td>1.75</td>
<td>1.68</td>
<td>1.78</td>
</tr>
<tr>
<td>2.02</td>
<td>2.18</td>
<td>1.71</td>
<td>1.91</td>
</tr>
</tbody>
</table>

Table 11.1: Relative electronic energies (∆E) and free energies (∆G) for the different conformers of Ac-Phe-OMe investigated here (in kcal.mol\(^{-1}\)). ∆G is calculated at 50 K, and results given by two levels of electronic representations are provided.

The two lower energy conformers are β\(L(g^+\)) and β\(L(a)\) with ~1 kcal/mol difference in energy (∆G and ∆E, BLYP and B3LYP electronic representations). β\(L\) refers to the backbone amide φ and ψ angles (see scheme 11.3) within the following ranges in the Ramachandran plot, φ ~[-120°,+120°], ψ ~[-120°,+120°], corresponding to a β-sheet type organisation of the backbone. There is a C5 interaction between the N-H\(_{Phe}\) and the C=O amide backbone groups, i.e. a weak N-H\(_{Phe}\)···O=C hydrogen bond interaction leading to the formation of an internal five membered ring between these two groups. g\(^+\), a and g\(^-\) refer to the orientation of the Phenylalanine (Phe) aromatic ring with respect to the backbone, defined by the dihedral angle \(\chi_1\) (see scheme 11.3, \(\chi_1 = N-C_{\alpha}$-C\(_{\beta}$-C\)): \(\chi_1 = 60^\circ\) for β\(L(g^+\)), 180° for β\(L(a)\) and -60° for β\(L(g^-\)). This latter β\(L(g^-)\) conformation is found roughly 2 kcal/mol (∆G and ∆E) higher in energy than the β\(L(g^+)\) lowest energy conformer. Sandwiched inbetween the three β\(L\) type conformers of Ac-Phe-OMe, one can find other symmetries of the δ\(D\) and γ\(L\) type, i.e. with φ and ψ angles in the ranges [-120°,+120°] & [-120°,0°] for δ\(D\) and [-120°,0°] & [0°,+120°] for γ\(L\) (γ-turn). These structures also have an internal N-H\(_{Phe}\)···O=C C5 interaction. In the previous work by Gerhards and coworkers\(^{69}\), only the two lower energy conformers β\(L(g^+)\) and β\(L(a)\) have been investigated (the same list of conformations has been found by Gerhards and coworkers), under the assumption that only the lower energy conformers could be formed in the experiment. The three β\(L\) conformers are depicted in figure 11.4. Relative energies are similar to ref.\(^{69}\). In β\(L(g^+)\), the C5 H-bonded N-H\(_{Phe}\) and C=O groups are located roughly above the Phe ring (but not interacting with it), only the C=O in β\(L(a)\), and these two groups are pointed outwards the Phe ring in β\(L(g^-)\).

Previous spectroscopic results from Gerhard’s group\(^{69,70}\) show that none of the signatures in the 3000-4000 and 1450-1800 cm\(^{-1}\) ranges are able to distinguish β\(L(g^+)\) and β\(L(a)\) conformations, while the latest published results in the fingerprint region indicate that the 1000-1450 cm\(^{-1}\) might be conformer selective\(^{71}\). We reproduce here in the 1000-2000 cm\(^{-1}\) fingerprint region a comparison between experiment (remeasured in the present work) and (scaled) static harmonic spectra (B3LYP-D3/6-311+G(d,p) level) for the five optimised structures of Ac-Phe-OMe (table 11.1). This is presented in figure 11.5.

One can see that the five conformers provide very similar signatures in the 1400-1800 cm\(^{-1}\) while the 1100-1400 cm\(^{-1}\) signatures offer slightly more conformational selectivity. In partic-
Figure 11.3: Chemical scheme of Ac-Phe-OMe monomer with the labels used in the text for the atoms and the definition of dihedral angles (φ, ψ, ω and χ) that correspond to the Ramachandran notation\textsuperscript{110}. The caps 'Ac' and 'OMe' correspond respectively to an acetyl group CH\textsubscript{3}-CO and to a O-CH\textsubscript{3} group.

Figure 11.4: Ac-Phe-OMe and (Ac-Phe-OMe)\textsubscript{2} peptides optimised geometries in the β\textsubscript{L} general structural organisation. The "β\textsubscript{L}(g\textsuperscript{+})" conformation corresponds to the assigned one for the monomer (in our experimental conditions). The "β\textsubscript{L}(g\textsuperscript{+})β\textsubscript{L}(g\textsuperscript{-})" dimer structure corresponds to the assigned one (in our experimental conditions). Free energies have been calculated at 50K at the BLYP-D3/6-311+G(d,p) level. See scheme 11.3 for more notations.
Figure 11.5: Experimental IR-UV ion dip spectra of the Ac-Phe-OMe capped amino acid (in black) compared with the calculated static harmonic B3LYP-D3/6-311+G(d,p) spectra (in color). The theoretical spectra are scaled by 0.967 and a 1% gaussian convolution is applied.

ular, the experimental broad peak at 1209 cm\(^{-1}\) and the smaller peak at 1246 cm\(^{-1}\) provide a clear marquer in favour of assigning the two lower energy conformers \(\beta_L(g^+)\) and \(\beta_L(a)\), with \(\beta_L(g^-)\) maybe providing a global better agreement with experiment. Bands that fit these two experimental features correspond to modes composed of CCH and CNH bendings and CC, CO and CN strechings (roughly the Amide III vibrational bands). All other conformers, including the \(\beta_L(g^-)\) conformer from the same \(\beta\)-sheet family, have too many dispersed signatures in this domain, not matching the experimental patterns.

We now provide in figure 11.6 the far-IR experimental spectrum of Ac-Phe-OMe gas phase peptide (100-800 cm\(^{-1}\)) together with dynamical DFT-MD spectra (averaged over three trajectories of 20 ps) for the three \(\beta_L(g^+)\), \(\beta_L(a)\), \(\beta_L(g^-)\) conformations. We choose here to present only the signatures of the three \(\beta_L\)-type structures, as according to the above comparison in the 1000-2000 cm\(^{-1}\) fingerprint region, one expects \(\beta_L(g^+)\) and \(\beta_L(a)\) to be good conformer candidates, while \(\beta_L(g^-)\) conformer (as well as \(\delta_D\) and \(\gamma_L\)) is not.

Based on number of bands and absolute positions of the theoretical bands (positions are shifted by 7 cm\(^{-1}\) on average from experiment), on their shapes and relative intensities (both
11.1. CONFORMATIONAL ASSIGNMENT OF THE MONOMER: AC-PHE-OMe

Figure 11.6: Experimental IR-UV ion dip spectrum of the Ac-Phe-OMe gas phase monomer (in blue) compared with the calculated dynamical DFT-MD IR spectra (in red) for the three monomeric isomers $\beta_L(g^+)$, $\beta_L(a)$, and $\beta_L(g^-)$ in the far infrared/THz domain. The green boxes highlight missing spectral features in the theoretical spectra that are however present in the experimental one. Conversely, the orange boxes highlight extra spectral features in the theoretical spectrum that are absent in the experimental one. The blue boxes show the assignment of the $\omega(NH)_{Phe}$ wagging motion.

coming from the dynamics, no *a posteriori* model applied), one can conclude that $\beta_L(g^+)$ is the conformation indeed formed in the experiment. Especially well given by the theoretical spectrum are the 680-750 cm$^{-1}$ ($\omega(CH)_{Ring}$ waggings) and 550-620 cm$^{-1}$ (backbone and side-chain bending motions around the $C_\alpha$ link) double bands, the 450-510 cm$^{-1}$ massif ($\omega(NH)_{Phe}$ wagging). Same very good agreement below 400 cm$^{-1}$, with however some theoretical bands missing some intensity. The 260-300 cm$^{-1}$ double band massif (coupled bending motions along the backbone and around $C_\alpha$-$C_\beta$-$C$ side chain) is remarkably well given by the theoretical spectrum of $\beta_L(g^+)$ (the higher frequency band is slightly shifted in position), while the two other conformations display different active band-shapes in this domain. Note also that the theoretical band-widths provide as well defined and resolved bands as in the experiment in the whole far-IR domain.

Highlighted in green in figure 11.6, one can see missing peaks (or peaks too much shifted in position) in the theoretical spectra that are present in the experimental spectrum, in disfavour of $\beta_L(a)$ and $\beta_L(g^-)$ conformational assignment. We observe no counterpart for the experimental peak at $135\text{cm}^{-1}$ in the $\beta_L(g^-)$ theoretical DFT-MD spectrum. For the experimental peak at $100\text{cm}^{-1}$ neither $\beta_L(a)$ nor $\beta_L(g^-)$ provide counterpart.

Highlighted in orange in figure 11.6 are extra peaks present in the theoretical spectra of $\beta_L(a)$ and $\beta_L(g^-)$ conformations that are absent in the experimental spectrum. For both $\beta_L(a)$ and $\beta_L(g^-)$ a theoretical peak at $\simeq 345\text{ cm}^{-1}$ find no counterpart in the experiment.
Finally the experimental peak at 710 cm\(^{-1}\) is shifted at \(\simeq 725\) cm\(^{-1}\) in the \(\beta_L(a)\) theoretical spectrum.

We do not observe the same drawbacks for \(\beta_L(g^+)\) for which the DFT-MD spectrum definitely provide the best agreement with the experiment spectrum.

In blue, we have highlighted the \(\omega(NH)_{Phe}\) wagging motion band of the N-H\(_{Phe}\) amide. Position of this band is identical for \(\beta_L(g^+)\) and \(\beta_L(a)\) conformers (respectively located at 463 and 465 cm\(^{-1}\), vs 471 cm\(^{-1}\) in the experiment), but the band shape for \(\beta_L(g^+)\) is remarkably in even better agreement with the experiment. As already pointed out in chapter 9 there is one band observed per NH group for small peptides (only one group for Ac-Phe-Ome), and the position obtained here for the three conformers is compatible with the range given in chapter 9 and ref.\(^{65}\) for a NH group involved in a C5 weak interaction. The lower frequency observed for \(\beta_L(g^-)\) (451 cm\(^{-1}\)) is also in line with what was highlighted in chapter 9 concerning lower frequencies for N-H groups being more free of intermolecular interactions. Here the \(\beta_L(g^-)\) conformer is the one with the longer C5 H-bond formed, i.e. 2.51±0.12 Å to be compared to 2.30-2.40±0.07 Å for the two other \(\beta_L\) conformers, thus providing this lower frequency of the NH\(_{Phe}\) wagging motion. The \(\omega(NH)_{Phe}\) wagging is part of a more complex massif, which shape- and band-positions are in remarkable agreement with the experiment for \(\beta_L(g^+)\), these other bands being due to backbone bending motions around the amide NH\(_{Phe}\) group.

As already discussed in ref\(^{105}\) and in chapter 6, the range 90-400 cm\(^{-1}\) appears to be more conformer selective than the 400-800 cm\(^{-1}\). In that respect, the theoretical spectrum of \(\beta_L(g^+)\) provides a great agreement with the experiment. All bands in this domain are due to bending motions delocalised along the backbone and along the side chain. Even more delocalised and larger amplitude dihedral motions start to kick-in at roughly 150 cm\(^{-1}\) and these motions are systematically dominant in all peaks located below 100 cm\(^{-1}\) in our theoretical spectra. Of special interest is the 100 cm\(^{-1}\) peak located in experiment and in the theoretical spectrum of \(\beta_L(g^+)\), absent in the spectrum of the two other conformers, that records rather highly delocalised bending motions along the backbone and the side chain of Ac-Phe-Ome (with some torsions also participating). This records the relative angular motion of the side-chain versus the backbone, corresponding to rather large-amplitude motions. Equivalent motions, but far less delocalised over the whole molecule, are seen for the two other conformers of Ac-Phe-Ome at higher frequencies. The 50-80 cm\(^{-1}\) bands obtained in the spectrum of \(\beta_L(g^+)\) are due to delocalized torsional motions along the whole backbone while the \(\sim 25\) cm\(^{-1}\) band also contains torsions of the side chain around \(\chi_1\) and \(\chi_2\) dihedrals. These bands can not be recorded yet at the FELIX beamline.

We believe we have made the demonstration that the \(\beta_L(g^+)\) structure can be unambiguously assigned to the gas phase Ac-Phe-Ome model peptide on the basis of the far infrared/terahertz spectral signatures.
11.2 Conformational assignment of the dimer: (Ac-Phe-OMe)$_2$ $\beta$-sheet

For the construction of the (Ac-Phe-OMe)$_2$ $\beta$-sheet structure, considering the results obtained for the Ac-Phe-OMe monomer, one would expect a conformation in the form of $\beta_L(g^+)$-$\beta_L(g^+)$ under the assumption that the dimer is solely assembled from $\beta_L(g^+)$ monomers. Such assumption does not take into account that, while they assemble to form a dimer, the monomers could however isomerise into different $\beta_L$ orientation as a result of intermolecular interactions. Alternatively, all possible combinations of dimers might be produced during the initial stage of the experiment and only the lowest free energy dimer will remain once passing through the supersonic expansion, following thermodynamics rules. Looking at the relative electronic and free energies of the six possible (Ac-Phe-OMe)$_2$ $\beta$-sheet conformations built (by hand) on $\beta_L$ orientations of the monomers (Table 11.2), one can see that the $\beta_L(g^+)$-$\beta_L(g^-)$ and $\beta_L(g^+)$-$\beta_L(g^-)$ conformers are quasi-isoenergetic, while the $\beta_L(g^+)$-$\beta_L(a)$ conformer is only $\sim$1.1-1.5 kcal/mol higher in energy. All six possible conformers are found within less than 3 kcal/mol of energy. No search for energy barriers has been conducted here.

<table>
<thead>
<tr>
<th>Structure</th>
<th>$\Delta E$ (kcal/mol)</th>
<th>$\Delta G$ 50K (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_L(g^+)$-$\beta_L(g^+)$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\beta_L(g^+)$-$\beta_L(a)$</td>
<td>1.54</td>
<td>1.49</td>
</tr>
<tr>
<td>$\beta_L(g^-)$-$\beta_L(a)$</td>
<td>2.80</td>
<td>2.73</td>
</tr>
<tr>
<td>$\beta_L(g^{-})$-$\beta_L(g^-)$</td>
<td>3.07</td>
<td>2.45</td>
</tr>
<tr>
<td>$\beta_L(a)$-$\beta_L(a)$</td>
<td>3.37</td>
<td>2.70</td>
</tr>
</tbody>
</table>

Table 11.2: Relative electronic energies ($\Delta E$) and free energies ($\Delta G$) for the different conformers of (Ac-Phe-OMe)$_2$ $\beta$-sheet investigated here (in kcal.mol$^{-1}$). $\Delta G$ is calculated at 50 K, and results given by two levels of electronic representations are provided.

In the pioneering work of Gerhards and coworkers$^{69-71}$, only the symmetrical structures $\beta_L(g^+)$-$\beta_L(g^+)$ and $\beta_L(a)$-$\beta_L(a)$ were investigated, because the $\beta_L(g^+)$ and $\beta_L(a)$ monomeric conformations were found lower in energy. We here investigate all the 6 possible $\beta_L$-$\beta_L$ structures.
Before discussing the far-IR spectroscopy and conformational assignment, let us discuss the vibrational spectra of the 6 conformers in the 1000-2000 cm\(^{-1}\) fingerprint region, employing scaled static harmonic spectra (B3LYP-D3/6-311+G(d,p), scaling factor of 0.967, 1% gaussian convolution applied for the plots), as presented in fig. 11.7. The mid-IR experimental spectrum has been recorded in the present work.

As already observed for the Ac-Phe-OMe monomer, this fingerprint region is hardly conformer selective and does not allow an easy conformational assignment. All theoretical spectra roughly provide the same IR spectrum for the \(\nu(C=O)\) Amide I and \(\delta(N-H)\) Amide II bands (same number of bands, same band-positions and very little difference in the band-shapes and intensities), and one can only start to observe differences in the theoretical signatures below 1300 cm\(^{-1}\), very similar to what was already obtained for the Ac-Phe-OMe monomer. These bands correspond to Amide III motions, backbone C-C and C-N stretching motions and N-H \(\text{Phe}\) and CCH bendings. These modes are mainly localised over one single strand. In that respect, \(\beta_L(g^+)-\beta_L(a)\), \(\beta_L(g^+)-\beta_L(g^-)\) and \(\beta_L(g^+)-\beta_L(g^-)\) appear as three better candidates for explaining the experimental features in the Amide III domain. The signatures in the mid-IR are however not sufficiently selective to unambiguously discriminate the conformation adopted by the \(\beta\)-sheet.
This is not true anymore once the far-IR signatures are taken into account, as presented in figure 11.8. We here have performed DFT-MD simulations on all the 6 possible $\beta_L-\beta_L$ structures and extracted the associated anharmonic dynamical vibrational spectra in the far-IR, to be compared to the experimental spectrum. As for Ac-Phe-OMe, each theoretical spectrum presented in figure 11.8 is generated as an average over three separate trajectories (each one with different initial conditions for the dynamics).

Figure 11.8: Experimental IR-UV ion dip spectrum of the gas phase (Ac-Phe-OMe)$_2$ $\beta$-sheet (in blue) compared with the calculated dynamical DFT-MD IR spectra (in red) of the 6 possible isomers of gas phase (Ac-Phe-OMe)$_2$ investigated here. The green boxes highlight missing spectral features in the theoretical spectra that are however present in the experimental spectrum. The orange boxes highlight extra spectral features that are present in the theoretical spectra but are absent in the experimental spectrum.

We especially show the 90-400 cm$^{-1}$ domain where the final conformational assignment is clear. Analysing the number of bands in this domain in this figure, their absolute and relative positions, their band-shapes and intensities with respect to the well resolved bands in the experimental spectrum, one can conclude that the DFT-MD theoretical spectrum of the $\beta_L(g^+)-\beta_L(g^-)$ isomer of the gas phase (Ac-Phe-OMe)$_2$ $\beta$-sheet structure provides an extremely precise account of the experimental features. The 300-400 cm$^{-1}$ triple-bands are in excellent agreement with the experiment (one however note that the middle-band intensity appears slightly too low), the 250-300 cm$^{-1}$ double-bands is not present with such accuracy in any other isomer (band-positions, shapes and intensities), the 150-250 cm$^{-1}$ triple-bands are just located on-spot (although presumably too low in absolute intensities), isomer $\beta_L(g^+)-\beta_L(g^-)$ being the
only other isomer displaying the same triple-bands. The 100 cm\(^{-1}\) band seems less conformer-selective as all 6 isomers spectra display this particular band, shifted by a few cm\(^{-1}\) from one isomer to the other.

As done in the previous section, we have highlighted in this figure bands that are missing in the theoretical spectra in order to match the experiment (green boxes) and bands that are appearing in the theoretical spectra but that are not present at these particular frequencies in the experiment (orange boxes). Strikingly, only the theoretical spectrum of \(\beta_L(g^+)-\beta_L(g^-)\) does not have any of these boxes highlighted. We unambiguously assign isomer \(\beta_L(g^+)-\beta_L(g^-)\) to the (Ac-Phe-Ome)\(_2\) β-sheet, which is also the lowest energy conformer.

We found that the 100-400 cm\(^{-1}\) range is the most conformational selective, thus because of the nature of the vibrational modes in this spectral range. We find there delocalised modes over the whole backbone and side chain residue mainly combination of bendings and dihedrals motions. In particular, we found the dihedral angle \(\chi_1\), see figure 11.3, involved in most of the modes below 200 cm\(^{-1}\). This internal coordinate directly drives the rotation of the phenyl ring (rotation that induces the conformational transitions between the \(g^+\), \(a\) and \(g^-\) orientations) and we expect the signatures of this internal coordinate to be extremely efficient for conformational differentiation between the different structures.

While NH or CO stretchings are excellent probes of the hydrogen bonds in which these functions can be involved, delocalised modes are very well suited to probe subtle differences in conformations on the side chain for example because these modes somehow probe all spatial locations. Far infrared delocalised modes share a common spirit with the "fingerprint range" for which modes involving CC, CO, CN stretchings are found on the whole system (but far less spatially delocalised). However, the far infrared domain appears to be more conformational selective than the "fingerprint" range since only this range provides a definitive assignment for the (Ac-Phe-Ome)\(_2\) dimer.

### 11.3 Comparisons of vibrational modes between the monomer and the dimer

We will conclude our investigations on the Ac-Phe-Ome monomer and its dimeric β-sheet form (Ac-Phe-Ome)\(_2\) by a discussion on the far infrared vibrational modes. The discussion somehow overlaps with chapter 9 (chapter especially focused on vibrational mode analyses) but we will here investigate the dimerisation process and its influence on the vibrational modes.

In figure 11.9 are plotted the DFT-MD infrared spectrum of the \(\beta_L(g^+)-\beta_L(g^-)\) dimer (in blue) and DFT-MD infrared spectrum of the \(\beta_L(g^+)^\) and \(\beta_L(g^-)\) monomer structures (in black). The red spectrum is the direct sum of the two black spectra. The blue boxes show the assignment of the \(\omega(NH)_{phe}\) wagging motions. For a correct comparison between vibrational modes, these theoretical spectra need to be in good agreement with the experimental one (and provide correct vibrational modes) which is definitely the case here as demonstrated in this chapter or in chapter 10.

We observe in figure 11.9 that, in the range 400-800 cm\(^{-1}\), (at the exception of the \(\omega(NH)\)
11.3. COMPARISONS OF VIBRATIONAL MODES

Figure 11.9: DFT-MD infrared spectrum of the $\beta_L(g^+)-\beta_L(g^-)$ dimer (in blue) and DFT-MD infrared spectrum of the $\beta_L(g^+)$ and $\beta_L(g^-)$ monomer structures (in black). The red spectrum is the direct sum of the two black spectra. The blue boxes show the assignment of the $\omega(N-H)_Phe$ wagging motion.

Figure 11.10: For each normal mode of (Ac-Phe-OMe)$_2$ ($\beta_L(g^+)-\beta_L(g^-)$ assigned structure), the mass weighted percentage of the motion over each strand according to a harmonic normal modes decomposition. When the two values are at 50%, the mode is perfectly delocalised over the two strands of the dimer. When the couple of value is 0 and 100%, the mode is perfectly localised over one strand only.
modes) the $\beta_L(g^+)$-$\beta_L(g^-)$ dimer spectrum is a good approximation of the sum of the $\beta_L(g^+)$ and $\beta_L(g^-)$ monomer spectra (red curve). A probable explanation is that we observe the same vibrational modes in the monomers and in the dimer (at the exception of the $\omega$(NH) modes since the two NH functions form hydrogen bonds in the dimer). If now we focus on the range below 150 cm$^{-1}$, we observe that the sum of the $\beta_L(g^+)$ and $\beta_L(g^-)$ monomer spectra (red curve) is now a very bad approximation of the $\beta_L(g^+)$-$\beta_L(g^-)$ dimer spectrum. One explanation could be that we have delocalised and coupled modes in the dimer that can not be seen as the direct sum of the vibrational modes in the individual monomers. The range 150-400 cm$^{-1}$ is intermediate in that respect.

Let us now focus on figure 11.10. The same figure is presented in chapter 9. For the dimer (Ac-Phe-OMe)$_2$ ($\beta_L(g^+)$-$\beta_L(g^-)$ assigned structure), see figures 11.3 and 11.4 for the structures, the mass weighted percentage of the motion over each strand according to the harmonic normal mode decomposition has been plotted for each normal mode. We therefore have two values, both at 50% for a mode perfectly delocalised over the two strands or close to 0 and 100% for modes perfectly localised over one strand only. Beyond 600 cm$^{-1}$, where we expect to find the wagging motions, we can see different kinds of modes, some are localised over one strand while some others are delocalised over the two strands. The $\omega$(CH) motions of the two aromatic rings, couple together and the associated modes for example around 690 cm$^{-1}$ are well delocalised over the two strands. But below 600 cm$^{-1}$, where delocalised modes are expected for this system, there is a really good correlation between the frequency of the mode and its spatial localisation, perfectly localised over one strand between 400 and 600 cm$^{-1}$ and almost perfectly delocalised over the two strands below 150 cm$^{-1}$. The domain 150-400 cm$^{-1}$ is intermediate.

Regrouping the information from both figures 11.9 and 11.10, and please note that we use the DFT-MD representation for figure 11.9 and the harmonic one for figure 11.10 (with good reasons, see chapter 10 on methods comparison), we see that when the vibrational modes are localised over one strand, the sum of the $\beta_L(g^+)$ and $\beta_L(g^-)$ monomer spectra (red curve) is a good approximation to the $\beta_L(g^+)$-$\beta_L(g^-)$ dimer spectrum, while when the modes are delocalised over the two strands, this is not the case anymore. In this second case, it means that we are observing vibrational modes that can mechanically not exist in the individual monomers.

11.4 Conclusion

This chapter has presented the conformational assignment of the $\beta$-sheet dimeric peptide (Ac-Phe-OMe)$_2$ (and its monomeric building block). This conformational assignment was not possible in the mid infrared range 1000-4000 cm$^{-1}$. It was impossible because as shown here, there are several low free energy conformations differing from each other only by the orientation of the phenyl group but having identical signatures in this infrared domain. The far infrared/THz spectra were shown here pivotal in order to assign the right conformer of each of the Ac-Phe-OMe and (Ac-Phe-OMe)$_2$ peptides.

The most efficient modes for conformational assignments were shown here to be the far in-
frared spatially delocalised modes because they involve internal coordinates that directly probe the side chain residue concerned by these conformational differences. In particular, we found the $\chi_1$ internal coordinate, see figure 11.3, to be involved in several conformational selective modes. This is not surprising because this $\chi_1$ internal coordinate directly drives the rotation of the phenyl ring and links our different low free energy conformations. We also briefly discussed the vibrational modes observed in the $\beta$-sheet dipeptide, especially in terms of involving motions over one strand only or other the two strands. We found the collective modes to be really delocalised over the two strands below $\simeq 200 \text{ cm}^{-1}$. 
Chapter 12

Conclusions and outlook

The work presented in this thesis is a contribution to the broad subject of the far infrared/Tera-Hertz (10-800 cm\(^{-1}\); 0.3-24 THz) gas phase vibrational spectroscopy of biological building blocks, with a synergy between original experiments and original theoretical calculations. We have hence characterized peptides, DNA base pairs, and phenol derivatives (without and with water). The investigated systems have been presented in chapter 2, and the introduction to that manuscript has given an overview of the reasons to characterise these particular systems by far infrared/Tera-Hertz spectroscopy in the gas phase.

IR-UV ion dip action vibrational spectroscopy has been performed in cold and isolated conditions on these neutral species. Principles of this spectroscopy as well as the laser desorption/heating source and molecular beams allowing these experimental conditions have been presented in chapter 3. This spectroscopy has been only recently made accessible in the far infrared/Tera-Hertz domain thanks to the development of far infrared Free Electron Lasers (FEL), the only tool intense enough to allow gas phase experiments in molecular beams. The work presented in this thesis has been done in the context of a collaboration between the theory group in Evry and the group of experimentalists of Dr. Anouk M. Rijs who has developed a setup at the FELIX laboratory at the Radbout University, Nijmegen, The Netherlands. Equivalent setups have been developed by the groups of Prof. Asmis and Dr. Filiecke at the Fritz Haber Institute in Berlin, more dedicated to characterise water\(^96,97\) or metal\(^98–104\) clusters. All the experimental spectra presented in this thesis have been measured by Daniël J. Bakker, Sander Jaeqx, Faady Sioury and myself, in a collaborative work. We systematically observed well resolved experimental far infrared spectra for all systems investigated, which is already an achievement for gas phase spectroscopy.

For conformational assignments as well as vibrational mode analyses, the experimental approach is systematically coupled with calculations. Our main theoretical method is DFT-MD simulations (Density Functional Theory - Molecular Dynamics), that has been used in Prof. Gaigeot’s group for more than a decade with large success in the mid infrared vibrational range. This methodology is based on a classical molecular dynamics for the nuclei and on a quantum description of the electrons using the DFT formalism (and in particular the BLYP-D3\(^123–125\) functional in this work), see chapter 4 for all details. This method takes into account modes couplings and anharmonicities of the potential energy surface and of the dipole surface
(contrary to the harmonic representation for which the description of the potential and of the dipole is restricted to be harmonic). We found the anharmonic DFT-MD/BLYP-D3 far IR/THz spectra to be in rather excellent agreement with the experimental spectra at the exception of the overtones and combinations bands unfortunately systematically missing in the dynamical spectra. I have performed most of the DFT-MD trajectories and their interpretations, some of them have also been performed in collaboration by Daniël J. Bakker in the group of A. M. Rijs (especially for the phenol derivatives).

This thesis whole work follows the seminal investigation: "Gas-Phase Peptide Structures Unraveled by Far-IR Spectroscopy: Combining IR-UV Ion-Dip Experiments with Born-Oppenheimer Molecular Dynamics Simulations" from the two collaborative groups of M.-P. Gaigeot and A. M. Rijs that showed that the signatures recorded in the far infrared spectral domain allowed distinguish between two structurally close low free energy conformations of the Ac-Phe-Gly-NH₂ dipeptide (which proved impossible in other vibrational domains).

All investigations in this thesis follow the same general conclusion: the vibrational signatures recorded in the far infrared/Tera-Hertz domain provide a clear means to unambiguously assign three dimensional conformations. That spectra are extremely well resolved helps a lot to that end. In this respect, we have in this work made the demonstration that indeed far infrared/Tera-Hertz spectroscopy allows to remove the spectral congestion observed in the other infrared domains (1000-4000 cm⁻¹). This spectral congestion was discussed in the introduction to this manuscript and was one of the motivation to probe vibrational signatures in another spectral domain than the usual 1000-4000 cm⁻¹ mid-IR. One of our initial goal is very nicely achieved.

Part of the characterisations done in this work were related to molecular systems which three dimensional conformations had already been assigned through signatures in the mid infrared. For those, the far infrared/Tera-Hertz signatures confirmed the assignments, and these investigations demonstrated that our combined strategies of IR-UV ion dip spectroscopy experiments and DFT-MD dynamical spectra were working and were relevant. On the other hand, a large part of our characterisations were done on systems for which there were still no final consensus in the literature based on the 1000-4000 cm⁻¹ mid-IR spectral range. One example is the β-turn conformation of Ac-Phe-Cys-NH₂ dipeptide (characterised within a whole series of dipeptides) for which two different investigations in the mid infrared provided two opposite conclusions regarding the conformational assignment. We demonstrated in chapter 6 that the correct conformation is the one assigned by Alauddin et al. based on the far IR/Tera-Hertz signatures. We believe that our best demonstration is presented in chapter 11 for the β-sheet model (Ac-Phe-OMe)₂. Only the far infrared range provided the definitive structural assignment to the β₁(\(g^+\))-β₁(\(g^-\)) structure. Chapter 8 also presented the conformational assignment of saligenin-water clusters using both the 3000-4000 and 0-800 cm⁻¹ spectral ranges, and we believe that our work provides a demonstration of the strength of the far infrared signatures for the final assignments.

One of our initial questions was to understand how far we can go in terms of conformational assignment using the far infrared/THz vibrational signatures. We actually did not
answer to this question during the course of this thesis work, for the simple reason that we have not reached the limits of the far infrared yet, i.e. we managed to systematically provide a conformational assignment for all systems investigated (for which the structure was unknown) combining experiments and simulations.

Instrumental to the successful conformational assignments done here is the synergy with theoretical calculations in terms of dynamical DFT-MD anharmonic spectra. Looking at the far infrared/Tera-Hertz spectra recorded experimentally, in all chapters of this thesis, one can see that the spectra are far from simple to analyse without the help of calculations. The spectra in this domain are not congested but they are composed of a very large number of (well resolved) bands, not so easy to interpret without the help of spectra simulations in this low frequency spectral domain.

Once we have demonstrated the relevance of the far infrared/Tera-Hertz domain for conformational assignments, two issues have been addressed in my work of interest/relevance to the community of computational chemists for spectroscopy (however not restricted to this community alone as we will see below).

The first issue is the unveiling of the vibrational modes in the far IR/THz domain. Isotopic substitution is an experimental way to deal with peaks assignments, but as presented in chapter 9, this tool is far less efficient in the far infrared because the modes are much more collective/delocalised than typically observed in the 3000-4000 cm$^{-1}$ range, thus complicating the isotopic interpretations. We have therefore used our DFT-MD dynamical spectra and their interpretation in terms of molecular motions in order to unravel vibrational modes in the low frequency domain. The methodology has been described in details in chapters 4 and 9 where some limitations of the methodology have also been discussed. We have analysed in details the vibrational modes of all systems investigated in this thesis, and as such the vibrational mapping achieved here is representative of these systems. Some modes are presumably missing, that could be unravelled only by investigating even more diverse molecular systems. Chapter 6 presented a mapping of the vibrational modes for a whole Ac-Phe-'AA'-NH$_2$ dipeptide series (where AA is replaced by one amino acid among a series of 6). A more complete mapping is presented in chapter 9, the mapping being now extended to most of the molecular systems/building blocks that I characterised during my PhD. We believe that such mapping might be useful for experimentalists, providing a tool to assign vibrational modes and to obtain insights on three dimensional structures without using theoretical methods.

An overview of the mapping is presented in figure 12.1. In a nutshell, we found two kinds of modes in the far infrared/THz domain, i.e. local modes (i.e. composed of only one major contribution involved in the motion) and delocalised/collective modes (i.e. composed of a large number of contributions). Local modes are found to be out of plane motions of hydrogen atoms, i.e. $\omega$(NH), $\omega$(OH), $\omega$(CH) in the systems investigated here, located inbetween 200 and 900 cm$^{-1}$. These local modes are found to be the most intense active modes in the far infrared spectra, and are probably not coupled with other internal coordinates because of the lightness of the hydrogen atom involved. Delocalised modes are less intense in the IR spectra, they are
Figure 12.1: Mapping of the far infrared/Tera-Hertz gas phase vibrational modes achieved in this work. Analyses are based on peptides, base pairs and phenol derivatives molecules.

located all over the far infrared but typically for peptides, the bendings of the backbone mainly arise below 400 cm\(^{-1}\) while the torsional motions of the backbone arise below 100 cm\(^{-1}\). We systematically found participations of the hydrogen bond stretchings below 250 cm\(^{-1}\), but as described in chapter 9 these motions are not easy to unravel as many other motions participate to these H-Bond motions, and there is no unique way to extract H-Bond vibrational signatures (at least with our methodology).

This mapping has been achieved more specifically for peptides, base pairs and phenol derivatives gas phase molecules. We believe our results are transferable to other systems, however incomplete they might be. For an even more complete vibrational modes mapping, this work should be completed with supplementary molecular systems like water clusters\(^{96,97}\) or metal clusters\(^{98-104}\) for instance, measured by Asmis and Fielicke’s groups in the far infrared.

Beyond the actual modes mapping, such vibrational analyses are also important to understand why the far infrared/Tera-Hertz domain is more conformational selective than the other spectral domains (at least for the systems investigated here). While NH or CO stretchings are excellent probes of the hydrogen bonds in which these functions can be involved, delocalised modes are very well suited to probe subtle differences in conformations, for instance for backbones or side chains in peptides, because these modes probe a large amplitude of
spatial locations. For the \((\text{Ac-Phe-OMe})_2\) system, see chapter 11, we found for example that the dihedral angle that drives the rotation of the phenyl ring (motion linking the two lowest free energy conformers) is involved in most of the modes below 200 cm\(^{-1}\), and we believe this internal coordinate to be highly (maybe the most) conformational selective for this system.

Another issue addressed in my work is to understand which theoretical methods are able to reproduce the experimental far infrared signatures and why. As mentioned above, the dynamical DFT-MD/BLYP-D3 spectra are systematically in good agreement with the experimental ones with in particular an average deviation of \(\sim 8\) cm\(^{-1}\) between the experimental and DFT-MD/BLYP-D3 features (considering all systems characterised in this work), as well as band-shapes in extremely good agreement with the experiments as well as band intensities ratio. There are however systematic missing bands in the DFT-MD/BLYP-D3 spectra in the case of DNA base pairs and phenol derivatives. We think that these missing features (in the DFT-MD spectra) arise from overtones and combination bands. At the time of writing, we do not understand why they are missing and how to obtain these extra features in our DFT-MD anharmonic spectra (while they should be present by construction of the method). Note that the VPT2 anharmonic method (see below) that provides the frequencies and intensities of overtones and combination bands by construction of the method, did not help us to assign these features either.

In chapter 7, a comparison between experimental, anharmonic MD, harmonic and anharmonic VPT2 spectra has been presented for the specific case of phenol derivatives. Chapter 10 extends this discussion to all systems investigated in my thesis work. We refer the readers to these chapters for the demonstration of the quality and robustness of the DFT-MD/BLYP-D3 spectra.

I believe the most important point of these chapters is probably the comparison between DFT-MD and harmonic spectra (calculated with the same DFT functional and equivalent electronic representation for the basis sets). The DFT-MD (anharmonic) spectra are much more expensive to calculate than the harmonic spectra and in this respect it is important to understand what are the vibrational modes that are indeed anharmonic (and "deserve"/require an anharmonic treatment) and which modes are intrinsically harmonic and for which one can use the cheaper harmonic approximation for vibrational spectroscopy (DFT-MD is roughly 200 times more expensive than the harmonic approximation, time given here for a 20 ps trajectory). Comparisons between DFT-MD/BLYP-D3 and BLYP-D3 harmonic spectra have been presented in chapter 10 and the OH wagging motions (O-H groups here present in the phenol derivatives molecules) have been identified as the most anharmonic modes in the far infrared domain (with different degrees of anharmonicities depending on the OH function and its interactions with its environment). The \(\omega(\text{NH})\) wagging motions of the NH\(_2\) functions in peptides and base pairs are the second source of large anharmonicities, while the backbone Amide V \(\omega(\text{NH})\) waggings are far less anharmonic. Anharmonicities have been shown here for band-positions and band-shapes, and also for the composition of the modes. All other \(\omega(\text{CH})\) wagging modes and backbone/ring angular/dihedral angular motions have been found to behave harmonically. We have also nicely shown that the anharmonicities in the large amplitude wagging motions of
the O-H/N-H groups are correlated with the strength in the H-Bonds formed by these groups. Weak (long) H-Bonds give rise to floppy H-Bond motions, which in turn are highly anharmonic, while strong (short) H-Bonds lead to far less flexible H-Bond motions, thus very harmonic. It is interesting to see that the backbone/ring angular/dihedral motions are essentially harmonic, which is presumably not what the whole community would initially have had in mind.

Using VPT2 anharmonic spectroscopy (as implemented in the Gaussian package by Barone\textsuperscript{134–136}), we found that VPT2 spectra provided theoretical anharmonic spectra of less good quality than the DFT-MD spectra when compared to experiments for a prize yet equivalent (DFT-MD is roughly twice more expensive as VPT2 calculation for a 20 ps trajectory). The match with experiment is less good than the one provided by MD for all the systems investigated here. The only advantage of the VPT2 method over the DFT-MD is that the overtones and combination bands are included by construction in the formalism, while it should be present in the dynamical spectra but failed to appear in all our calculations here (maybe because of the too low temperature of the dynamics). However, we found many times that these transitions are not intense enough in the VPT2 spectra (in comparison with the fundamental transitions) so that the assignments of these bands to experimental features still remain difficult and ambiguous.

We have also tested if a hybrid functional like B3LYP could provide any improvement with respect to the BLYP-D3 one. We have tested here DFT-MD/B3LYP-D3 infrared spectra for four phenol derivatives. The hybrid electronic representation was not found to provide better or worse spectra than the GGA one. This is somehow disappointing because the DFT-MD/B3LYP-D3 spectra are roughly 5-6 times more expensive than the DFT-MD/BLYP-D3 spectra, but it is reassuring as the GGA functional already works extremely well. Before any definitive conclusion about the quality of the B3LYP-D3 functional can be definitely given, a more complete analysis on more gas phase molecular systems should be however performed.

We believe that the present combined experimental and theoretical work using the far infrared/Tera-Hertz spectral domain is opening new avenues for the characterisation of complex gas phase molecular systems. The first obvious perspective to my work would be to investigate larger systems than the ones characterised here. One issue is to see how far we can go in terms of conformational assignment using the far infrared/THz vibrational signatures of these larger systems and test whether the signatures are still so well resolved and still provide so clear separate peaks. We have already touched a bit this point with the Z-Ala\textsubscript{6}-NH\textsubscript{2} peptide, where we could see that the far IR/THz signatures are still well-resolved indeed. For even larger peptides one would like in particular to manage to provide unambiguous far IR/THz vibrational signatures of a helix \textit{versus} a strand structure for instance, or even manage to provide a direct unraveling of random coils. Same outcomes for DNA/RNA based structures. Characterising larger and larger peptides building blocks (similarly for DNA/RNA larger and larger building blocks) would also allow record far IR/THz spectra for more diverse H-Bonded motifs that can be encountered in the more organized proteins (resp. DNA/RNA). One typically thinks of the C5 motif found in $\beta$ strands, the C7 found in $\gamma$-turns, C10 in $\beta$-turns and $3_{10}$ helices, C13 in $\alpha$-helices and $\beta$-sheets. We have in this work already provided a few signatures of some
of these motifs, but the building blocks were somehow too small to provide unambiguous fi-
nal signatures. Larger peptides are now accessible to the gas phase spectroscopy experiments
made by our partners, and they are also accessible to DFT-MD/BLYP-D3 simulations without
increasing too much the computational cost. We are especially interested in characterising
β-sheet structures (composed of larger and larger number of strands), as these peptides are
involved in terrible diseases (like Alzheimer to cite only one disease).

From the theoretical point of view, there are a number of further investigations to perform.
For a lack of time, some of these obvious investigations could not be achieved yet. Let me
cite for instance further investigations on DFT-MD spectra employing other GGA functionals
and other hybrid functionals. I would have liked to test PBE, PBE0 and ωB97D functionals
in order to see if any further improvements could be obtained over the BLYP-D3 dynamical
spectra calculated here. More tests on dispersion interactions in the modern functionals would
be useful as well. I would also have liked to spend more time on the VPT2 equations for
the intensities of the bands, especially the ones related to the overtones/composition bands,
to understand if there is anything that could be done to improve these intensities and get non
ambiguous assignments. I would also have liked to calculate anharmonic spectra with the more
advanced VSCF method developed by Gerber and Bowman’s groups\textsuperscript{189,190} and understand how
these methods stand with respect to the dynamical spectra calculated in my work.

Another obvious investigation concerns the way we extract the vibrational modes from the
dynamics. Such vibrational modes assignments have been done through the ICDOS method
described in chapter 4, and as shown there, there are a few issues that could be investigated
and presumably would lead to new developments. Although these issues did not prevent
us from interpreting the vibrational peaks in my works, more refined techniques based on
effective normal modes extracted from the dynamics might be of better use. The group in
Evry has provided such a method in the past, see ref.\textsuperscript{202,203}, there are a few other groups who
also gave some developments in this area. These methods might be more quantitative in their
interpretation of the vibrational modes, but they suffer some defaults that are intrinsic of the
gas phase underlying dynamics (equipartition of energy), which might limit the relevance of
such developments. It is therefore not clear whether such developments (however exciting
they might be) would be so useful in the end.

My impression is that a great perspective to this work would be to develop clever ways to
mix harmonic and non-harmonic methodologies within the molecular dynamics simulations
of vibrational spectra. Let me explain. We have indeed seen that a lot of the vibrational
modes in the far IR/THz domain were found harmonic in nature (typically the delocalised
backbone/ring angular/torsional deformations), while only the O-H/N-H (and actually only
the N-H of NH\textsubscript{2} groups) wagging motions were found with substantial anharmonic charac-
ters. Large amplitude vibrational motions of these groups were found even more anharmonic
when weak and floppy H-Bonds were formed. A harmonic treatment of the vibrations would
therefore be enough for the delocalised backbone/ring deformations while the dynamical an-
harmonic treatment is essential for the wagging motions. Treating these harmonic/anharmonic
motions through different theoretical approaches within the same molecular dynamics would
be relevant in order to possibly reduce computational costs of the theoretical spectra. At the
extreme, one could say that the dynamics could be applied only on the anharmonic O-H/N-H
groups, while another method could be applied to the harmonic motions. Alternatively, the dynamical representation could maybe be used in order to extract/refine the vibrational potential/dipole surface of the anharmonic motions/modes, to be injected afterwards in another theoretical method less computationally costly. One could hence think in terms of injecting these potentials into the VPT2 anharmonic method for instance. The advantage of this later being that the overtones/combination bands are included in the method by construction. These overtones/combination bands are anharmonic couplings that should also naturally appear in the dynamics, but the trajectories performed here did not succeed in getting them. One should work on that particular issue and understand the current reasons for these modes not to be appearing in the final spectra.
Bibliography


[109] Frisch, M. J. *et al.* “Gaussian 09 and Revision C.02 and Gaussian and Inc. and Wallingford CT and 2009”, .


[183] Sulpizi, M.; Gaigeot, M.; Sprik, M. 2012, 8, 1037.


Remerciements/Acknowledgements

J’aimerais commencer par remercier Jean-Yves Salpin et Marie-Pierre Gaigeot pour m’avoir accueilli au laboratoire LAMBE en 2014 pour un stage de M2 dans un premier temps poursuivi par cette thèse qui a abouti à ce manuscrit.

Je remercie également tous le personnel du LAMBE pour leur accueil et leur gentillesse. En particulier Vincent pour m’avoir formé et avec qui j’ai eu la chance de travailler en arrivant au laboratoire.

Mais également Yannick, Sacha et Riccardo avec qui j’ai eu grand plaisir à partager ces trois années, travailler, discuter, manger.

Merci à Ketty qui a rendu mon travail si facile au cours de ces trois années!

Je n’oublie pas mes amis de la cantine: Alan, Jean et Maud qui me manqueront plus que la cantine.

Je souhaite le meilleur aux "nouveaux" arrivants: Louis, Sana, Daria, Simone, Veronica, Fabrizio, Flavio et Delphine (et Sjors!) et un bon courage pour la suite.

J’ai eu la chance de pouvoir réaliser des expériences en plus des activités de théoriciens menées à l’université d’Évry. J’aimerais donc remercier Anouk avec qui nous avons travaillé tout le long de ces trois années et qui m’a régulièrement accueilli au laboratoire FELIX aux Pays-Bas.

Ce qui m’a permis de travailler avec Daniël J. Bakker qui m’a aidé à réaliser les expériences présentées dans cette thèse. Grâce à lui, chaque visite aux Pays-Bas a été un grand plaisir.

Merci également à Britta Redlich, Miriam Heijmerink et les membres de FELIX qui ont grandement facilité mes séjours à Nijmegen.

J’ai eu la chance de collaborer avec de nombreuses personnes que j’aimerais également remercier: David Lauvergnat, Jim Lisy, Haochen Ke.

Un grand merci à Sandra Luber et Knut Asmis d’avoir accepté d’être rapporteurs pour la thèse et pour leur retours.

Merci à mes parents, mon frère et mes grand-parents et au reste de ma famille ainsi qu’à mes amis: Dorian, Alex, Maxime, Olivier. Leur présences m’a beaucoup aidé.

J’aimerais terminer par remercier Nathalie pour son soutien et sa présence. Je n’aurais pas pu réussir à terminer ce travail sans elle.
Titre: Spectroscopie infrarouge en phase gazeuse dans le domaine de l'infrarouge lointain/Téra-hertz : expériences et théorie.

Mots clés: Spectroscopie vibrationnelle infrarouge, infrarouge lointain, peptides, FELIX.

Résumé: Cette thèse présente des travaux combinés théorie et expériences de spectroscopie infrarouge en phase gazeuse dans le domaine de l’IR lointain/THz, menés au laboratoire LAMBE (UMR8587) de l’université d’Evry val d’Essonne & Paris-Saclay en collaboration avec les groupes du Dr. Anouk Rijs (Nijmegen, Pays-Bas) et du Prof. Mattaniah DeVries (Santa Barbara, USA). La spectroscopie infrarouge est utilisée ici pour l’attribution conformationnelle de molécules et clusters en phase gazeuse. Le principe est de comparer le spectre mesuré expérimentalement avec celui calculé pour les structures 3D identifiées de plus basses énergies, le meilleur accord donnant la structure présente dans les conditions expérimentales. Pour cette thèse les expériences sont de type IR-UV ion-dip et les spectres théoriques sont simulés par dynamiques moléculaires ab initio de type DFT-MD, qui prennent en compte la température et les anharmonicités vibrationnelles. Les systèmes considérés sont des dipéptides, un modèle de feuillet β, des dérivés du phénol (et complexés à l’eau) des paires de bases de l’ADN, dont les structures sont bâties sur des liaisons hydrogènes intra- et inter-moléculaires.

Nous démontrons que le domaine de l’IR lointain/THz (<800 cm⁻¹, <24 THz) permet d’identifier sans ambiguïté la structure 3D de ces molécules et clusters en phase gazeuse, là où les signatures du domaine 1000-4000 cm⁻¹ sont insuffisantes. Nous répondons également aux trois questions suivantes:

- Quels sont les modes de vibration présents dans le domaine de l’infrarouge lointain/THz? Une carte des modes de vibration a été mise en place sur les systèmes présentés ci-dessus. Deux types principaux de modes de vibrations ont été identifiés dans le domaine IR lointain/THz: des modes locaux de wagging des atomes d’hydrogène qui se couplent peu avec les autres coordonnées internes, et des modes délocalisés pour lesquels beaucoup de coordonnées internes sont couplées pour former des modes de vibration délocalisés le long des squelettes des molécules.

- Quels sont les méthodes théoriques les plus adaptées pour reproduire un spectre IR lointain/THz? L’outil théorique principal utilisé ici est la méthode DFT-MD prenant en compte les anharmonicités de la surface d’énergie potentielle et de la surface du moment dipolaire, ainsi que le couplage entre modes, donnant d’excellents accords théorie/expérience. Les spectres DFT-MD ont été comparés aux spectres calculés par approximation harmonique (bien moins coûteuse en temps de calcul), qui bien souvent peuvent se montrer suffisamment de bonne qualité pour une attribution de structure moléculaire, et aux spectres VPT2 anharmoniques qui donnent des résultats systématiquement moins bons que ceux DFT-MD pour un coût de calcul équivalent.

- Jusqu’où les signatures IR lointain/THz peuvent-elles permettre une attribution non ambiguë des structures moléculaires 3D? La meilleure illustration concerne le modèle de feuillet β (Ac-Phe-OMe)₂. Sa conformation exacte ne peut être extraite des signatures spectrales des domaines habituels 1000-4000 cm⁻¹ mais peut être identifiée sans ambiguïté avec les signatures IR lointain/THz, comme démontré ici.
Title: Far infrared/Tera-Hertz spectroscopy in the gas phase: experiments and theory

Keywords: Spectroscopy, far infrared, peptides, FELIX

Abstract: We present here combined experimental and theoretical investigations of gas phase far infrared/THz spectroscopy, performed at the LAMBE Laboratory (UMR8587) - University of Evry val d’Essonne & Paris-Saclay in collaboration with the groups of Dr. Anouk M. Rijs (Radbout University & FELIX Laboratory, Nijmegen, The Netherlands) and Prof. Mattaniah DeVries (University of Santa Barbara, USA). Far infrared/THz spectroscopy is applied for conformational assignments (3D structures) of molecules and clusters in the gas phase. In practice, experimentally measured spectra are compared with theoretical spectra for low energy structures, the best match providing the information about the structure(s) present in the experimental conditions. For this thesis, experimental spectra are measured using IR-UV ion dip spectroscopy and theoretical anharmonic spectra are calculated using the DFT-MD method (Density Functional Theory - Molecular Dynamics) taking into account anharmonicities of the potential energy surface, of the dipole surface, and mode couplings. Investigated systems are dipeptides, a β-sheet model, phenol derivatives (also complexed with water molecules) and DNA base pairs, all structures being based on intra-/inter-molecular hydrogen bonds.

We demonstrate for several systems that the far infrared/THz spectral domain (<800 cm⁻¹, <24 THz) provides conformer selectivity for gas phase molecules and clusters, without structural ambiguities contrary to the more classical 1000-4000 cm⁻¹ spectral range. In this thesis we also answer the three following questions:

- What are the vibrational modes that can be found in the far infrared/THz domain? A map of vibrational modes has been built for the systems introduced above. Two kinds of vibrational modes are hence found: local wagging modes in which one single internal coordinate dominates the motions, and delocalised/more collective modes for which several internal coordinates couple and lead to spatially delocalised modes over molecular backbones.

- What are the most adapted theoretical methods for calculating far-infrared/THz spectra? Our main theoretical tool is the DFT-MD anharmonic method, taking into account anharmonicities of the potential energy surface, of the dipole surface, and mode couplings. This method provides excellent agreements with respect to the measured experimental spectra. DFT-MD spectra are compared with harmonic spectra (computationally far less costly), that are often found good enough to obtain reliable conformational assignments, and to VPT2 anharmonic spectra that systematically provide theoretical spectra in less good agreement than MD anharmonic spectra, for an equivalent computational cost.

- How far can we go in terms of conformational assignment using the far infrared/THz domain? The conformational assignment of the (Ac-Phe-OMe)₂ β-sheet model certainly illustrates the best the strength of the far IR/THz spectral domain for conformational assignment. This cannot be achieved using the ‘classical’ 1000-4000 cm⁻¹ spectral signatures only, but can be achieved without ambiguities once the far infrared/THz signatures are taken into account.
**Titre :** Spectroscopie infrarouge en phase gazeuse dans le domaine de l’infrarouge lointain/Téra-hertz : expériences et théorie.

**Mots clés :** Spectroscopie vibrationnelle infrarouge, infrarouge lointain, peptides, FELIX.

**Résumé :** La spectroscopie infrarouge permet d’identifier la structure 3D de systèmes moléculaires, par comparaison des spectres mesurés et simulés. Nous travaillons en phase gazeuse, où les molécules et clusters sont libres d’interactions intermoléculaires. Notre travail combine les expériences IR-UV ion-dip et le calcul de spectres IR anharmoniques par la méthode DFT-MD. Le spectre IR est calculé pour les structures 3D de plus basses énergies, le meilleur accord donnant la connaissance de la structure présente dans les conditions expérimentales.

Nous démontrons que le domaine de l’IR lointain/THz (<800 cm⁻¹, <24 THz) permet d’identifier sans ambiguïté la structure 3D de molécules et clusters en phase gazeuse, là où les signatures du domaine 1000-4000 cm⁻¹ peuvent être limitées. Les systèmes considérés sont des dipeptides, un modèle de feuillette β, dérivés du phénol (et complexés à l’eau) des paires de bases de l’ADN, dont les structures sont bâties sur des liaisons hydrogène intra/inter-moléculaires.

---

**Title :** Far infrared/Tera-Hertz spectroscopy in the gas phase : experiments and theory

**Keywords :** Spectroscopy, far infrared, peptides, FELIX

**Abstract :** Infrared spectroscopy allows the assignment of three dimensional structures of molecular systems, by comparing experimental and theoretical spectra. Our investigations take place in the gas phase, where molecules and clusters are free of intermolecular interactions. Our work combines experimental IR-UV ion dip spectroscopy and theoretical DFT-MD anharmonic spectroscopy. The infrared spectrum is calculated for low energy 3D structures and the best match between theory and experiment provides the information about the structure present in the experimental conditions.

We demonstrate for several systems that far-infrared/THz spectroscopy (<800 cm⁻¹, <24 THz) allows conformational assignment without ambiguities, contrary to the more traditional 1000-4000 cm⁻¹ range. Systems investigated here are dipeptides, a β-sheet model, phenol derivatives (also complexed with water molecules), DNA base pairs, all these structures being built on intra-/inter-molecular hydrogen bonds.