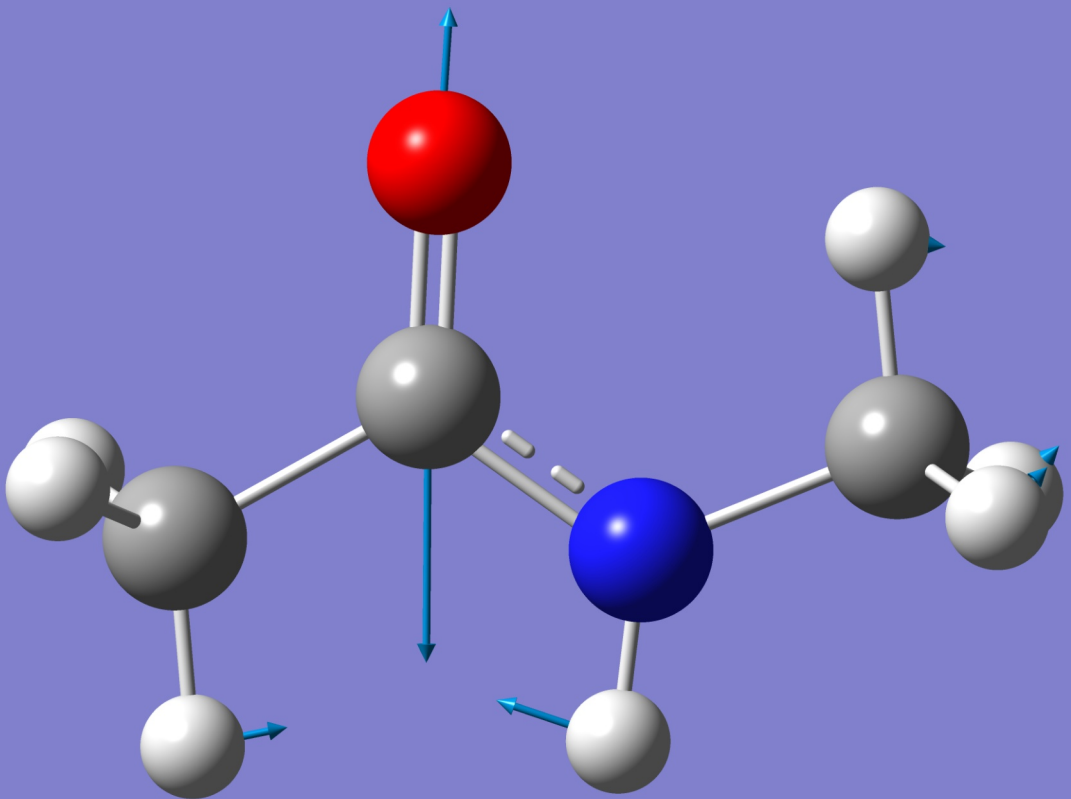


# Computational infrared spectroscopy

## Calculation of the amide I absorption of proteins

# Cesare Michele Baronio





# Computational infrared spectroscopy

## Calculation of the amide I absorption of proteins

Cesare Michele Baronio

Academic dissertation for the Degree of Doctor of Philosophy in Biophysics at Stockholm University to be publicly defended on Wednesday 29 April 2020 at 10.00 in Magnélisalen, Kemiska övningslaboratoriet, Svante Arrhenius väg 16B.

### Abstract

Infrared spectroscopy is an important technique that allows to retrieve structural information from the analysis of absorption spectra. The main application of infrared spectroscopy within life science is the study of the amide I band, which is correlated with protein backbone conformation and, consequently, with the secondary structure of proteins. However, band assignment and interpretation of the infrared spectra is not straightforward.

Therefore, several simulation methods were developed to guide the interpretation of experimental amide I spectra. In this thesis, one of these methods is a normal mode analysis, which is based on the evaluation of the intrinsic vibration of the amide groups and the interactions between them. The calculation considers several effects: transition dipole coupling, nearest neighbor interaction, the local environment effect and the effect of hydrogen bond. From the normal mode analysis, it is possible to obtain the simulated infrared spectrum and the contribution of each amide group to a specific spectral range of the spectrum.

The aim of this thesis and of the included publications is to explain this approach, to improve it and to show its potential. Results from simulations were compared with experimental data for different proteins of interest: amyloid- $\beta$  oligomers and  $\beta$ -helix proteins. Simulated and experimental infrared spectra showed similar bands. Simulations also provided additional conclusions: they confirmed the random mixing of amyloid- $\beta$  peptides in oligomers; they suggested that amyloid- $\beta$  peptides contribute at least two strands in the structure of the oligomers; they revealed that the high wavenumber band, typical of antiparallel  $\beta$ -sheets, can be caused by other secondary structures, but not by parallel  $\beta$ -sheets. In addition, to verify and to improve the accuracy of this approach, simulation results were also put in a direct comparison with results from density functional theory calculations. From this comparison, a new optimal set of parameters for the calculations is suggested.

**Keywords:** *infrared spectroscopy, FTIR, simulation, calculation, amide I, transition dipole coupling, F matrix, protein, amyloid  $\beta$ , oligomers,  $\beta$ -helix.*

Stockholm 2020

<http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-179912>

ISBN 978-91-7911-074-1  
ISBN 978-91-7911-075-8



Stockholm  
University

Department of Biochemistry and Biophysics

Stockholm University, 106 91 Stockholm



COMPUTATIONAL INFRARED SPECTROSCOPY

Cesare Michele Baronio



# Computational infrared spectroscopy

Calculation of the amide I absorption of proteins

Cesare Michele Baronio

©Cesare Michele Baronio, Stockholm University 2020

ISBN print 978-91-7911-074-1

ISBN PDF 978-91-7911-075-8

Cover image: representation of a NMA molecule. The atomic displacements during the amide I vibration are shown by arrows. Image created with GaussView.

Printed in Sweden by Universitetsservice US-AB, Stockholm 2020

To the people who  
support me

“Computers don’t solve  
problems. People do.”  
F. Jensen, Introduction to  
computational  
chemistry



# Contents

List of Publications.....	1
List of Abbreviations.....	3
1. Introduction.....	5
2. Proteins.....	7
2.1 Composition.....	7
2.2 Structure.....	7
2.3 Peptide aggregation and amyloid $\beta$ .....	11
2.4 $\beta$ -helix proteins.....	13
3. Infrared spectroscopy.....	15
3.1 Classical description.....	16
3.2 Quantum mechanical description.....	17
3.3 Instrumentation.....	19
3.4 Interpretation of infrared spectra.....	20
4. Methods for the simulation of the amide I band.....	23
4.1 Theoretical description of the $F$ matrix.....	23
4.2 Transition dipole coupling.....	27
4.3 Nearest neighbor coupling.....	29
4.4 Local environment.....	29
4.5 Hydrogen bond.....	31
4.6 From $F$ matrix to simulated spectrum.....	32
5. Density functional theory and amide I band.....	35
5.1 Introduction to quantum chemistry calculations.....	35
5.2 Density functional theory.....	37
5.3 Hessian matrix reconstruction.....	39
6. Results and discussion.....	41
6.1 Paper I.....	41
6.2 Paper II.....	42

6.3 Paper III.....	43
6.4 Paper IV.....	45
7. Conclusions and outlook.....	47
Acknowledgements.....	49
Sammanfattning (Summary in Swedish).....	51
Bibliography.....	53

# List of Publications

- Paper I: Maurizio Baldassarre, Cesare M. Baronio, Ludmilla A. Morozova-Roche and Andreas Barth, Amyloid  $\beta$ -peptides 1–40 and 1–42 form oligomers with mixed  $\beta$ -sheets, ***Chem. Sci.***, 2017, **8**, 8247-8254.  
<https://doi.org/10.1039/C7SC01743J>
- Paper II: Cesare M. Baronio, Maurizio Baldassarre and Andreas Barth, Insight into the internal structure of amyloid- $\beta$  oligomers by isotope-edited Fourier transform infrared spectroscopy, ***Phys. Chem. Chem. Phys.***, 2019, **21**, 8587-8597.  
<https://doi.org/10.1039/C9CP00717B>
- Paper III: Cesare M. Baronio, Andreas Barth, The amide I spectrum of proteins - Optimization of transition dipole coupling parameters using density functional theory calculations, ***J Phys Chem B***, 2020, **124**, 9, 1703-1714.  
<https://doi.org/10.1021/acs.jpcc.9b11793>
- Paper IV: Cesare M. Baronio, Huimin Li, Guangyuan Song, Henry Anneck, Robert Gustafsson, Markel Martinez, Pål Stenmark, Andreas Barth, The amide I spectrum of parallel  $\beta$ -sheet proteins, manuscript in preparation.



# List of Abbreviations

A $\beta$	Amyloid $\beta$
DFT	Density Functional Theory
IR	Infrared
NMR	Nuclear Magnetic Resonance
FTIR	Fourier Transform Infrared
PDB	Protein Data Bank
SAS	Solvent Accessible Surface
TDC	Transition Dipole Coupling
TDM	Transition Dipole Moment



# 1. Introduction

Simulations and modeling are becoming more and more important in scientific research. The ability to reproduce real systems through mathematical and physical models is a powerful tool to analyze many important physical and chemical processes.

A powerful feature of using simulations is that simulations usually rely on specific parameters to describe the system of interest. The parameters can easily be changed in the simulations, which make possible to observe the consequences that this change implies on the system.

In some cases, the use of simulation is even preferred over empirical experiments because the latter are too risky or require too much time. The choice of the simulation approach can also be a way to reduce the costs of scientific projects.

Also in infrared (IR) spectroscopy, simulations are helpful. IR spectroscopy is a technique used to retrieve structural information, in particular the secondary structure of proteins, through the study of the absorption in the amide I band ( $1700\text{-}1600\text{ cm}^{-1}$ ). The vibrations in the amide I spectral range are mainly due to the stretching vibrations of the C=O bonds in the amide groups of the protein backbone, then the IR spectrum of the amide I band is sensitive to the backbone conformation of proteins. However, the interpretation of the experiments is not always easy because of several factors, such as the possibility of overlapping bands and an incomplete theoretical description of the amide I band.

Several computational approaches were suggested in the literature to help in understanding the amide I band and of the experimental results. One of these methods is based on the floating oscillator model<sup>1</sup> and on a normal mode analysis of the vibrations in the amide I band. The wavenumbers and the intensities of the amide I oscillators are retrieved through determination of the eigenvalues and the eigenvectors of the  $F$  matrix, which describes the intrinsic frequencies of vibration of the amide groups and the coupling between the amide groups.<sup>2</sup>

This approach is explained in detail in this thesis and was used in all included publications for the simulations of the IR spectra. All calculations were performed with a Matlab program.

In Paper I, IR spectra were collected on mixtures of two species of amyloid- $\beta$  (A $\beta$ ) peptide, which aggregated and formed hetero-oligomers or homo-

oligomers in solution. Experiments with combinations of unlabeled and  $^{13}\text{C}$ -labeled peptides revealed that the two species mixed in the same structures. The calculations, carried out as support of the experimental results, were performed on  $\beta$ -sheet models. The  $F$  matrices were modified to include the presence of labeled peptides in the model structures, simulating the experimental molar fractions of labeled peptides. The simulation results confirmed not only the mixed oligomers, but also a random mixing of the two species in the same structures.

In Paper II we explored the homo-oligomers of A $\beta$  peptides both with experiments and simulations. The aim of the calculations was to identify the building block of the oligomers. This time, the calculations were performed on several antiparallel  $\beta$ -sheet models, on several antiparallel  $\beta$ -barrels from published PDB entries and a dodecamer structure. The presence of labeled peptides was simulated according to the experimental molar fractions using different labeling patterns, related to hypothetical structures of individual peptide molecules, i.e. to the building blocks of the oligomers. The simplest considered building blocks were single strand, 2-strand block and 3-strand block. Through the comparison of experimental and simulated  $^{13}\text{C}$ -band positions, we suggested that individual A $\beta$  peptides contribute at least two adjacent strands to the structure of the oligomers.

The aim of Paper III was the optimization of the parameters of transition dipole coupling, one of the computed interactions to describe the coupling between amide groups. The optimization of the parameters was the result of a comparison of the coupling constants between our Matlab calculations and Density functional theory calculations. Both types of calculation were performed on two  $\alpha$ -helices, two antiparallel  $\beta$ -sheets and two parallel  $\beta$ -sheets. The optimization procedure produced a good match of the coupling constants and the dipole strengths of the two types of calculations. From these, we suggested a new set of parameters for the calculation of the transition dipole coupling and other sets of parameters specific for each secondary structure.

In Paper IV, two proteins, Pent and SV2, were studied both by IR spectroscopy and simulations. The characteristic secondary structure of these two proteins is the parallel  $\beta$ -sheet. Surprisingly, the experiments revealed that both the proteins showed a small high wavenumber band, typically the hallmark of antiparallel  $\beta$ -sheets. Through simulations, we discovered that the high wavenumber band is mainly caused by the vibrations of the turns and not by the parallel  $\beta$ -sheet.

## 2. Proteins

Proteins are macromolecules, present in all living organisms. Proteins are involved in essentially all biological processes and carry out many different roles. For example, they serve as catalyst for chemical reactions, they transport or store small molecules (such as oxygen or metal ions), they can transmit signals, they are involved in DNA replication, they provide structural stability to cells through the cytoskeleton.

Proteins are widely described and illustrated in many biochemistry textbooks.<sup>3-5</sup> The two following sections present a general description of the composition and the structure of proteins. Then, a few examples of proteins, which were studied in Paper I, II and IV, are presented in the remaining sections of this chapter.

### 2.1 Composition

Proteins are polymers, consisting of one or more long chains of monomers, called amino acids. An amino acid is composed of a central carbon atom (called  $\alpha$ -carbon or  $C_\alpha$ ), linked to an amino group ( $\text{NH}_3^+$ ), to a carboxylic group ( $\text{COO}^-$ ), a hydrogen atom and a distinctive R group, which is different for each type of amide group and is usually referred to side chain. The amino acids can be distinguished according to their properties, such as size, shape, charge or solubility. Amino acids and their properties are listed in many textbooks.<sup>3-5</sup>

Proteins may also contain other functional groups (such as alcohols, carboxylic acids or basic groups) or *cofactors* (non-protein chemical compounds or metal ions). Most of them are reactive and are needed for the fulfillment of the biological function of proteins. It is also possible that a protein interacts with another and form complex assemblies.

### 2.2 Structure

Polypeptide structure can be described in term of primary, secondary, tertiary and quaternary structure. The primary structure is simply the sequence of the amino acids, which form a polypeptide chain when bonded together. It is

possible to use the term *protein* instead of *peptide* when the number of amino acids is higher than 50.<sup>3</sup> Each protein has a unique amino acid sequence. In the chain, the carboxylic group of one amino acid is linked to the amino group of another amino acid: this type of bond is called *peptide bond* (or *amide bond*). Each amino acid in the polypeptide chain is then called *residue*, while the group composed of the C=O of the first amino acid and the N-H of the second amino acid is called *peptide group* or *amide group*. This polypeptide chain is conventionally represented from the amino-terminal residue (N-terminus) to the carboxyl-terminal residue (C-terminus). The chain consists of a regularly repeating part, called *backbone* or *main chain*, and a variable part, called *side chain*, which is characteristic for each residue of the chain.

The peptide bond is typically planar and in trans configuration; instead, the torsional angles (dihedral angles) around the  $\alpha$ -carbon can assume different values. These torsion angles can be represented in the Ramachandran plot, where  $\phi$  is the angle between N and  $C_\alpha$  and  $\psi$  is the angle between  $C_\alpha$  and C. Not all combinations of dihedral angles are possible because of steric hindrance between the atoms. An example of a polypeptide chain is represented in Fig. 1.

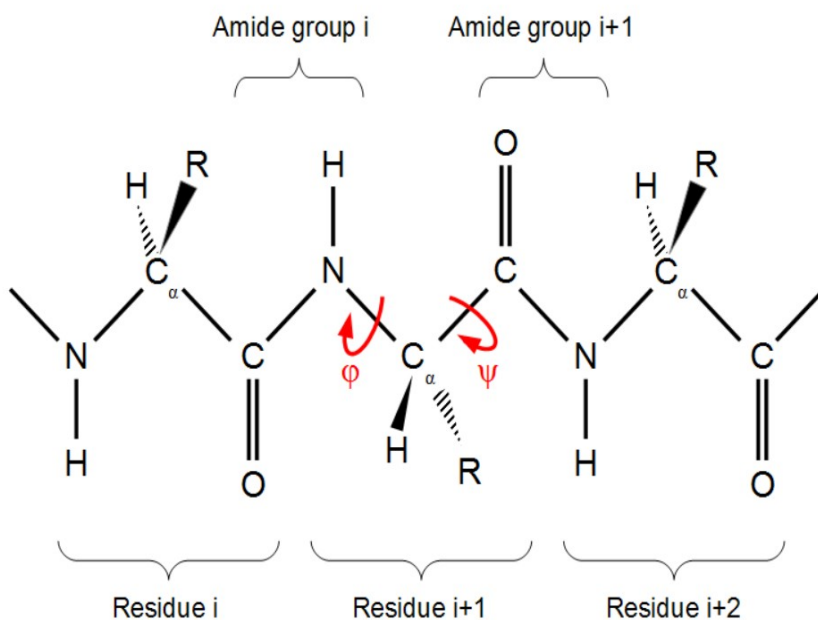


Figure 1. Representation of a section of a polypeptide chain in a protein. The dihedral angles are shown in red.

The polypeptide chain can fold into regular structures (secondary structures), stabilized by hydrogen bonds between the amide groups of the polypeptide chain. Certain residues (or group of residues) have the predisposition to arrange themselves into specific secondary structures or 3D conformations. The most common secondary structures, which are characterized by different dihedral angles, are  $\alpha$ -helix,  $\beta$ -sheet and *turn*. The  $\alpha$ -helix structure is similar to a screw, where the hydrogen bond is present between residue  $i$  and residue  $i+4$ . Further types of helices can be distinguished according to their dihedral angles, length and diameter.

In  $\beta$ -sheet structures, several strands are aligned and the hydrogen bond is located between adjacent strands. According to the two possible directions of the strands (antiparallel or parallel), two different hydrogen bond patterns between the strands can be found.  $\beta$ -sheets can be distinguished by the number of strands, the number of residues per strand and the planarity of the sheet.  $\beta$ -sheets can also bend themselves, forming a toroidal structure called  *$\beta$ -barrel*.

The *turn* is a particular motif that changes the direction of the polypeptide chain; in many turns, the hydrogen bond is present between residue  $i$  and residue  $i+3$ . *Loops* are composed of residues without any regular secondary structure and hydrogen bonds can be absent. The location of the secondary structures in a Ramachandran plot is illustrated in Fig. 2.

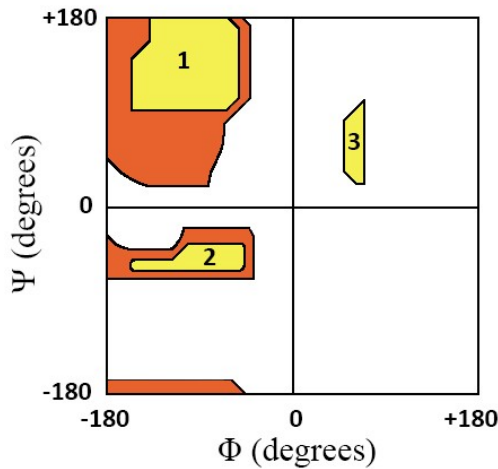


Figure 2. Ramachandran plot for all combinations of dihedral angles. The possible combinations to avoid steric hindrance are colored in orange. The region for the characteristic secondary structures are colored in yellow: 1)  $\beta$ -sheet; 2) right-handed  $\alpha$ -helix; 3) left-handed  $\alpha$ -helix. Adapted from Berg et al., 2004.<sup>3</sup>

The tertiary structure of a protein is the arrangement of the primary structure into a three dimensional (3D) shape. This process is called folding and the final state of this process is the native / functional state of the protein. Protein misfolding or protein aggregation can be the cause of neurodegenerative and neurological diseases (further explanation in Chapter 2.3).<sup>6,7</sup> The tertiary structure is based on the amino acid sequence. According to the hydrophobic effect, hydrophilic polar residues are exposed to the solvent and hydrophobic non polar residues are buried inside the protein.

The quaternary structure is the geometrical assembly of two or more polypeptide chains, stabilized by non covalent interactions. Examples of primary, secondary, tertiary and quaternary structures are shown in Fig. 3. The tertiary and quaternary structure in Fig. 3 are extracts of PDB entry 1OPF.<sup>8</sup>

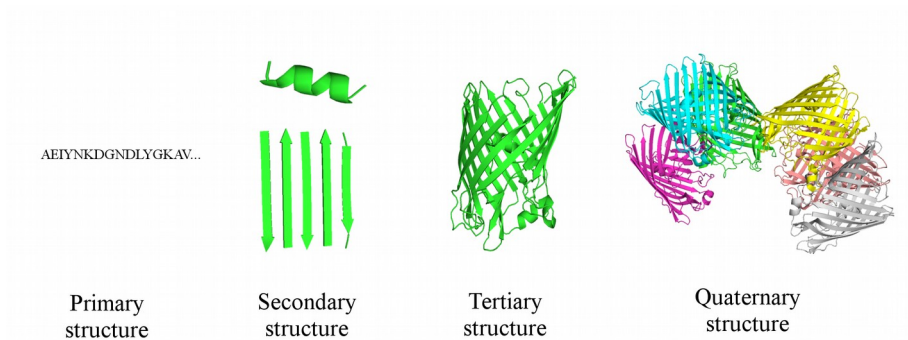


Figure 3. Examples of primary, secondary, tertiary and quaternary structures. In the secondary structure panel, an  $\alpha$ -helix and an antiparallel  $\beta$ -sheet are shown. Secondary, tertiary and quaternary structure panels were created with PyMol.

The 3D structures of proteins are usually determined using the techniques of X-ray crystallography, cryogenic electron microscopy or Nuclear Magnetic Resonance (NMR). Structural information can also be obtained from IR spectroscopy and circular dichroism.<sup>4,5</sup> The atomic coordinates are stored in many online databases, such as RSCB Protein Data Bank (PDB).<sup>9,10</sup>

## 2.3 Peptide aggregation and amyloid $\beta$

The human brain can be damaged by neurodegenerative diseases. In order to understand the possible reasons that lead to the neurodegeneration in the brain, neuropathologists started to study morphological lesions around 150 years ago.<sup>11</sup> Tissue deposits that were microscopically, and in severe cases macroscopically, visible in different organs (e.g. in the brain) have been analyzed and linked to various and apparently unrelated human disorders. Indeed, protein aggregation is a phenomenon that is common to many diseases, as Parkinson's, type 2 diabetes, Huntington's, prion and Alzheimer's disease.<sup>6,11–13</sup>

Amongst them, Alzheimer's disease is a neurodegenerative disorder distinguished by a progressive loss of cognitive abilities and typical neuropathological features. One of these peculiarities is the presence of specific toxic peptide aggregates, called amyloid plaques. The plaques arise in the neocortical area of preclinical Alzheimer's disease and spread into further brain areas in symptomatic patients.<sup>14</sup> The plaques are composed of fibrils, which are aggregates of monomeric A $\beta$  peptide.<sup>6,11,12,15–18</sup>

The A $\beta$  peptide is composed of 36–45 residues and created by the  $\beta$ - and  $\gamma$ -secretase-dependent cleavage of the amyloid precursor protein.<sup>6,7</sup> It is recov-

ered from the human body (serum and cerebrospinal fluid).<sup>6,7</sup> The most abundant isoform of A $\beta$  is composed of 40 amino acids, A $\beta$ <sub>40</sub>. Another important species is formed by 42 amino acids, A $\beta$ <sub>42</sub>; this is a less common species, but it is dominant in the plaques and it shows higher toxicity than A $\beta$ <sub>40</sub>.<sup>15,17,18</sup>

Aggregation of A $\beta$  and the structure of the aggregates have been studied for years using many different techniques and approaches, both experimental and computational.<sup>6,7,18-24</sup> It was then possible to study many morphological features of A $\beta$ , such as the formation of fibrils, protofibrils and oligomers depending on external factors, for instance temperature and solution salinity. Several structures of A $\beta$  have been observed using different techniques and experimental conditions. A $\beta$  species have been classified by size (number of monomers in the oligomer) and by shape.

Because of their physical properties, it is difficult to determine the structure of amyloid fibrils using X-ray crystallography or liquid state NMR.<sup>13,15,21,25,26</sup> Data obtained by X-ray diffraction revealed that the fibrils are arranged as a cross- $\beta$  structure, in other words an extended  $\beta$ -sheet (two sheets stacked upon each other) where the  $\beta$ -strands are perpendicular to the fibril axis and the hydrogen bonds between the  $\beta$ -strands are parallel to the same axis.

Very important results were also obtained using solid state NMR.<sup>13,15,16,21,22,25,27</sup> Many experiments in the last few years confirmed that the structure for the A $\beta$  fibrils is the parallel  $\beta$ -sheet.<sup>13,16,25,27</sup> It is possible that the soluble aggregates could be a mix between parallel and antiparallel  $\beta$ -sheets, however with a preponderance of the second type of structure. Results about the  $\beta$ -sheet structure for the A $\beta$  aggregates are confirmed by the techniques of Fourier Transform Infrared Spectroscopy (FTIR) and circular dichroism.<sup>14,17,19,20,22,24,28</sup> A very important result obtained by FTIR spectroscopy is the antiparallel  $\beta$ -sheet conformation found in A $\beta$  oligomers.<sup>24,28</sup> A representation of the aggregates is shown in Fig. 4. The monomer structure is an A $\beta$ <sub>42</sub> peptide and it is obtained from solution NMR (PDB entry 1Z0Q).<sup>29</sup> The oligomer structure is a tetramer of A $\beta$ <sub>42</sub> peptides and it is obtained from solution NMR (PDB entry 6RHY).<sup>30</sup> The shown structure for the fibril is an extract of twelve A $\beta$ <sub>42</sub> peptides, where their disordered residues from 1 to 10 are omitted, and it is obtained from solid state NMR (PDB entry 2MXU).<sup>31</sup>

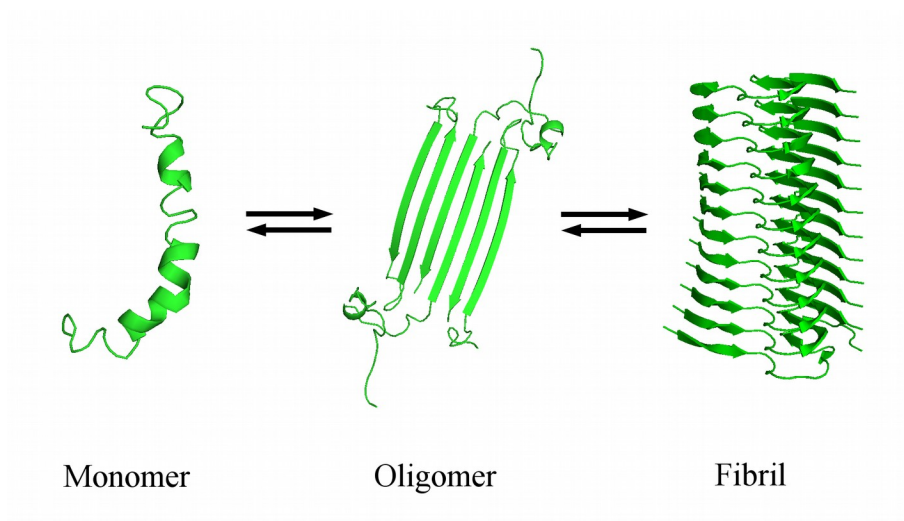


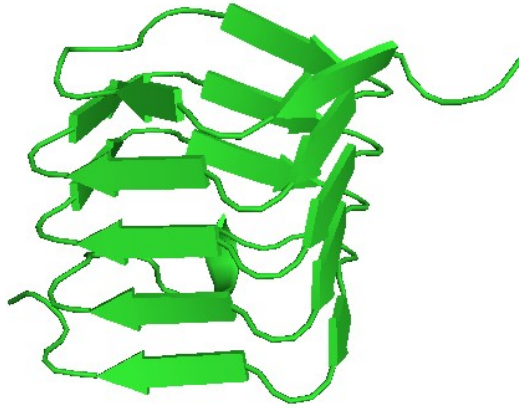
Figure 4. Representation of the amyloid- $\beta$  aggregates. The structures were created with PyMol.

## 2.4 $\beta$ -helix proteins

The  $\beta$ -helix is a particular type of secondary structure. It is composed of parallel  $\beta$ -strands that are arranged in a helical pattern. This sort of helix can have several faces. An example of a  $\beta$ -helix is shown in Fig. 5. Similarly to  $\beta$ -sheets, the hydrogen bonds are located between adjacent strands. In Paper IV we studied two  $\beta$ -helix proteins, SV2<sup>32</sup> and Pent.<sup>33</sup>

SV2 is a protein located in synaptic vesicles, i.e. vesicles that contain neurotransmitters in synapses. It is therefore involved in the transmission of nerve impulses between neurons. This protein has three different isoforms: SV2A, SV2B and SV2C. The structure used in our calculations and shown in Fig. 5 is the SV2C isoform. In this study, this protein receptor forms a complex with Botulinum neurotoxin, a toxin which can cause botulism to humans; the PDB structure was obtained from x-ray diffraction. The  $\beta$ -helix, located in the luminal domain of SVC, is composed of four faces, each one of them has four/five strands.

Pent is a pentapeptide repeat protein (also called PRP). Each protein of this family has four pentapeptide repeats that form one of the helical turns of the  $\beta$ -helix. It is extracted from *Clostridium botulinum*, the bacterium that produces Botulinum neurotoxin. Pent has 217 residues and its  $\beta$ -helix is composed of four faces with eight strands.



*Figure 5. Example of a  $\beta$ -helix. The protein is SV2 and it is an extract of PDB entry 6ESI.<sup>32</sup> The structure was created with PyMol.*

### 3. Infrared spectroscopy

IR spectroscopy is a powerful technique that allows one to retrieve a lot of structural information about the proteins of interest. The main application of IR spectroscopy in protein science is the analysis of the secondary structure, but this technique can also be useful to understand the processes of protein reactions such as protein folding and protonation. This method is also useful because it can be applied to proteins in many environmental conditions, in aqueous and non-aqueous solution or in membranes.

IR spectroscopy is widely described in several textbooks and reviews.<sup>34–38</sup> In the following sections, the description of IR spectroscopy is presented according to classical physics and quantum physics. Then, a brief discussion about the recording and the interpretation of the spectra is given.

In IR spectroscopy, the IR light is absorbed when the frequency of the light is the same as that of the vibration. Thus, IR spectra are related to the frequencies of the molecular vibrations. In consequence, we can deduce information about the structure from spectral parameters such as band position, bandwidth and absorption coefficient.

There are many types of vibration in a protein. In stretching vibrations, the length of the bond changes during the vibration. In bending vibrations, the angle between two bonds changes during the vibration. In torsional vibrations, the torsional angle changes during the vibration.

Bond lengths and angles can be called *internal coordinates*, because they can be used to define the structure instead of Cartesian coordinates. When the vibrations of several internal coordinates couple to generate a movement with the same frequency of vibration and for each atom and if the involved atoms pass through the equilibrium position at the same time, we can call the resulting pattern *normal mode of vibration*. Using the normal modes, we can approximate each vibrational mode to a harmonic oscillator. In the harmonic approximation, a non-linear protein with  $N$  atoms has  $3N-6$  vibrational degrees of freedom ( $3N$  degrees of freedom, minus 3 translations and 3 rotations of the whole protein), i.e.  $3N-6$  normal modes of vibration. In addition, in the harmonic approximation the vibration of a normal mode is independent from the vibration of the other normal modes. For each normal mode, it is convenient to use *normal coordinates*, which are proportional to the deviation from the equilibrium position of each atom, i.e. proportional to linear combinations of Cartesian of internal coordinates.

### 3.1 Classical description

An important parameter of the electromagnetic radiation in the description of IR spectroscopy is the wavenumber  $\tilde{\nu}$ , which is defined as:

$$\tilde{\nu} = \frac{\nu}{(c/n)} = \frac{1}{\lambda} \quad E = h\nu = \frac{hc\tilde{\nu}}{n} ,$$

where  $c$  is the speed of light,  $n$  is the refractive index of the medium,  $\nu$  is the frequency,  $\lambda$  is the wavelength and  $h$  is the Planck's constant. The IR spectra are usually plotted from high to low wavenumbers, because they are then similar to spectra from short to long wavelengths.

We can consider two atoms in the molecule as a system of two oscillating masses joined by a massless spring. The frequency of this harmonic oscillator is given by the formula:

$$\nu = \frac{1}{2\pi} \sqrt{k \left( \frac{1}{m_1} + \frac{1}{m_2} \right)} = \frac{1}{2\pi} \sqrt{\frac{k}{m_r}} ,$$

where  $\nu$  is the frequency in cycles per second,  $k$  is the force constant in N/m,  $m_1$  and  $m_2$  are the masses of the atoms in g and  $m_r$  is the reduced mass in g. The previous formulas show that the frequency is a function of the masses of the two oscillating atoms and of the force constant.

In the classic description, the interaction between the vibrating bond and the electromagnetic radiation is seen as the interaction between the partial charges of the vibrating bond and the electric field component of the radiation. The amplitude of vibration is increased when the frequency of the electric field matches the frequency of the vibrating bond. This interaction does not change the frequency of vibration.

The presence of the partial charges in the bond creates a dipole moment:

$$\mu = qR ,$$

where  $\mu$  is the dipole moment,  $q$  is the value of the partial charges and  $R$  is the distance between the atoms. The change in the dipole moment with respect to the distance is:

$$\frac{\partial \mu}{\partial R} = \frac{\partial \mu}{\partial R}(R_0) = q .$$

The larger the partial charges, the larger the absorption. One of the *selection rules* for the absorption in IR spectroscopy states that a vibration is infrared active if the dipole moment changes during the vibration. An example of application of this selection rule is the stretching of the linear molecule of CO<sub>2</sub>. In the symmetric stretching vibration, the dipole moments of the individual C=O bonds have the same values but opposed signs; so, the overall dipole moment is zero at any time and the vibration is infrared inactive. Instead, in the antisymmetric stretching vibration, the dipole moments of the individual C=O bonds have different signs and different amplitudes; so, there is an

overall dipole moment that changes with time and the vibration is infrared active.

## 3.2 Quantum mechanical description

We know that the potential energy for a harmonic oscillator of a diatomic system is given by this formula:

$$V = \frac{1}{2} k x^2 \quad .$$

This parabolic potential is an approximation of the Morse potential, which describes the potential energy curve of a covalent bond. During the vibration of the bond, the oscillator moves along the curve, exchanging potential energy and kinetic energy. The total energy is always constant and equal to the maximum potential energy.

From the solution of the time-independent Schrödinger equation, it is possible to calculate the wavefunctions of the vibrational states for each quantum number and the quantum mechanical probability of finding the oscillator at a specific distance.

The corresponding discrete energy levels, eigenvalues of the Schrödinger equation, are:

$$E_n = \left( n + \frac{1}{2} \right) h\nu \quad ,$$

where  $\nu$  is the classical vibrational frequency of the oscillator and  $n$  is a quantum number (expressed as integer values). Energy levels, wavefunctions and probability functions are shown in Fig. 6.

According to Fermi's golden rule, the probability ( $P$  in the following) of transition from one vibrational level to the next one is proportional to this expression:

$$P(|\Psi_0\rangle \rightarrow |\Psi_1\rangle) \propto |\langle \Psi_1 | V | \Psi_0 \rangle|^2 \quad ,$$

where  $|\Psi_0\rangle$  is the initial state,  $|\Psi_1\rangle$  is the final state and  $V$  is the operator of the perturbation. The probability is then proportional to the projection of vector  $V|\Psi_0\rangle$  on vector  $|\Psi_1\rangle$ ; i.e. proportional to how much the two vectors are similar. The maximum of probability is reached when the two vectors have the same direction, instead the minimum of probability is obtained when the two vectors are orthogonal.

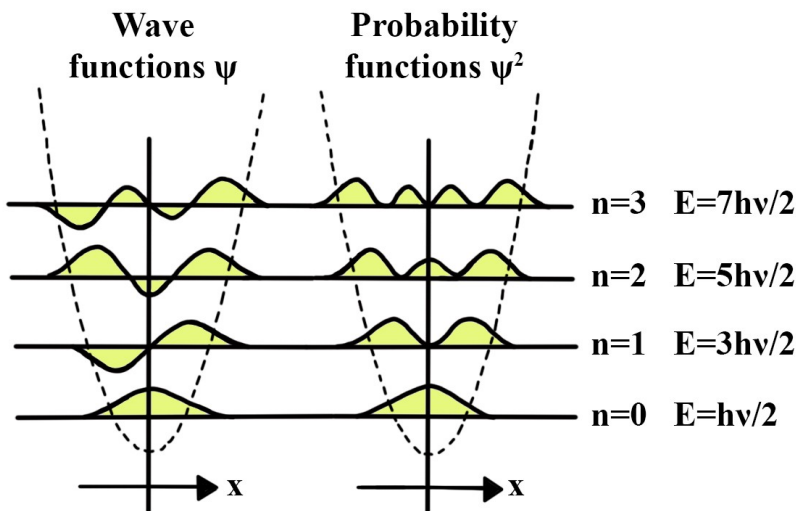


Figure 6. Quantum numbers, energy levels, wave functions and probability functions for the quantum mechanical harmonic oscillator. Adapted from Colthup et al, 1990.<sup>34</sup>

Similarly to the classic description, the interaction between the electric field of the IR radiation and the vibrating bond is approximated by the interaction with the dipole moment, which is described by the following operator:

$$\mu = \sum_{i=1}^N q_i \mathbf{r}_i \quad ,$$

where  $q_i$  is the charge of the particle,  $\mathbf{r}_i$  the position operator and  $N$  is the number of charged particles. Likewise the definition of energy of the interaction of a dipole in an electric field in the classical description, the interaction energy between the bond and the radiation can be expressed by:

$$V(t) = -\mu E(t) \quad ,$$

where  $V$  is the interaction energy (the perturbation) and  $E$  the oscillating electric field vector of the IR radiation.

The formula for the probability of transition changes accordingly:

$$\begin{aligned} \langle \Psi_1 | V | \Psi_0 \rangle &= -\langle \Psi_1 | \mu | \Psi_0 \rangle E \\ P(|\Psi_0\rangle \rightarrow |\Psi_1\rangle) &\propto |\langle \Psi_1 | \mu | \Psi_0 \rangle|^2 E^2 \cos^2 \alpha \quad , \end{aligned}$$

where  $\alpha$  is the angle between the electric field  $E$  and  $\langle \Psi_1 | \mu | \Psi_0 \rangle$ . Two quantities can then be defined, the *Transition Dipole Moment (TDM)* and the *dipole strength*, respectively:

$$\begin{aligned} \text{TDM: } \mu_{10} &= \langle \Psi_1 | \mu | \Psi_0 \rangle = \int \Psi_1^* \mu \Psi_0 \\ \text{Dipole strength: } D_{10} &= |\langle \Psi_1 | \mu | \Psi_0 \rangle|^2 \quad . \end{aligned}$$

The transition probability is highest if the TDM and the electric field of the IR radiation have the same direction and is zero when they are orthogonal. Considering the Born-Oppenheimer approximation, the nuclei are much heavier and slower than the electrons. A vibrational state  $|\Psi\rangle$  can be then written as a product of an electronic and a nuclear state vector:  $|\psi\phi\rangle$ . As consequence, the TDM can be written as:

$$TDM: \mu_{mn} = |\langle \psi_0 \phi_m | \mu | \psi_0 \phi_n \rangle|$$

After the transition, the electrons are still in the electronic ground state  $\psi_0$ ; the different indices for the nuclear wavefunctions  $\phi$  indicate the transition from vibrational state  $n$  to vibrational state  $m$ . A further *selection rule* states that the difference between the quantum numbers in vibrational transitions is  $\pm 1$ .

Usually, the oscillators are in the ground state and the absorption of IR radiation leads to a transition to the first excited state. In this situation, the TDM of the oscillators can be expressed as:

$$TDM = \frac{\partial \mu}{\partial R} (R_0) \left( \frac{h}{8 \pi^2 c m_r \tilde{\nu}} \right)^{0.5},$$

where  $\partial \mu / \partial R(R_0)$  is the change of the dipole moment at the equilibrium position  $R_0$  (*dipole derivative*), while the factor can be calculated as  $4.1058/\tilde{\nu}^{0.5}$  u<sup>0.5</sup> Å cm<sup>-0.5</sup>.

### 3.3 Instrumentation

The IR spectrum is registered by a classical IR spectrometer or by FTIR spectrometers. The modern instrumentation is based on the latter type of spectrometers. The core of the instrument is a Michelson interferometer (shown in Fig. 7), which has a movable and a fixed mirror. The IR light, emitted from a source, reaches a beam splitter. Part of the light is sent to the fixed mirror, is reflected back to the splitter and passes through it; another part is sent to the movable mirror, is reflected back to the splitter and then re-reflected in direction of the detector. The sample is located between the beam splitter and the detector. The two parts of the light are recombined after the second passage through the beam splitter and there can be constructive or destructive interference according to the position of the movable mirror, i.e. according to the length difference of the two light paths. The instrument measures the light intensity as function of the position of the movable mirror (*interferogram*), then makes a Fourier transformation to calculate the IR spectrum.

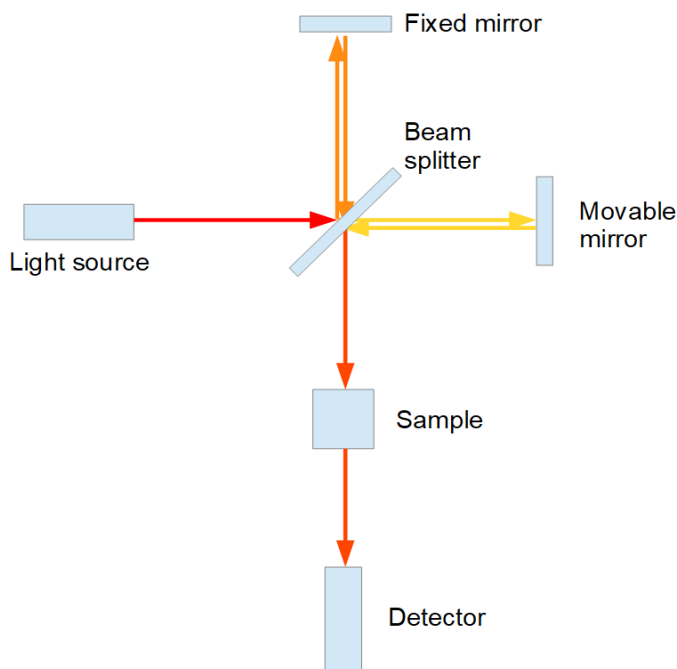


Figure 7. Scheme of a Michelson interferometer.

### 3.4 Interpretation of infrared spectra

The spectrum of IR light spectrum is wide; if we consider the mid-infrared region, it extends from  $4000\text{ cm}^{-1}$  to  $400\text{ cm}^{-1}$ . The protein backbone has several absorption bands in this region.<sup>35,37,39,40</sup> In the following list, the main bands are listed, with their approximate wavenumbers and their main internal coordinate contributions:

- Amide A ( $3300\text{ cm}^{-1}$ ): N-H stretching;
- Amide B ( $3100\text{ cm}^{-1}$ ): N-H stretching;
- Amide I ( $1650\text{ cm}^{-1}$ ): C=O stretching, C-N stretching;
- Amide II ( $1550\text{ cm}^{-1}$ ): C-N stretching, N-H bending;
- Amide III ( $1300\text{ cm}^{-1}$ ): C-N stretching, N-H bending;
- Amide IV ( $625\text{ cm}^{-1}$ ): C-O bending, C-C stretching;
- Amide V ( $650\text{ cm}^{-1}$ ): C-N torsion, out-of-plane N-H bending;
- Amide VI ( $550\text{ cm}^{-1}$ ): out-of-plane C=O bending;
- Amide VII ( $200\text{ cm}^{-1}$ ): out-of-plane N-H bending, C-N-C deformation.

Since the main application of IR spectroscopy is the analysis of the secondary structures of proteins, the IR spectrum between  $1800\text{ cm}^{-1}$  and  $1500\text{ cm}^{-1}$  is the region of interest, in particular the amide I band ( $1700\text{--}1600\text{ cm}^{-1}$ ). The amide I band is highly sensitive to the protein backbone conformation, so to the secondary structure of the protein. Plus, few other vibrations do not absorb in this range, so the absorption in amide I band is not influenced by them and it is easier to analyze the spectrum.

Each secondary structure absorbs in a typical spectral range, listed below:<sup>36,41</sup>

- $\alpha$ -helix:  $1650\text{ cm}^{-1}$ ;
- $\beta$ -sheet:  $1630\text{ cm}^{-1}$  and  $\sim 1690\text{ cm}^{-1}$  (in  $\text{H}_2\text{O}$ );
- turn:  $1670\text{ cm}^{-1}$ ;
- irregular:  $1650\text{ cm}^{-1}$ .

These are typical values, but in some cases the secondary structures absorb in different ranges. Antiparallel  $\beta$ -sheets shows two bands in the listed spectral range.<sup>42,43</sup> However, the IR spectrum of a parallel  $\beta$ -sheet does not usually reveal the high wavenumber band (this argument will be discussed thoroughly in Paper IV). Parallel  $\beta$ -sheets present their main absorption band at  $1630\text{ cm}^{-1}$ , with smaller bands between  $1680\text{ cm}^{-1}$  and  $1650\text{ cm}^{-1}$ .<sup>1,44</sup>  $\alpha$ -helices show a main absorption band at  $1650\text{ cm}^{-1}$ , but its position can increase until  $1680\text{ cm}^{-1}$ , depending on the length of the helix.<sup>41</sup> Turns absorb in the region between  $1700\text{ cm}^{-1}$  and  $1630\text{ cm}^{-1}$ , depending on their dihedral angles.<sup>37,45,46</sup> Amide groups that cannot be classified as  $\alpha$ -helix and  $\beta$ -sheets can absorb throughout the entire amide I spectral range.<sup>1</sup>

These characteristic IR spectra for the different secondary structures can be influenced by several factors. For example, the IR spectrum of  $\beta$ -sheets is influenced by the planarity of the sheet and by the number of strands. A factor to consider is the hydrogen bond effect, that makes the bands shift. Amino acid side chains from Gly, Asn and Arg also absorb in the amide I region.<sup>47–50</sup> If the IR spectroscopy experiment is performed in  $\text{H}_2\text{O}$ , the H-O-H bending vibration is located in the amide I band; to avoid this contribution, experiments can be performed in  $\text{D}_2\text{O}$ .

If different secondary structures are present in the tertiary structure of the protein of interest, their absorption band can overlap in the same amide I band region. There are two methods to separate the overlapping bands and to make an assessment of the contributions of the secondary structures. The first one is to calculate the second derivative of the absorption spectrum. In this way the width of the main band is reduced and the minima correspond to the maxima absorbance of the original IR spectrum. An example is shown in Fig. 8.

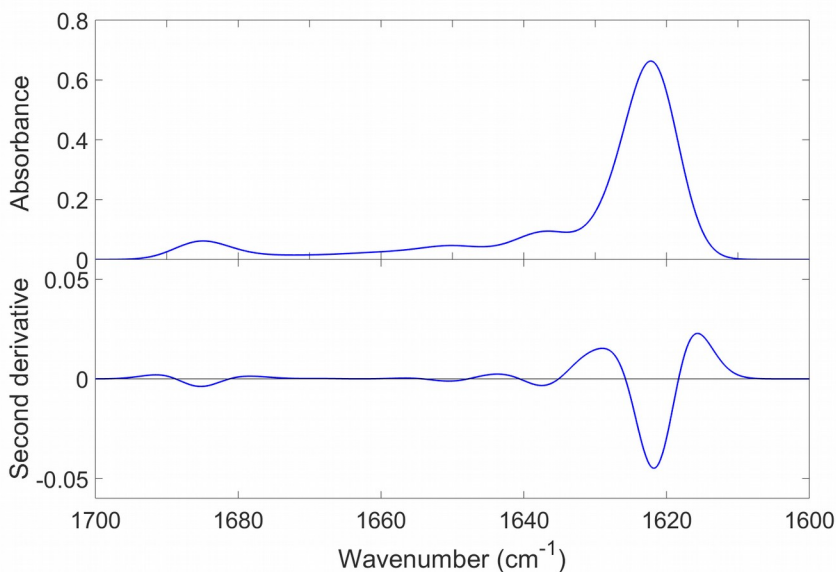


Figure 8. Example of an IR spectrum and its second derivative calculated for an antiparallel  $\beta$ -sheet with 8 strands and 10 residues per strand.

The second method is Fourier self-deconvolution.<sup>51–53</sup> In this method, the interferogram (Fourier transform of the IR spectrum) is multiplied by an increasing function; applying a second time a Fourier transform, we will obtain a narrower spectrum than the original one.

## 4. Methods for the simulation of the amide I band

The interpretation of the amide I band can provide important structural information, but it is not always straightforward. To help in the understanding of the experimental results, several theoretical approaches were developed to simulate the amide I band, one of which is based on normal mode analysis.<sup>1,37,42,44,54–57</sup> A normal mode analysis allows the study of the vibrations of a molecule, providing information on the frequencies and on the atomic movements.

The method used in all included publications is based on the floating oscillator model:<sup>1</sup> in this approximation, each amide group is seen as a vibrating oscillator, characterized by an intrinsic frequency of vibration and a TDM. The coupling between the oscillators and the intrinsic frequencies are influenced by different interactions, which will be described in the following sections. This information is needed to calculate the  $F$  matrix, a matrix whose elements are force constants obtained from Newton's equations. From the  $F$  matrix, the wavenumbers and intensities of the amide I oscillators can be retrieved by solving its eigenvalue and eigenvector problem.<sup>2</sup> The IR spectrum is then calculated by creating Gaussian or Lorentzian bands with the amide I intensities for each amide I normal mode and by summing them.

In all the included publications, the calculation of the  $F$  matrix elements, the diagonalization of the  $F$  matrix and the creation of the simulated IR spectrum are performed in a Matlab program.<sup>58,59</sup> This program only requires as input to specify atomic coordinates (generally using a PDB file) and which parameters to use in the calculation.

### 4.1 Theoretical description of the $F$ matrix

The number of amide I oscillators in a protein is equal to the number of amide groups. We can consider them as harmonic oscillators, which are coupled to each other. This coupling creates collective vibrations, which are delocalized over several amide groups.

The main internal coordinate contribution in the amide I band is provided by the vibration of the C=O bond of the amide group. We can consider this bond as a diatomic system, i.e. two atoms with masses  $m_1$  and  $m_2$ , and the corresponding equations of motion for a harmonic diatomic oscillator:

$$\begin{aligned} m_1 \ddot{x}_1 &= -f(x_1 - x_2) = -fx_1 + fx_2 \\ m_2 \ddot{x}_2 &= -f(x_2 - x_1) = fx_1 - fx_2 \end{aligned}$$

In the equations,  $x_1$  and  $x_2$  are the displacements of the two atoms from the equilibrium positions along the  $x$  axis and  $f$  is the force constant. The solutions of the equations are:

$$\begin{aligned} x_1 &= a_1 \sin(\omega t) \\ x_2 &= a_2 \sin(\omega t) \end{aligned}$$

Inserting these solutions into the equations of motion gives:

$$\begin{aligned} -m_1 \omega^2 a_1 &= -f a_1 + f a_2 \\ -m_2 \omega^2 a_2 &= +f a_1 - f a_2 \end{aligned}$$

These equations can be rearranged in the following way:

$$\begin{aligned} (f - m_1 \omega^2) a_1 - f a_2 &= 0 \\ -f a_1 + (f - m_2 \omega^2) a_2 &= 0 \\ \downarrow \\ \begin{pmatrix} f - m_1 \omega^2 & -f \\ -f & f - m_2 \omega^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned}$$

The matrix is called *coefficient matrix*. This system of equations can be solved by setting the determinant of the coefficient matrix to zero; this gives two solutions for  $\omega$ ,  $a_1$  and  $a_2$ :

$$\begin{aligned} \omega_1^2 = 0 &\rightarrow \frac{a_1}{a_2} = 1 \\ \omega_2^2 = f \left( \frac{1}{m_1} + \frac{1}{m_2} \right) &= \frac{f}{m_r} \rightarrow \frac{a_1}{a_2} = -\frac{m_2}{m_1} \end{aligned}$$

The first solution means that there is a translation of the atoms along the  $x$  axis; the second solutions means that the two displacements have opposite signs and that the ratio of the displacements is inversely proportional to the ratio of the masses.

We can rewrite the coefficient matrix and model the equations into an eigen-value problem for the  $\omega$  values using the  $F$  matrix:

$$\begin{aligned}
\omega^2 a_1 &= (f/m_1) a_1 - (f/m_1) a_2 \\
\omega^2 a_2 &= -(f/m_2) a_1 + (f/m_2) a_2 \\
&\downarrow \\
\begin{pmatrix} f/m_1 - \omega^2 & -f/m_1 \\ -f/m_2 & f/m_2 - \omega^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
&\downarrow \\
\mathbf{F} \mathbf{a} &= \omega^2 \mathbf{a}
\end{aligned}$$

In the following, the normal coordinate  $q$  is related to the displacements  $x_1$  and  $x_2$  and to the internal coordinate  $s$ . This is done by expressing the potential energy in the normal and internal coordinates.

$$\begin{aligned}
s &= x_1 - x_2 \\
V &= 1/2 \omega^2 q^2 = 1/2 f s^2 \\
f (x_1 - x_2)^2 &= \omega^2 q^2 = \frac{f}{m_r} q^2 \\
q &= m_r^{1/2} (x_1 - x_2) = m_r^{1/2} s
\end{aligned}$$

We can expand the same reasoning to the coupling between two two-atomic oscillators. The equations of the motion using the internal coordinates are:

$$\begin{aligned}
m_{r1} \ddot{s}_1 &= -f_1 s_1 + f_{12} s_2 \\
m_{r2} \ddot{s}_2 &= -f_2 s_2 + f_{12} s_1
\end{aligned}$$

In the previous equations,  $s_1$  and  $s_2$  are the internal coordinates for the two oscillators,  $f_1, f_2$  and  $f_{12}$  are the force constants:

$$f_{ij} = \frac{\partial^2 V}{\partial s_i \partial s_j}.$$

The solutions of the equations of the motion using the internal coordinates are:

$$\begin{aligned}
s_{1w} &= a_{1w} \sin(\omega t) \\
s_{2w} &= a_{2w} \sin(\omega t)
\end{aligned}$$

Consequently, we can insert these solutions into the equations of the motion. We obtain this system of equations and the coefficient matrix for the diatomic can be found in this way:

$$\begin{aligned}
-m_{r1} \omega^2 a_{1w} &= -f_1 a_{1w} + f_{12} a_{2w} \\
-m_{r2} \omega^2 a_{2w} &= -f_2 a_{2w} + f_{12} a_{1w} \\
&\downarrow \\
(f_1 + m_{r1} \omega^2) a_{1w} - f_{12} a_{2w} &= 0 \\
-f_{12} a_{1w} + (f_2 - m_{r2} \omega^2) a_{2w} &= 0 \\
&\downarrow \\
\begin{pmatrix} f_1 - m_{r1} \omega^2 & -f_{12} \\ -f_{12} & f_2 - m_{r2} \omega^2 \end{pmatrix} \begin{pmatrix} a_{1w} \\ a_{2w} \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\end{aligned}$$

Similarly to the single two-atomic oscillator, we can define the  $F$  matrix and its eigenvalue problem.

$$\begin{aligned}
 \omega^2 a_{1\omega} &= (f_1/m_{r1}) a_{1\omega} - (f_{12}/m_{r1}) a_{2\omega} \\
 \omega^2 a_{2\omega} &= (f_2/m_{r2}) a_{2\omega} - (f_{12}/m_{r2}) a_{1\omega} = - (f_{12}/m_{r2}) a_{1\omega} + (f_2/m_{r2}) a_{2\omega} \\
 &\downarrow \\
 \begin{pmatrix} f_1/m_{r1} - \omega^2 & -f_{12}/m_{r1} \\ -f_{12}/m_{r2} & f_2/m_{r2} - \omega^2 \end{pmatrix} \begin{pmatrix} a_{1\omega} \\ a_{2\omega} \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
 &\downarrow \\
 \mathbf{F} \mathbf{a} &= \omega^2 \mathbf{a}
 \end{aligned}$$

The  $F$  matrix is not symmetrical if the reduced masses are different. Converting to normal coordinates, we obtain a new symmetrical  $F$  matrix:

$$\begin{aligned}
 m_r^{1/2} s &= q & m_r^{1/2} a &= q_0 \\
 &\downarrow \\
 \omega^2 q_{01\omega} &= (f_1/m_{r1}) q_{01\omega} - f_{12} q_{02\omega} / (m_{r1} m_{r2})^{1/2} \\
 \omega^2 q_{02\omega} &= -f_{12} q_{01\omega} / (m_{r1} m_{r2})^{1/2} + (f_2/m_{r2}) q_{02\omega} \\
 &\downarrow \\
 \begin{pmatrix} f_1/m_{r1} & -f_{12}/(m_{r1} m_{r2})^{1/2} \\ -f_{12}/(m_{r1} m_{r2})^{1/2} & f_2/m_{r2} \end{pmatrix} \begin{pmatrix} q_{01\omega} \\ q_{02\omega} \end{pmatrix} &= \omega^2 \begin{pmatrix} q_{01\omega} \\ q_{02\omega} \end{pmatrix}
 \end{aligned}$$

The diagonal elements of the  $F$  matrix are related to the intrinsic vibrations of the amide I oscillators of each amide group, i.e. the vibrations where there is no coupling between the amide groups; the non-diagonal elements are related to the interactions / coupling between these intrinsic vibrations of the amide groups.

In the  $F$  matrix, it is convenient to express the masses in u and the force constants in mdy  $\text{\AA}^{-1}$ . Using these units, the diagonal elements can be converted into wavenumbers using the following formula:<sup>2,34</sup>

$$\tilde{\nu}_0 = 1303 (f/m_r)^{1/2}$$

For example, using  $1650 \text{ cm}^{-1}$  as wavenumber, the corresponding diagonal element is  $1.605 \text{ mdy } \text{\AA}^{-1} \text{ u}^{-1}$ . At the same time, the conversion formula to obtain the coupling constants (non-diagonal elements) in the unit of  $\text{cm}^{-1}$  is:<sup>37</sup>

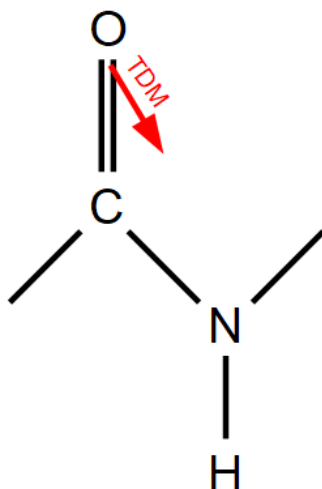
$$\tilde{\nu} = \frac{848619}{\tilde{\nu}_0} F_{ij} \quad ,$$

where  $F_{ij}$  is the  $F$  matrix element in  $\text{mdyn } \text{\AA}^{-1} \text{ u}^{-1}$ ,  $\tilde{\nu}_0$  is the wavenumber in  $\text{cm}^{-1}$ . Suitable values for  $\tilde{\nu}_0$  are  $1707 \text{ cm}^{-1}$  or  $1717 \text{ cm}^{-1}$  (amide I wavenumber of NMA in gas phase)<sup>60,61</sup> or  $1650 \text{ cm}^{-1}$  (average wavenumber of an amide I oscillator in a protein).<sup>1,2</sup>

## 4.2 Transition dipole coupling

Transition Dipole Coupling (TDC) is an approximation to describe the electrostatic coupling of individual oscillators. It was suggested by Krimm and Abe in the 1970s,<sup>62,63</sup> and is widely used to calculate the amide I band.<sup>58,64–71</sup>

In the calculation of this interaction, each amide group has its own TDM (as shown in Fig. 9) and the coupling between the amide groups can be approximated by a dipole-dipole interaction (as shown in Fig. 10). The TDM can be described by position, magnitude and direction, but there is still no agreement on the optimal values of the TDM parameters in literature.<sup>1,42,44,54,72–76</sup>



*Figure 9. Illustration of the TDM (in red) of an amide group.*

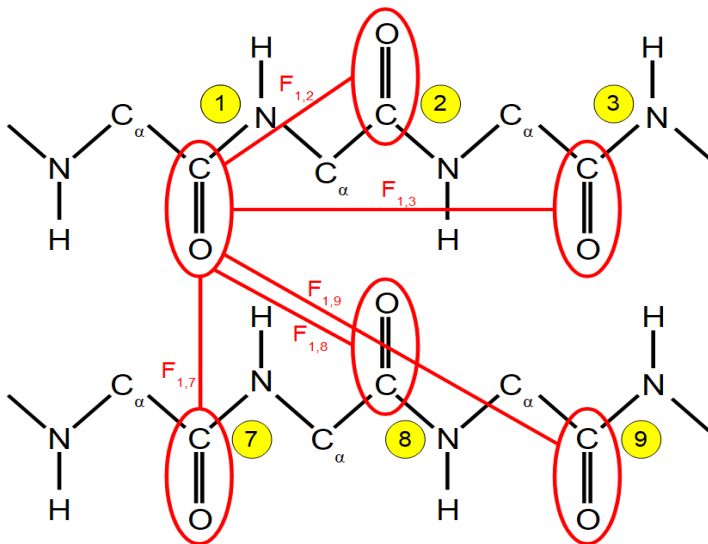


Figure 10. Scheme of TDC interactions (in red) between amide groups inside a parallel  $\beta$ -sheet (only the backbone is depicted). The numbering of the amide groups is shown in yellow circles.

The coupling is more effective when the amide groups vibrate with the same frequency. TDC is a function not only of the magnitude of the TDMs, but also of their orientations. According to TDC, the calculation of the coupling between the amide groups, i.e. the calculation of the non-diagonal elements of the  $F$  matrix, is expressed in the following formula:

$$F_{jk} = \frac{0.1}{\epsilon} \frac{\frac{\partial \mu_j}{\partial q_j} \cdot \frac{\partial \mu_k}{\partial q_k} - 3 \left( \frac{\partial \mu_j}{\partial q_j} \cdot \mathbf{n}_{jk} \right) \left( \frac{\partial \mu_k}{\partial q_k} \cdot \mathbf{n}_{jk} \right)}{R_{jk}^3} .$$

In the formula,  $\partial \mu_\alpha / \partial q_\alpha$  is the dipole derivative of the amide group  $\alpha$  in  $\text{D} \text{ \AA}^{-1} \text{ u}^{-1/2}$ ,  $R_{jk}$  is the distance between the two dipole derivatives in  $\text{\AA}$ ,  $\mathbf{n}_{jk}$  is the unit vector of the distance,  $\epsilon$  is the dielectric constant (assumed to be unity) and 0.1 is the factor for conversion from  $\text{D}^2 \text{ \AA}^{-5}$  to  $\text{mdyn} \text{ \AA}^{-1}$ . The TDM, as explained in Section 3.2, is proportional to the dipole derivative.

TDC is used in many calculations and it usually provides good agreement with experimental results.<sup>66,69,77</sup> It describes long range electrostatic interactions well, but it is inadequate to describe short range interactions, such as interaction between neighboring amide groups.<sup>55,56,72,73,78,79</sup> This limitation can be overcome using data from quantum chemical calculations (see the following Section on Nearest neighbor coupling). In addition, TDC coupling does not include multipoles interactions and mechanical interactions.

TDC was considered in the calculations of all included publications. In particular, in Paper III we suggested our optimized set of parameters for the TDM and we compared the results with sets from literature.

### 4.3 Nearest neighbor coupling

The interaction between amide groups has an electrostatic and a mechanical component. As explained in Section 4.2, when the interaction is long range, the approximation used in TDC, i.e. the mechanical component is not included, is appropriate to describe the interaction.

However, this is proven not to be true when the interaction is short range.<sup>55,56,72,73,78,79</sup> In addition, it was shown that  $C_\alpha$  vibrates along the other atoms involved in amide I vibrations.<sup>80</sup> So, TDC (as well as other approaches that rely only on electrostatic coupling) is not sufficient to correctly describe the interaction between nearest neighbors and mechanical coupling cannot be neglected.

A better way to describe short range interactions is to perform quantum chemical calculations because they consider both the mechanical and the electrostatic coupling.<sup>55,56,70,72,73,78,81–84</sup> These calculations were performed on small molecules (usually dipeptides) after a geometry optimization. The coupling between nearest neighbors ( $F_{i,i+1}$ ) can be expressed as a function of the dihedral angles  $\phi$  and  $\psi$  of the amide groups. Consequently, the nearest neighbor couplings were calculated for pairs of dihedral angles, covering the entire Ramachandran plot. Several maps were created using different levels of theory and different basis sets.<sup>55,56,78,85</sup> Stock and co-workers stated that these coupling constants maps are not sensitive to these different computational details.<sup>78</sup> The coupling constant maps can be used to describe nearest neighbor interactions in larger structures.<sup>78,81</sup> Nearest neighbor coupling was included in the calculations of Paper I, II and IV.

### 4.4 Local environment

It was proven that the local environment, i.e. the nearest neighbors on both sides of an amide group, can influence the intrinsic frequencies of the amide groups.<sup>55,56,78,85</sup> The effect of the local environment can be described using the pairs of dihedral angles  $\phi$  and  $\psi$  of the amide groups before and after each amide group, as shown in Fig. 11.

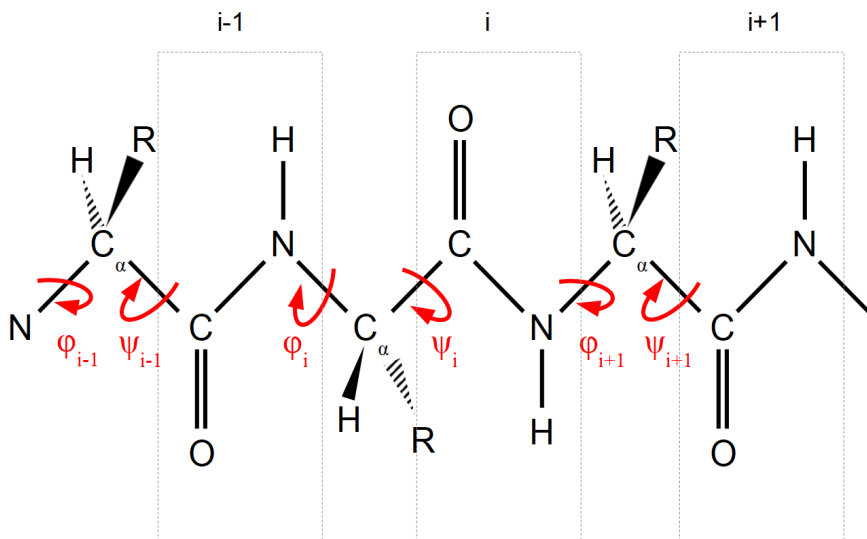


Figure 11. Dihedral angles (in red) which are involved in the effect that the local environment (amide groups  $i-1$  and  $i+1$ ) has on the intrinsic frequency of the vibration of amide group  $i$ . The amide groups are indicated using dashed boxes.

The influence of the nearest neighbors can be expressed by a modification of the diagonal elements of the  $F$  matrix; this corresponds to a shift of the intrinsic frequencies of the amide groups. Similarly to the nearest neighbor coupling, these shifts are obtained from quantum chemical calculations on small molecules such as dipeptides or tripeptides (to test the consistency of the shifts to longer chains).<sup>55,56,78,85,86</sup> From these calculations, two maps were created, where the shift is a function of the dihedral angles: one map is to retrieve the shift caused by the preceding amide group ( $\Delta\tilde{\nu}_{pre}$  in the following), the second map by the following amide group ( $\Delta\tilde{\nu}_{post}$  in the following). Each diagonal element in the  $F$  matrix can therefore be modified in this way:

$$\tilde{\nu} = \tilde{\nu}_0 + \Delta\tilde{\nu}_{pre}(\phi_i, \psi_i) + \Delta\tilde{\nu}_{post}(\phi_{i+1}, \psi_{i+1}) ,$$

where  $\tilde{\nu}$  is the wavenumber that will be converted to the diagonal element of the  $F$  matrix and  $\tilde{\nu}_0$  is the wavenumber calculated from the intrinsic frequency of vibration. If the amide group is at the C-terminus or the N-terminus of the chain, only one of the two shifts shall be considered.

Differently from the nearest neighbor coupling, Stock and co-workers noted that the calculation of the shifts is dependent on the computational details (level of theory and basis set),<sup>78</sup> so the implementation of this effect should be carefully evaluated.

The local environment effect was included in the calculations of Paper IV.

## 4.5 Hydrogen bond

As explained in Chapter 2, the hydrogen bond is one of the interactions that stabilize secondary and tertiary structures of proteins, since the amide hydrogen (N-H) and the carbonyl oxygen (C=O) of different amide groups bond to each other. IR spectroscopy is very sensitive to hydrogen bonding, both regarding the hydrogen bonds inside the protein of interest and regarding the hydrogen bonds between the protein and the solvent: a large redshift of the amide I wavenumbers was noted because of hydrogen bonding.<sup>36,61,87,88</sup> Each amide group can form at maximum three hydrogen bonds (two with the carbonyl oxygen and one with the amide hydrogen). Each hydrogen bond influences the intrinsic frequency of the involved amide groups in an approximately additive way:

$$\tilde{\nu} = \tilde{\nu}_0 + \Delta \tilde{\nu}_{CO_1}^H + \Delta \tilde{\nu}_{CO_2}^H + \Delta \tilde{\nu}_{NH}^H ,$$

where  $\tilde{\nu}$  is the wavenumber that will be converted to the diagonal element of the  $F$  matrix and  $\tilde{\nu}_0$  is the wavenumber calculated from the intrinsic frequency of vibration. The redshift was studied using clusters of NMA, in some cases surrounded by solvent (typically water molecules), in order to study intrapeptide hydrogen bond and solvation.<sup>61,87-93</sup> Typical values for the redshift are 20-25 cm<sup>-1</sup> for each hydrogen bond of the carbonyl oxygen and 10-15 cm<sup>-1</sup> for the hydrogen bond of the amide hydrogen.<sup>61,88</sup>

The shift is dependent on the strength of the hydrogen bond and on the geometrical arrangement of the involved amide groups. Attempts to model the shift due to the hydrogen bond to the geometrical properties were suggested in the literature. One of the formulas used to calculate the shift is the following.<sup>94,95</sup>

$$\Delta \tilde{\nu} = -\alpha (2.6 - r_{OH}) ,$$

where  $\Delta \tilde{\nu}$  is the shift of the wavenumber caused by the hydrogen bond in cm<sup>-1</sup>,  $r_{OH}$  is the hydrogen bond length in Å and  $\alpha$  is a fitting parameter, typically assumed to be 30 cm<sup>-1</sup>. This simple approach does not include the hydrogen bond strength and the hydrogen bond made to the amide hydrogen. But, it was used in simulations of the amide I band in the literature.<sup>96,97</sup>

Another approach is based on the Kabsch-Sander energy,<sup>98</sup> which calculates the energy of the hydrogen bond based on the distances between the involved atoms:

$$E_{KS} = f q_1 q_2 (r_{ON}^{-1} + r_{CH}^{-1} - r_{OH}^{-1} - r_{CN}^{-1}) .$$

In the formula,  $f$  is equal to 332 e<sup>-2</sup> Å kcal/mol,  $q_1$  is the partial charge on the C=O bond and is equal to 0.42 e,  $q_2$  is the partial charge on the N-H bond and is equal to 0.20 e and all distances are expressed in Å. Ge and coworkers suggested that the redshift is proportional to the Kabsch-Sander energy:<sup>99</sup>

$$\begin{aligned}\Delta \tilde{\nu}_{CO_1} &= \Delta \tilde{\nu}_{CO_2} = \xi^O E_{KS} \\ \Delta \tilde{\nu}_{NH} &= \xi^H E_{KS}\end{aligned},$$

where the coefficients  $\xi^O$  and  $\xi^H$  are respectively equal to 2.4 cm<sup>-1</sup>/kcal and 1.0 cm<sup>-1</sup>/kcal.

In addition to the hydrogen bond between amide groups, it is possible to calculate the solvation effect using the solvent accessible surface (SAS).<sup>100</sup> In the protein, each atom can be represented by a sphere with a radius equal to the Van Der Waal's radius. Using a water solvent probe sphere with a radius equal to 1.4 Å, we can define the accessible surface as the trace of the center of the probe sphere rolled over the protein surface. In our program, the SAS area is correlated to the redshift caused by the hydrogen bond between the solvent and the amide hydrogen or the carboxyl oxygen.<sup>58,101</sup> Similarly to the intrapeptide hydrogen bond, the redshift caused by solvation is approximately additive:

$$\tilde{\nu} = \tilde{\nu}_0 + \Delta \tilde{\nu}_{CO_1}^S + \Delta \tilde{\nu}_{CO_2}^S + \Delta \tilde{\nu}_{NH}^S,$$

where  $\tilde{\nu}$  is the wavenumber that will be converted to the diagonal element of the  $F$  matrix and  $\tilde{\nu}_0$  is the wavenumber calculated from the intrinsic frequency of vibration.

Intrapeptide hydrogen bond (using the method based on Kabsch-Sander energy) and solvation effect on the intrinsic frequency were included in the calculations of Paper IV.

## 4.6 From $F$ matrix to simulated spectrum

The combination of the interactions described in the previous sections fills the  $F$  matrix. The wavenumbers ( $\tilde{\nu}_k$ ) and the intensities ( $I_k$ ) of the amide I oscillators can be respectively found from the eigenvalues and the eigenvectors of the  $F$  matrix.<sup>2</sup> The equation for the intensity is:

$$I_k = \sum_{i=1}^3 \left( \frac{\partial \mu_i}{\partial Q_k} \right) = \sum_{i=1}^3 \left| \sum_{j=1}^N \left( \frac{\partial \mu_i}{\partial q_j} \right) \cdot \left( \frac{\partial q_j}{\partial Q_k} \right) \right|^2.$$

In this formula,  $\partial q_j / \partial Q_k$  is the vibrational amplitude of the amide group  $j$  in the  $k$ th normal mode (obtained as eigenvector of the  $F$  matrix),  $\partial \mu_i / \partial q_j$  is the  $i$ th Cartesian component of the dipole derivative of amide group  $j$  and  $N$  is the total number of amide groups.

The dipole strengths can be obtained from the intensities  $I_k$  and the wavenumbers  $\tilde{\nu}_k$  using the following formula:

$$D_k = \frac{I_k}{\tilde{\nu}_k} \cdot 4.1058^2 \cdot 10^4,$$

where  $D_k$  is the dipole strength of the normal mode  $k$  in  $D^2$ ,  $I_k$  is the intensity of the normal mode  $k$  in  $D^2 \text{ u}^{-1} \text{ \AA}^{-2}$ ,  $\tilde{\nu}_k$  is the wavenumber of the normal mode  $k$  in  $\text{cm}^{-1}$ , and  $10^4$  is a factor of conversion between  $\text{esu}^2 \text{ cm}^2$  and  $D^2$ .

Wavenumbers and intensities are used to simulate the spectrum. The band shape of each normal mode is approximated to a Gaussian or a Lorentzian curve, described with an appropriate full width at half maximum. Gaussian and Lorentzian bandshapes can be observed respectively when there is heterogeneous or homogeneous line broadening. Then, the IR spectrum is obtained by the sum of all generated curves.



## 5. Density functional theory and amide I band

Computational chemistry is a field of chemistry that solves chemical problems using the implementation of mathematical methods in computer simulations. As we know, molecules are composed of atoms and, in turn, of charged particles: negative electrons and positive nuclei. Briefly, the structure and the properties of molecule are obtained from the interactions between the charged particles through physical and chemical models. The aim of quantum chemistry, a subfield of computational chemistry, is to retrieve these properties through the postulates of quantum mechanics, such as the Schrödinger equation. Unfortunately, there are only a few limited situations where it is possible to solve the Schrödinger equation analytically and the electronic structure of the molecule. This is why several approaches were developed to overcome this problem, using different levels of approximation.

Quantum chemistry theory and several approaches are explained in textbooks.<sup>102,103</sup> A basic introduction of quantum chemistry calculations is presented in the following section. Then, I will focus on density functional theory (DFT), the method we chose to use for the calculations in Paper III. These calculations were performed using Gaussian 09.<sup>104</sup> From the DFT calculations, we collected the amide I wavenumbers and the dipole strengths; it was possible to retrieve the  $F$  matrices using the Hessian matrix reconstruction method,<sup>79</sup> explained in Section 5.3.

### 5.1 Introduction to quantum chemistry calculations

The simplest case to consider is a single charged particle (an electron, for example), which has a mass  $m$  and is moving through space and time. This particle has a potential energy  $V$  and a kinetic energy  $T$ , which can be expressed by:

$$H = T + V$$
$$T = \frac{\mathbf{p}^2}{2m} = -\frac{\hbar^2}{2m} \nabla^2 \quad ,$$

where  $\mathbf{p}$  is the momentum operator,  $\hbar$  is the reduced Planck constant and  $\nabla^2$  is the Laplace operator. The Hamiltonian operator  $H$  is the total energy operator. Time-dependent Schrödinger equation for a wavefunction is:

$$\mathbf{H} \Psi(\mathbf{r}, t) = i \hbar \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} ,$$

where  $\Psi$  is the wavefunction that describes the motion of the particle. The probability of observing the particle at a position  $\mathbf{r}$  and time  $t$  is given by the square of the wavefunction. If the potential energy is time-independent, the wavefunction can be written by the product of a spatial factor and a time factor:

$$\Psi(\mathbf{r}, t) = \Psi(\mathbf{r}) f(t) .$$

Then, the Schrödinger equation assumes a time-independent form:

$$\mathbf{H} \Psi(\mathbf{r}) = E \Psi(\mathbf{r}) ,$$

where  $E$  is the energy of the particle. The energy levels can be calculated from the wavefunctions using the following formula:

$$E = \frac{\int \Psi^* \mathbf{H} \Psi}{\int \Psi^* \Psi} .$$

The energy levels are then the eigenvalues of the  $\mathbf{H}$  operator. The Schrödinger equation can be solved analytically only in a few cases: a particle in a box, a harmonic oscillator (briefly mentioned in Section 3.2) and atoms with one electron (hydrogen or helium cation). We can consider, for example, the latter case. In addition to the kinetic energy of the electron, the potential energy is equal to the Coulomb interaction between the electron and the proton (protons, if it is the helium cation):

$$\mathbf{T} = \frac{\mathbf{p}^2}{2m} = -\frac{\hbar^2}{2m} \nabla^2 \quad \mathbf{V} = -\frac{Z e^2}{4\pi \epsilon_0 \mathbf{r}} ,$$

where  $Z$  is the nuclear charge,  $e$  is the elementary charge and  $\epsilon_0$  is the dielectric constant. Using polar coordinates instead of Cartesian coordinates, the solution of Schrödinger equation is provided by the product of a radial function and an angular function:

$$\Psi_{nlm} = R_{nl}(r) Y_{lm}(\theta, \phi) .$$

The wavefunctions are also called *orbitals* and are characterized by three quantum numbers  $n$ ,  $l$  and  $m$ :  $n$  is the principal quantum number and assumes integer values (0, 1, 2...);  $l$  is the azimuthal quantum number and assumes the integer values between 0 and  $n-1$  (0, 1, ...  $n-1$ );  $m$  is the magnetic quantum number and assumes the integer values between  $-l$  and  $l$  ( $-l$ ,  $-l+1$ , ... 0, ...  $l-1$ ,  $l$ ).  $R_{nl}$  is the radial function and is the product of a factor and a Laguerre polynomial.  $Y_{lm}$  is the angular function and is the product of a factor and a Legendre polynomial. The energy of each solution is only a function of the principal quantum number; orbitals with the same  $n$  and different  $m$  or  $l$  are considered degenerate.

When we are facing a many-body system, i.e. when there are more than one electrons, the Schrödinger equation cannot be solved analytically. Solutions which are approximations of the real solution can be suggested. In addition,

the electron spin has to be considered in a system with more than one electron. One of the approximations to include in the theoretical description is the Born-Oppenheimer approximation: the masses of the nuclei are much greater than the masses of the electrons, so the electrons can adjust almost immediately to any position changes of the nuclei. This allows to consider the total wavefunction as a product of a wavefunction for the electrons and a wavefunction for the nuclei. Consequently, the total energy will be the sum of the nuclear energy and the electron energy. Another simple approximation is to ignore the interaction between electrons in the Hamiltonian. Then, the total Hamiltonian becomes the sum of the Hamiltonian of each electron and the wavefunction becomes a linear combination of the wavefunctions of each electron, to respect the indistinguishability (antisymmetry) criterion. Molecular orbitals can be expressed in terms of known functions from a set, called *basis set*. Each basis function is called *atomic orbital* and the molecular orbital (one-electron wavefunction) is a linear combination of atomic orbitals:

$$\psi_i = \sum_{\mu=1}^K c_{\mu i} \phi_{\mu} \quad ,$$

where  $\psi_i$  is a molecular orbital,  $c_{\mu i}$  is a coefficient,  $\phi_{\mu}$  is one of the  $K$  atomic orbitals. Any type of basis function can be accepted, as long as it respects physics (convergence if many basis functions are included; basis function decreases to zero if the distance between nucleus and electron increases) and it can be easily integrated. Basis sets can be distinguished by the type and by the number of basis functions. A higher number of basis functions makes the calculations more accurate, but it requires a higher computational cost. Each atomic orbital is a linear combination of functions, typically Slater type or Gaussian type. In the *minimal basis set*, each atomic orbital can be described by a single Gaussian or Slater function. However, calculations with minimal basis sets usually contain deficiencies; so, the number of basis function can be increased. Polarization functions (indicated by the symbol \*) or diffuse functions (indicated by the symbol +) can be also added to the basis set.

## 5.2 Density functional theory

Different methods were implemented to solve the electronic structure and the Schrödinger equation for many-body systems. For example, *ab initio* calculations are based solely on the laws of quantum mechanics, or semi-empirical methods avoid the calculations of certain integrals using experimental data. Force fields were also created to convert electronic energy into parametric functions. Another method is suggested by density functional theory (DFT) calculations.

The first description of DFT was introduced by Hohenberg and Kohn<sup>105</sup> and perfected by Kohn and Sham in the 1960s;<sup>106</sup> however, this method started to be widely used from the 1980s.<sup>107,108</sup>

DFT calculations are based on the assumption that the total electronic energy is related to the overall electronic density, i.e. the energy is a functional of the density. The density functional in DFT is usually written as:

$$E[\rho(\mathbf{r})] = \int V_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} + F[\rho(\mathbf{r})] \quad .$$

The first term in the integral is related to the interaction of the electrons with an external potential  $V_{\text{ext}}$  (typically derived from the Coulomb interaction between the nuclei); the second term is the sum of the kinetic energy of the electrons and the contribution from the interactions between the electrons.

Kohn and Sham suggested that this second term  $F[\rho(\mathbf{r})]$  can be approximated by three terms:

$$F[\rho(\mathbf{r})] = E_{KE}[\rho(\mathbf{r})] + E_H[\rho(\mathbf{r})] + E_{XC}[\rho(\mathbf{r})] \quad .$$

The first term  $E_{KE}$  is the kinetic energy and it is defined as the kinetic energy of a system of non-interacting electrons with the same density  $\rho(\mathbf{r})$ . The second term  $E_H$  is the electron-electron Coulombic energy and it is defined as the classical electrostatic interaction between two charged densities. The third term  $E_{XC}$  contains contributions from exchange and correlation. Substituting the correct formulas for these energy terms, we obtain the expression of the total energy for a system with N electrons and M nuclei:

$$\begin{aligned} E[\rho(\mathbf{r})] = & - \sum_{A=1}^M \int \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \rho(\mathbf{r}) d\mathbf{r} + \sum_{i=1}^N \psi_i(\mathbf{r}) \left( -\frac{\nabla^2}{2} \right) \psi_i(\mathbf{r}) d\mathbf{r} \\ & + \frac{1}{2} \iint \frac{\rho(\mathbf{r}_1) \rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 + E_{XC}[\rho(\mathbf{r})] \quad . \end{aligned}$$

In the description of Kohn and Sham, the density  $\rho(\mathbf{r})$  is the sum of square moduli of a set of one-electron orbitals  $\psi_i$ :

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2 \quad .$$

Combining the definition of the density and the formula of the total energy, we obtain the one-electron Kohn-Sham equation:

$$\left\{ -\frac{\nabla^2}{2} - \left( \sum_{A=1}^M \frac{Z_A}{r_{1A}} \right) + \int \rho \frac{(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + V_{XC}[\mathbf{r}_1] \right\} \psi_i(\mathbf{r}_1) = \epsilon_i \psi_i(\mathbf{r}_1) \quad ,$$

where  $\epsilon_i$  are the orbital energies and  $V_{XC}$  is the exchange-correlation functional, defined as:

$$V_{XC}[\mathbf{r}] = \frac{\delta E_{XC}[\rho(\mathbf{r})]}{\delta [\rho(\mathbf{r})]} \quad .$$

Kohn-Sham equations are solved through a self-consistent approach: an initial guess of the density leads to the calculation of a set of orbitals, which in

turn leads to a new and improved value for the density. The iteration continues until the convergence is reached.

Many types of empirical or semi-empirical functionals for the exchange-correlation energy were suggested in the literature.<sup>109</sup> We decided to choose BPW91<sup>110–113</sup> as density functional and the 6-31G\*\* as basis set for the geometry optimizations and the frequency calculations in Paper III. This combination was chosen because it is a good compromise between accuracy of the vibrational frequencies and computational cost.<sup>91,114–116</sup> In addition, the amide I and amide II dipole strengths for NMA calculated with this combination are close to the experimental values;<sup>91</sup> it also reproduces well the vibrational coupling in the description of the effects of site-specific <sup>13</sup>C labeling.<sup>117</sup> Another popular functional is B3LYP;<sup>118</sup> however, it seems to perform worse because it gives higher amide I frequencies.<sup>114,116</sup>

### 5.3 Hessian matrix reconstruction

Amongst several properties, a frequency calculation performed on an optimized geometry gives amide I wavenumbers and dipole strengths. Cho and coworkers suggested the Hessian reconstruction method to retrieve the Hessian matrix (and, as a consequence, the  $F$  matrix) from the DFT amide I wavenumbers.<sup>79</sup> The reconstructed Hessian  $H$  can be retrieved using the following formula:

$$H = U^{-1} \Lambda U \quad ,$$

where  $\Lambda$  is the wavenumber eigenvalue matrix obtained from the DFT calculations.  $U$  is a matrix that can be obtained from two different methods: carbonyl population analysis or carbonyl coordinate displacement.

The first method approximates the contribution of each amide group to a particular amide I normal mode by the carbon and oxygen displacements from the DFT results. The elements of the  $U$  matrix are obtained from the following formula:

$$U_{\alpha j}^2 = N_j (u_{C_{\alpha x} j}^2 + u_{C_{\alpha y} j}^2 + u_{C_{\alpha z} j}^2 + u_{O_{\alpha x} j}^2 + u_{O_{\alpha y} j}^2 + u_{O_{\alpha z} j}^2) \quad ,$$

where  $u_{C_{\alpha x} j}$  is the displacement of the carbonyl carbon atom of amide group  $\alpha$  along the  $x$  direction for the  $j$ -th normal mode and  $u_{O_{\alpha x} j}$  is the displacement of the carbonyl oxygen atom of the amide group  $\alpha$  along the  $x$  direction for the  $j$ -th normal mode. The normalization constant  $N_j$  is calculated according to the formula:

$$\sum_{\alpha=1}^N U_{\alpha j}^2 = 1 \quad ,$$

where  $N$  is the number of amide groups. The sign of the elements of the  $U$  matrix is assigned according to whether the carbonyl bond length increases or decreases according to the displacements.

The second method approximates the contribution of each amide group to a particular amide I normal mode by the carbonyl bond length change. The elements of the  $U$  matrix are obtained from the following formula:

$$U_{\alpha j}^2 = N_j (r_{\alpha j} - r_{\alpha j}^0) \quad ,$$

where  $r_{\alpha j}$  is the carbonyl bond length of amide group  $\alpha$  after evaluating the atom displacements for the  $j$ -th normal mode and  $r_{\alpha j}^0$  is the original carbonyl bond length of amide group  $\alpha$  in the optimized geometry. The normalization constant  $N_j$  is calculated as in the carbonyl population analysis; however, the sign of the elements of the  $U$  matrix is already determined.

Carbonyl population analysis and carbonyl coordinate displacement produce similar Hessian matrices. The reconstructed Hessian matrix is almost diagonally symmetric and the eigenvectors are fairly orthogonal to one another.<sup>74</sup> In Paper III, we used the carbonyl population analysis to retrieve the Hessian matrices. Then, these matrices were converted to  $F$  matrices through the transformation of the elements from wavenumbers to mass normalized force constants (formulas have already been reported in Section 4.1). The  $F$  matrices were then made symmetrical, by averaging the elements  $f_{ij}$  and  $f_{ji}$ .

## 6. Results and discussion

### 6.1 Paper I

The aim of the Paper was to study the co-aggregation of two species of A $\beta$  peptides, A $\beta_{40}$  and A $\beta_{42}$ . To do so, we used FTIR spectroscopy, combined with  $^{13}\text{C}$  labeling and simulations.

In the experiments, labeled A $\beta_{40}$  and unlabeled A $\beta_{42}$  (and vice versa) were co-incubated, then the IR spectrum was recorded (Paper I, Fig. 1). The resulting mixture spectrum shows the typical features of the antiparallel  $\beta$ -sheet: a strong band at low wavenumber (around  $1625\text{ cm}^{-1}$ ) and a weak band at high wavenumber (around  $1680\text{ cm}^{-1}$ ). The experimental bands correspond to the bands in the spectra of the two individual species. However, in the mixture spectrum, the main  $^{12}\text{C}$  band and  $^{13}\text{C}$  band positions show a clear upshift, when the isotopic dilution increases (Paper I, Fig. 2). The  $^{12}\text{C}$  band also shows a loss in intensity. These features are clearly evidence that the two species interact with each other and form  $\beta$ -sheets with mixed isotopic compositions. The behavior of the mixture spectrum and of the  $^{13}\text{C}$  band position is similar for homo-oligomers and hetero-oligomers. Therefore, we suggest that the two species mixed randomly in the same structures. To verify if A $\beta$  can form mixed structures with other peptides, we performed a control experiment, where we mixed one of the two A $\beta$  species with S100A9 (Paper I, Fig. 6). The latter protein is known to form amyloids and to interact with A $\beta$ ,<sup>119,120</sup> its IR spectrum is different from A $\beta$ , because the main band is located at  $1650\text{ cm}^{-1}$  (due to  $\alpha$ -helices). The resulting mixture spectrum is almost superimposable with the average spectrum of the individual species, so we can deduce that the two species do not form mixed  $\beta$ -sheets.

To add additional evidence, we performed a simulation on an antiparallel  $\beta$ -sheet model using TDC and nearest neighbor coupling (Paper I, Fig. 3). We simulated the presence of  $^{13}\text{C}$  peptides in the model structure by the modification of the elements of the  $F$  matrix using an appropriate mass ratio. For each experimental isotopic dilution, we considered 3000  $\beta$ -sheets with different isotopic compositions; to calculate errors in the band positions, we repeated the calculations 20 times. The simulated spectra show the same be-

havior as the experiments, so we can confirm our hypothesis of mixed  $\beta$ -sheets. Since labeled and unlabeled species had different carbon isotopes, preference for homogeneous aggregates was modeled in the calculations with the introduction of a parameter called *acceptance probability* (Paper I, Fig. S8). This probability represents the tendency of the peptides to stay closer to a peptide of the own kind: 0% probability means isotopically pure sheet, 100% means a complete random mixing of the two types of peptide. After a comparison of the simulated and experimental  $^{13}\text{C}$  band positions and after an analysis of the isotopic compositions of the simulated  $\beta$ -sheets, the simulations can confirm a behavior close to the random mixing of the peptides (acceptance probability higher than 40%).

## 6.2 Paper II

The aim of Paper II was to elucidate the building block of the  $\text{A}\beta$  oligomers by making a comparison of  $^{13}\text{C}$  band shift upon isotopic enrichment between experimental spectra of homo-oligomers and simulations.

In this set of simulations, neither the structure of the oligomers nor the parameters to use in the calculations were known. To overcome the first problem, we performed the calculations on a collection of different model structures: antiparallel  $\beta$ -sheets of different sizes, antiparallel  $\beta$ -sheets of different sizes and a dodecamer (Paper II, Fig. 1). To solve the second problem, we decided to adjust the magnitude of the transition dipole moment and the center wavenumber for each model structure in order to match the experimental high and low wavenumber band positions for the all-unlabeled sample. These adjustments were performed for  $\text{A}\beta_{40}$  homo-oligomers and  $\text{A}\beta_{42}$  homo-oligomers. The position of the transition dipole moment was kept fixed to Moore and Krimm's suggested position,<sup>76</sup> while the magnitude and the center wavenumber were obtained for three fixed TDM angles ( $10^\circ$ ,  $20^\circ$  and  $30^\circ$ ).

The labeling procedure of the model structures was performed using five different patterns: single strand, 2-strand block, 3-strand block, interlaced model 1 and interlaced model 2 (Paper II, Fig. 3). These patterns are related to hypothetical structures of individual peptide molecules, i.e. to the building block of the oligomers. We performed 20 repetitions of 3000 model structures with different isotopic composition for each experimental isotopic dilution and for each labeling pattern. The calculation of the  $F$  matrices was done according to TDC and nearest neighbor coupling. According to the isotopic composition, the elements of the  $F$  matrix were modified using an appropriate mass ratio (as done in Paper I).

The  $^{13}\text{C}$  band shifts in the simulations showed that the  $^{13}\text{C}$  band positions depends on the building block and that the downshift for each building block does not overlap with other downshifts (Paper II, Fig. 8). Similar  $^{13}\text{C}$  band

shifts were found for the three fixed angles of the TDM, indicating that our results do not depend on the choice of the angle (Paper II, Fig. 7). The experimental  $^{13}\text{C}$  band shifts for  $\text{A}\beta_{40}$  homo-oligomers are similar to those of the 2-strand block, so we can suggest that each peptide contributes 2 strands to the structure of the  $\text{A}\beta_{40}$  homo-oligomers. Instead, the experimental  $^{13}\text{C}$  band shifts for  $\text{A}\beta_{42}$  homo-oligomers are located close to the 3-strand block (Paper II, Fig. 9). However, we cannot exclude the 2-strand block because, when the model structure has a small number of strands, the  $^{13}\text{C}$  band shifts for the 2-strand block become smaller and closer to the experimental values. Therefore we suggest that each peptide contributes at least 2 strands to the structure of the  $\text{A}\beta_{42}$  homo-oligomers.

Different infrared studies in the literature provided additional conclusions on the  $^{13}\text{C}$  band positions. Matos et al.<sup>121</sup> studied aggregates of labeled pyroglutamylation  $\text{A}\beta_{3-42}$  with  $\text{A}\beta_{42}$ . Comparing the  $^{13}\text{C}$  band shifts at 0.5 molar ratio, we noted that their aggregates are different from our homo-oligomers and hetero-oligomers. Their  $^{13}\text{C}$  band shift seems to suggest a single strand as a building block of the  $\text{A}\beta_{3-42}$ - $\text{A}\beta_{42}$  oligomers. Moran et al.<sup>122</sup> calculated a larger  $^{13}\text{C}$  band shift than ours for amyloid fibrils of  $\gamma\text{D}$ -crystallin. The reason of this discrepancy can be found in their lower band positions, which revealed a stronger interstrand coupling. Buchanan et al.,<sup>20</sup> using an experimental procedure similar to the one performed in our Paper, found a  $^{13}\text{C}$  band shift close to our 2-strand block model.

Our 2-strand block can be seen as a  $\beta$ -hairpin; the series of multiple  $\beta$ -hairpins is in agreement with the antiparallel structure of the oligomers from the literature. Comparing the main band position of small antiparallel structures ( $\beta$ -hairpin or 3-stranded sheet),<sup>123-125</sup> with our experimental values we can deduce that the possible structure of the oligomers is the antiparallel  $\beta$ -sheet or  $\beta$ -barrel with at least 4 strands. Our conclusions on the structure and on the 2-strand block fit well with recent studies and suggested models.<sup>28,126-128</sup> Other models in literature that suggested the single strand as building block<sup>129,130</sup> are not in agreement with the isotope effect seen in Papers I and II.

## 6.3 Paper III

As anticipated in Section 4.2 of this thesis, there is still no agreement on the optimal set of parameters for the TDM to describe TDC. Therefore, the aim of Paper III is the optimization of the parameters used for the calculation of the transition dipole coupling. In order to find the optimal set of parameters, we decided to minimize the difference between the coupling constants from TDC calculations and from DFT calculations.

Our set of model structures was composed by 2  $\alpha$ -helices, 2 antiparallel  $\beta$ -sheets and 2 parallel  $\beta$ -sheets. Using the software Gaussian 09,<sup>104</sup> we performed a geometry optimization and a frequency calculation for the model structures. The BPW91<sup>111–113</sup> density functional and the 6-31G\*\* basis set were used for both types of calculations. The Hessian matrices were retrieved from the frequency calculation results using the carbonyl population analysis as Hessian reconstruction method. The Hessian matrices were then converted to  $F$  matrices and then made symmetrical.

TDC calculations used coupling constants from the diagonal and the nearest neighbor diagonals of the DFT  $F$  matrices; in this way, the TDC  $F$  matrices only depended on TDC and not on other interactions. The optimized parameters were the position of the TDM, the magnitude of the dipole derivative and the  $A$  parameter, which is a parameter we introduced to describe the proportionality between the hydrogen bond and the magnitude of the dipole derivative. We also tested two different positions for the TDM, one suggested by Moore and Krimm<sup>76</sup> and one suggested by Chirgadze and Nevskaya.<sup>42</sup>

We performed two procedures of optimization. In the first one, we considered all the model structures together. We minimized the squared difference between the DFT and the TDC  $F$  matrices (R value in Paper III, Tab. 1) for each model structure. Because of the smaller number of coupling constants, the squared difference for the three smaller structures was doubled. The optimal parameters are shown in Paper III, Tab. 1; the best choice for the TDM position is our optimized position or the position suggested by Chirgadze and Nevskaya.<sup>42</sup> A good agreement is reached for the wavenumbers of the normal modes between TDC and DFT calculations, but the strongest dipole strengths are underestimated in TDC calculations for the largest model structures (Paper III, Fig. 1). To verify that the underestimation was not a deficiency of the TDC, we performed a calculation using DFT  $F$  matrices (Paper III, Fig. 2) and the optimal TDM parameters for intensity calculations. The behavior was the same, so we suggest that the calculations of the TDC  $F$  matrix and the dipole strengths could be done with different sets of parameters. A possible explanation for the underestimation of the strongest dipole strengths can be an imperfect implementation of the hydrogen bonding effect. It is also possible that our optimization program focused on the strongest interactions, such as the interactions between close neighbors, where it is known that TDC is not completely reliable.<sup>55,56,72,73,78,79</sup>

Our optimized magnitude of the dipole derivative is smaller than the magnitudes calculated in literature,<sup>90,131–133</sup> whereas the optimized angle is in agreement with several studies.<sup>42,90,131,134,135</sup> The effect of the hydrogen bond on the magnitude of the dipole derivative was already revealed in literature.<sup>90,131,136</sup>

In the second procedure of optimization, we minimized the squared difference between the DFT and TDC  $F$  matrices (R value in Paper III, Tab. 2) for each secondary structure separately. Similarly to the first optimization proce-

ture, the squared difference for the smaller structures was doubled. This optimization procedure was performed fixing the optimized position of the first procedure. We found then the optimal set of parameters for each secondary structure. The magnitude of the TDM and the  $A$  parameter were similar to the values optimized in the first procedure, but the angle of the TDM was different. The optimized angle of the TDM was also found to be different for each secondary structure.

Using the optimal set of parameters for the calculation of TDC for each secondary structure, we successfully obtained a good agreement of the coupling constants of the DFT  $F$  matrices, but a perfect matching was still not reached. Our optimized angles for the secondary structures are in agreement with the ranges of the angles suggested in literature,<sup>66,75,131</sup> but they could not be the real optimal parameters because they could have compensated the deficiencies of the TDC model.

The accuracy of TDC is still debated and studied in literature.<sup>55,56,74,79,97,131,137,138</sup> A particular choice of the set of parameters can lead to underestimation or overestimation of specific interactions between the amide groups. However, despite of being an approximation which does not include multipoles and mechanical coupling, TDC can describe the amide I spectrum well. It is known that different secondary structures can be described more correctly by specific sets of parameters.<sup>84</sup> However, our suggestion is to use our optimized set of parameters from the complete set of model structures using the position suggested by this Paper or by Chirgadze and Nevskaya.<sup>42</sup> Even if the matching of the coupling constants is not perfect, they can reproduce correctly the spectral shapes of the DFT IR spectra.

## 6.4 Paper IV

The aim of Paper IV was to study the amide I spectrum of parallel  $\beta$ -sheet proteins by IR experiments and simulations. Typically, a small high wavenumber band is considered a signature band for antiparallel  $\beta$ -sheets (1680-1690  $\text{cm}^{-1}$ ) and it is not present in IR spectra of parallel  $\beta$ -sheets. This conclusion was proven for experiments on proteins with antiparallel  $\beta$ -sheets,<sup>49,139</sup> through calculations on antiparallel and parallel  $\beta$ -sheets<sup>42,44,65,140</sup> and experiments with site-specific labeling<sup>64,141</sup>. However, this high wavenumber band is a matter of controversy. A review showed that a high wavenumber band was present in the spectrum of A $\beta$  fibrils, suggesting that fibrils have an antiparallel  $\beta$ -sheet arrangement;<sup>24</sup> this conclusion was lately confuted, demonstrating the parallel  $\beta$ -sheet arrangement.<sup>6,18,142</sup> Furthermore, other work studied some parallel  $\beta$ -sheets proteins that showed a high wavenumber band in IR spectra.<sup>143,144</sup>

We decided to study two  $\beta$ -helix proteins (SV2 and Pent) using IR experiments and simulations. The simulations included the local environment effect<sup>78</sup> and the hydrogen bonding effect (both intrapeptide and between protein and solvent through the SAS method);<sup>58,100</sup> the coupling constants were obtained from DFT calculations for nearest neighbor interactions<sup>78</sup> or calculated from TDC. The contribution of each amide group was then calculated and collected according to the secondary structure of the residue containing the carbonyl group.

The experimental and the simulated second derivatives of the IR absorbance for SV2 and Pent are shown in Paper IV, Figs 2 and 3. For both of these proteins, the simulated spectra present a good agreement on the band positions with the experimental spectra. The component band at high wavenumbers ( $1690\text{ cm}^{-1}$ ) is clearly visible in both experimental and simulated second derivatives of SV2 and Pent. The intensities of experimental and simulated second derivatives are matched in the central region of the spectrum; in the other spectral ranges of the amide I band (such as the range where is located the high wavenumber band), simulations show higher intensities than experiments.

The calculated residue contribution reveals the spectral range in which the secondary structure categories contribute (Paper IV, Figs 4 and 5). The contributions in the region between  $1695$  and  $1685\text{ cm}^{-1}$ , where the high wavenumber band is located, reveal that the contribution of the  $\beta$ -sheets is limited; all the other secondary structures contribute largely, in particular the two types of bend in the analysis of SV2.

Therefore, our simulations correctly reproduced the experimental band positions in the second derivatives. Both the experimental and simulated second derivatives confirmed the presence of the high wavenumber band. However, after the residue contribution analysis, we can exclude the contribution of parallel  $\beta$ -sheets to the high wavenumber band. The secondary structures contribution to the amide I band are in accordance with previous studies,<sup>1,37,45,46</sup> in particular for  $\beta$ -sheets, bends and amide groups that cannot be considered as  $\beta$ -sheets or  $\alpha$ -helix.

We can conclude that the high wavenumber band is a hallmark of the antiparallel  $\beta$ -sheets, but other secondary structures may contribute in the same spectral range, such as in the case of SV2 and Pent. These proteins and other  $\beta$ -helix proteins show other prominent bands outside the region of  $\beta$ -sheet absorption. The presence of only two distinct bands, one in the main  $\beta$ -sheet region (lower than  $1640\text{ cm}^{-1}$ ) and one above ( $1680$ - $1690\text{ cm}^{-1}$ ), is an evidence for the presence of antiparallel  $\beta$ -sheets.

## 7. Conclusions and outlook

The listed papers presented in this thesis showed the benefits of using simulations in different aspects of IR spectroscopy. As summarized in Chapter 6, our approach revealed a good agreement with experimental and with DFT results. Even if in Paper I, II and III the calculations were performed on ideal model structures instead of real protein structures, the conclusions based on the simulations results were remarkable. However, even if the normal mode analysis explained in this thesis provided important results, it is characterized by a few deficiencies.

As anticipated in Sections 4.2 and 6.3, TDC is an electrostatic interaction that does not include higher multipoles and through-bond coupling. A more complete description of the electrostatic interaction is provided by the transition charge model,<sup>70</sup> which assigns a point charge flow to each atom in the amide group and calculates the interaction between the vibrating electric field (generated by the vibration of the amide group) and the other amide groups. This model has already been used in literature, but the expected improvement is not always reached.<sup>55,66,70,79,131,145</sup> Through-bond coupling is still not included in the description. However, the small contribution of the C-N stretching vibration to the amide I band, which is the main reason for the mechanical coupling, makes TDC a good approximation of the interaction between the amide groups.

Another possible improvement can be pointed out in the calculation of the effect of hydrogen bond. Recent studies performed by Torii suggested that the shift is not correlated with the strength of the hydrogen bond, but with geometrical properties.<sup>146,147</sup> Recently, we have been testing an implementation of a new method<sup>147</sup> in our program: partial charges were assigned to each atom in the amide group and the shift for each amide group is function of the electric field generated by the atoms in the amide group and of the electrostatic potential of the environment.

Furthermore, we have to consider that our calculations do not include dynamics, but they consider fixed, momentary geometries. Proteins, during dynamics, may have fluctuations and conformational changes in the structure that can affect the  $F$  matrix. A solution was implemented and used in Paper II: statistical variations were added to the elements of the  $F$  matrix.<sup>82,122,148</sup> The diagonal elements were subjected to a random variation between -1% and 1% of their original value and the off-diagonal elements be-

tween -10% and 10%. Another solution would be to make use of molecular dynamics calculations and to analyze the conformational dynamics, averaging spectra of snapshot structure.<sup>20,84,149</sup>

# Acknowledgements

I still remember the day of departure from Italy almost five years ago: this PhD was my first working/studying experience abroad. I was a little scared, since I would have been alone and far from home for a lot of time, without knowing anyone; but, at the same time, I was curious to see a new country, taking advantage of this huge opportunity and trying to improve my abilities and my knowledge in academia.

Then, first of all, I want to thank **Prof. Andreas Barth**, for believing in me and giving me this chance after a quick interview via Skype, without any encounter in person. You taught me a lot about IR spectroscopy and we discussed a lot about several other topics, not only related to science. Under your guidance, I think to have been a good coworker and to have become a better scientist (I hope so!).

**Prof. Aatto Laaksonen**, my assistant supervisor, thank you for your assistance during my PhD experience, in particular for your lessons and your help to perform the DFT calculations.

A general thank you to all the **members of the DBB**, from professors, post-docs, PhD students to the administration staff and everyone else. I think that our department has a nice working and social environment during everyday life and I feel happy to have been part of it.

Amongst all of them, a huge gratitude to the (past and present) members of the AB group. At first, **Maurizio** and **Chenge** welcomed me in the lab office for my first months; then, **Faraz**, **Evelyne** and **Suman** joined our group. Thank you for your company during working time and, in particular, during lunch time and dinners. I do not mention all the students that came during the last five years (I do not want to write a long list), but I enjoyed to meet and know each one of them.

I also want to thank all my **friends**, both the ones I made in Sweden and the ones I have (and have made) in Italy. I started my experience alone here in Sweden, so even a small message made me happy and distracted me from my work. Dinners, beers, movies and holidays together were even better!

Last, but not least, I want to thank my family: my mother **Giacinta** and my sisters **Roberta** (with **Kasper**) and **Silvia**. As the sign in our kitchen says, family is like branches on a tree: we all grow in different directions, yet our roots remain as one.



## Sammanfattning (Summary in Swedish)

Infrarödspektroskopi är en viktig teknik för att få strukturinformation genom analys av absorptionsspektra. Den huvudsakliga användningen av infrarödspektroskopi inom livsvetenskapen är studien av amid I-bandet som är korrelerat med konformationen hos proteinryggraden och således med proteinernas sekundärstruktur. Men bandtolkning och spektrumtydning är inte enkelt. På grund av detta utvecklades flera simuleringsmetoder för att hjälpa tolkningen av experimentella amid I-spektra. Ett exempel på en metod använd i denna avhandling är normalkoordinatanalysen som är baserad på en utvärdering av den inneboende amid I-vibrationen hos amidgrupper och av interaktioner mellan dessa. Beräkningen tar hänsyn till flera effekter: övergångsdipolmomentkoppling, närmaste grannars växelverkan, effekten av den lokala omgivningen och effekten av vätebindning. Genom normalmodanalysen är det möjligt att få det simulerade infrarödspektrumet och de enskilda amidgruppernas bidrag till ett visst spektralområde.

Syftet med den här avhandlingen och med de inkluderade publikationerna är att förklara metoden, förbättra den och att visa på dess potential. Simuleringsresultat jämfördes med experimentella data för olika proteiner av intresse: amyloid- $\beta$  oligomerer och  $\beta$ -helix proteiner. Simulerade och experimentella infrarödspektra visade liknande absorptionsband. Simuleringarna ledde också till ytterligare slutsatser: de bekräftade en slumpmässig blandning av amyloid- $\beta$  peptider i oligomerer; de antydde att amyloid- $\beta$  peptider bidrar med minst två strängar till oligomerstrukturen; de avslöjade att bandet vid höga vågtal, som är typisk för antiparallela  $\beta$ -flak, också kan orsakas av andra sekundärstrukturer, men inte av parallela  $\beta$ -flak. Dessutom, för att verifiera och förbättra metodens noggrannhet, jämfördes simuleringsresultat också direkt med resultat från täthetsfunktionalteori. Denna jämförelse ledde till ett förslag till en ny parameteruppsättning för simuleringarna.



# Bibliography

- 1 H. Torii and M. Tasumi, *J. Chem. Phys.*, 1992, **96**, 3379–3387.
- 2 H. Torii and M. Tasumi, in *Infrared spectroscopy of biomolecules*, eds. H. H. Mantsch and D. Chapman, Wiley-Liss, New York, 1996, pp. 1–18.
- 3 J. M. Berg, J. L. Tymoczko and L. Stryer, *Biochemistry*, New York: W.H. Freeman., 2004.
- 4 K. E. Van Holde, W. C. Johnson and P. S. Ho, *Principles of Physical Biochemistry*, Upper Saddle River, NJ, 2nd ed., 2006.
- 5 D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, New York, 5th ed., 1921.
- 6 F. Chiti and C. M. Dobson, *Annu. Rev. Biochem.*, 2017, **86**, 27–68.
- 7 E. N. Cline, M. A. Bicca, K. L. Viola and W. L. Klein, *J. Alzheimer's Dis.*, 2018, **64**, S567–S610.
- 8 S. Cowan, R. Garavito, J. Jansonius, J. Jenkins, R. Karlsson, N. König, E. Pai, R. Pauptit, P. Rizkallah, J. Rosenbusch, G. Rummel and T. Schirmer, *Structure*, 1995, **3**, 1041–1050.
- 9 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
- 10 RCSB Protein Data Bank (PDB), <http://www.rcsb.org>.
- 11 C. Haass and D. J. Selkoe, *Nat. Rev. Mol. Cell Biol.*, 2007, **8**, 101–112.
- 12 K. L. Viola and W. L. Klein, *Acta Neuropathol.*, 2015, **129**, 183–206.

- 13 A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio and R. Tycko, *Proc. Natl. Acad. Sci.*, 2002, **99**, 16742–16747.
- 14 D. R. Thal, J. Walter, T. C. Saido and M. Fändrich, *Acta Neuropathol.*, 2015, **129**, 167–182.
- 15 S. Chimon, M. A. Shaibat, C. R. Jones, D. C. Calero, B. Aizezi and Y. Ishii, *Nat. Struct. Mol. Biol.*, 2007, **14**, 1157–1164.
- 16 I. Bertini, L. Gonnelli, C. Luchinat, J. Mao and A. Nesi, *J. Am. Chem. Soc.*, 2011, **133**, 16013–16022.
- 17 I. Kuperstein, K. Broersen, I. Benilova, J. Rozenski, W. Jonckheere, M. Debulpaep, A. Vandersteen, I. Segers-Nolten, K. Van Der Werf, V. Subramaniam, D. Braeken, G. Callewaert, C. Bartic, R. D’Hooge, I. C. Martins, F. Rousseau, J. Schymkowitz and B. De Strooper, *EMBO J.*, 2010, **29**, 3408–3420.
- 18 R. Aleksis, F. Oleskovs, K. Jaudzems, J. Pahnke and H. Biverstål, *Biochimie*, 2017, **140**, 176–192.
- 19 S. D. Moran and M. T. Zanni, *J. Phys. Chem. Lett.*, 2014, **5**, 1984–1993.
- 20 L. E. Buchanan, J. K. Carr, A. M. Fluitt, A. J. Hoganson, S. D. Moran, J. J. de Pablo, J. L. Skinner and M. T. Zanni, *Proc. Natl. Acad. Sci.*, 2014, **111**, 5796–5801.
- 21 V. A. Streltsov, J. N. Varghese, C. L. Masters and S. D. Nuttall, *J. Neurosci.*, 2011, **31**, 1419–1426.
- 22 M. Schmidt, C. Sachse, W. Richter, C. Xu, M. Fändrich and N. Grigorieff, *Proc. Natl. Acad. Sci.*, 2009, **106**, 19813–19818.
- 23 S. Wärmländer, A. Tiiman, A. Abelein, J. Luo, J. Jarvet, K. L. Söderberg, J. Danielsson and A. Gräslund, *ChemBioChem*, 2013, **14**, 1692–1704.
- 24 R. Sarroukh, E. Goormaghtigh, J.-M. Ruyschaert and V. Raussens, *Biochim. Biophys. Acta - Biomembr.*, 2013, **1828**, 2328–2338.

- 25 T. Lührs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Döbeli, D. Schubert and R. Riek, *Proc. Natl. Acad. Sci.*, 2005, **102**, 17342–17347.
- 26 W. L. Klein, W. B. Stine and D. B. Teplow, *Neurobiol. Aging*, 2004, **25**, 569–580.
- 27 M. Ahmed, J. Davis, D. Aucoin, T. Sato, S. Ahuja, S. Aimoto, J. I. Elliott, W. E. Van Nostrand and S. O. Smith, *Nat. Struct. Mol. Biol.*, 2010, **17**, 561–567.
- 28 E. Cerf, R. Sarroukh, S. Tamamizu-Kato, L. Breydo, S. Derclaye, Y. F. Dufrêne, V. Narayanaswami, E. Goormaghtigh, J.-M. Ruyschaert and V. Raussens, *Biochem. J.*, 2009, **421**, 415–423.
- 29 S. Tomaselli, V. Esposito, P. Vangone, N. A. J. van Nuland, A. M. J. J. Bonvin, R. Guerrini, T. Tancredi, P. A. Temussi and D. Picone, *ChemBioChem*, 2006, **7**, 257–267.
- 30 S. Ciudad, E. Puig, T. Botzanowski, M. Meigooni, A. S. Arango, J. Do, M. Mayzel, M. Bayoumi, S. Chaignepain, G. Maglia, S. Cianferani, V. Orekhov, E. Tajkhorshid, B. Bardiaux and N. Carulla, *bioRxiv*, , DOI:10.1101/759472.
- 31 Y. Xiao, B. Ma, D. McElheny, S. Parthasarathy, F. Long, M. Hoshi, R. Nussinov and Y. Ishii, *Nat. Struct. Mol. Biol.*, 2015, **22**, 499–505.
- 32 R. Gustafsson, S. Zhang, G. Masuyer, M. Dong and P. Stenmark, *Toxins (Basel)*, 2018, **10**, 153.
- 33 L. Notari, M. Martínez-Carranza, J. A. Farías-Rico, P. Stenmark and G. von Heijne, *J. Mol. Biol.*, 2018, **430**, 5196–5206.
- 34 N. B. Colthup, L. H. Daly and S. E. Wiberley, in *Introduction to Infrared and Raman Spectroscopy*, 1990.
- 35 P. J. Larkin, *Infrared and Raman Spectroscopy*, Elsevier, 2011.
- 36 A. Barth and C. Zscherp, *Q. Rev. Biophys.*, 2002, **35**, 369–430.
- 37 S. Krimm and J. Bandekar, in *Adv. Prot. Chem.*, 1986, vol. 38, pp. 181–364.

- 38 J. L. R. Arrondo, A. Muga, J. Castresana and F. M. Goñi, *Prog. Biophys. Mol. Biol.*, 1993, **59**, 23–56.
- 39 J. Bandekar, *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.*, 1992, **1120**, 123–143.
- 40 R. D. B. Fraser and T. P. MacRae, *Conformation in Fibrous Proteins and Related Synthetic Polypeptides*, Elsevier, 1973.
- 41 E. Goormaghtigh, V. Cabiaux and J.-M. Ruyschaert, in *Subcell. Biochem.*, 1994, vol. 23, pp. 405–450.
- 42 Y. N. Chirgadze and N. A. Nevskaya, *Biopolymers*, 1976, **15**, 607–625.
- 43 J. Kubelka and T. A. Keiderling, *J. Am. Chem. Soc.*, 2001, **123**, 6142–6150.
- 44 Y. N. Chirgadze and N. A. Nevskaya, *Biopolymers*, 1976, **15**, 627–636.
- 45 S. Krimm and J. Bandekar, *Biopolymers*, 1980, **19**, 1–29.
- 46 P. Lagant, G. Vergoten, G. Fleury and M.-H. Loucheux-Lefebvre, *Eur. J. Biochem.*, 1984, **139**, 149–154.
- 47 A. Barth, *Prog. Biophys. Mol. Biol.*, 2000, **74**, 141–173.
- 48 Y. N. Chirgadze, O. V. Fedorov and N. P. Trushina, *Biopolymers*, 1975, **14**, 679–694.
- 49 S. Y. Venyaminov and N. N. Kalnin, *Biopolymers*, 1990, **30**, 1259–1271.
- 50 K. Rahmelow, W. Hübner and T. Ackermann, *Anal. Biochem.*, 1998, **257**, 1–11.
- 51 H. Susi and D. M. Byler, *Methods Enzym.*, 1986, **130**, 290–311.
- 52 J. K. Kauppinen, D. J. Moffatt, H. H. Mantsch and D. G. Cameron, *Appl. Spectrosc.*, 1981, **35**, 271–276.
- 53 W. K. Surewicz and H. H. Mantsch, *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.*, 1988, **952**, 115–130.

- 54 N. A. Nevskaya and Y. N. Chirgadze, *Biopolymers*, 1976, **15**, 637–648.
- 55 T. La Cour Jansen, A. G. Dijkstra, T. M. Watson, J. D. Hirst and J. Knoester, *J. Chem. Phys.*, 2006, **125**, 44312.
- 56 T. La Cour Jansen, A. G. Dijkstra, T. M. Watson, J. D. Hirst and J. Knoester, *J. Chem. Phys.*, 2012, **136**, 209901.
- 57 H. Torii, *J. Phys. Chem. A*, 2004, **108**, 7272–7280.
- 58 E.-L. Karjalainen, T. Ersmark and A. Barth, *J. Phys. Chem. B*, 2012, **116**, 4831–4842.
- 59 E. L. Karjalainen, H. K. Ravi and A. Barth, *J. Phys. Chem. B*, 2011, **115**, 749–757.
- 60 L. C. Mayne and B. Hudson, *J. Phys. Chem.*, 1991, **95**, 2962–2967.
- 61 H. Torii, T. Tatsumi, T. Kanazawa and M. Tasumi, *J. Phys. Chem. B*, 1998, **102**, 309–314.
- 62 S. Krimm and Y. Abe, *Proc. Natl. Acad. Sci.*, 1972, **69**, 2788–2792.
- 63 Y. Abe and S. Krimm, *Biopolymers*, 1972, **11**, 1817–1839.
- 64 W. R. W. Welch, T. A. Keiderling and J. Kubelka, *J. Phys. Chem. B*, 2013, **117**, 10359–10369.
- 65 W. R. W. Welch, J. Kubelka and T. A. Keiderling, *J. Phys. Chem. B*, 2013, **117**, 10343–10358.
- 66 E. G. Buchanan, W. H. James, S. H. Choi, L. Guo, S. H. Gellman, C. W. Müller and T. S. Zwier, *J. Chem. Phys.*, 2012, **137**, 094301.
- 67 J. Kessler, V. Andrushchenko, J. Kapitán and P. Bouř, *Phys. Chem. Chem. Phys.*, 2018, **20**, 4926–4935.
- 68 L. Wang, C. T. Middleton, M. T. Zanni and J. L. Skinner, *J. Phys. Chem. B*, 2011, **115**, 3713–3724.
- 69 J. P. Lomont, J. S. Ostrander, J.-J. Ho, M. K. Petti and M. T. Zanni, *J. Phys. Chem. B*, 2017, **121**, 8935–8945.

- 70 P. Hamm and S. Woutersen, *Bull. Chem. Soc. Jpn.*, 2002, **75**, 985–988.
- 71 S. J. Roeters, C. N. van Dijk, A. Torres-Knoop, E. H. G. Backus, R. K. Campen, M. Bonn and S. Woutersen, *J. Phys. Chem. A*, 2013, **117**, 6311–6322.
- 72 H. Torii and M. Tasumi, *J. Raman Spectrosc.*, 1998, **29**, 81–86.
- 73 S. Cha, S. Ham and M. Cho, *J. Chem. Phys.*, 2002, **117**, 740–750.
- 74 C. Lee and M. Cho, *J. Phys. Chem. B*, 2004, **108**, 20397–20407.
- 75 S. Hahn, S. Ham and M. Cho, *J. Phys. Chem. B*, 2005, **109**, 11789–11801.
- 76 W. H. Moore and S. Krimm, *Biopolymers*, 1976, **15**, 2439–2464.
- 77 J. Kubelka, J. Kim, P. Bour and T. A. Keiderling, *Vib. Spectrosc.*, 2006, **42**, 63–73.
- 78 R. D. Gorbunov, D. S. Kosov and G. Stock, *J. Chem. Phys.*, 2005, **122**, 224904.
- 79 S. Ham, S. Cha, J.-H. Choi and M. Cho, *J. Chem. Phys.*, 2003, **119**, 1451–1461.
- 80 P. Hamm, M. Lim, W. F. DeGrado and R. M. Hochstrasser, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 2036–2041.
- 81 J. Wang and R. M. Hochstrasser, *Chem. Phys.*, 2004, **297**, 195–219.
- 82 J.-H. Choi, S. Ham and M. Cho, *J. Chem. Phys.*, 2002, **117**, 6821.
- 83 T. Hayashi and S. Mukamel, *J. Phys. Chem. B*, 2007, **111**, 11032–11046.
- 84 T. M. Watson and J. D. Hirst, *Phys. Chem. Chem. Phys.*, 2004, **6**, 998–1005.
- 85 J. H. Choi and M. Cho, *J. Chem. Phys.*, 2004, **120**, 4383–4392.
- 86 R. D. Gorbunov and G. Stock, *Chem. Phys. Lett.*, 2007, **437**, 272–276.
- 87 H. Guo and M. Karplus, *J. Phys. Chem.*, 1992, **96**, 7273–7287.

- 88 H. Torii, T. Tatsumi and M. Tasumi, *J. Raman Spectrosc.*, 1998, **29**, 537–546.
- 89 H. Torii, *J. Mol. Struct.*, 2005, **735–736**, 21–26.
- 90 T. La Cour Jansen and J. Knoester, *J. Chem. Phys.*, 2006, **124**, 044502.
- 91 J. Kubelka and T. A. Keiderling, *J. Phys. Chem. A*, 2001, **105**, 10922–10928.
- 92 J. H. Kim and M. Cho, *Bull. Korean Chem. Soc.*, 2003, **24**, 1061–1068.
- 93 M. P. Gageot, R. Vuilleumier, M. Sprik and D. Borgis, *J. Chem. Theory Comput.*, 2005, **1**, 772–789.
- 94 P. Hamm, M. Lim and R. M. Hochstrasser, *J. Phys. Chem. B*, 1998, **102**, 6123–6138.
- 95 C. Scheurer, A. Piryatinski and S. Mukamel, *J. Am. Chem. Soc.*, 2001, **123**, 3114–3124.
- 96 P. Hamm and M. Zanni, *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, 2011.
- 97 H. S. Chung and A. Tokmakoff, *J. Phys. Chem. B*, 2006, **110**, 2888–2898.
- 98 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.
- 99 H. Maekawa, C. Toniolo, Q. B. Broxterman and N.-H. Ge, *J. Phys. Chem. B*, 2007, **111**, 3222–3235.
- 100 B. Lee and F. M. Richards, *J. Mol. Biol.*, 1971, **55**, 379–400, IN3–IN4.
- 101 M. Petukhov, G. Rychkov, L. Firsov and L. Serrano, *Protein Sci.*, 2004, **13**, 2120–2129.
- 102 F. Jensen, *Introduction to computational chemistry*, John Wiley & Sons, Chichester, UK, 2nd ed., 2007.
- 103 A. R. Leach, *Molecular modelling: principles and applications*, Pearson, Harlow, UK, 2nd ed., 2001.

- 104 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, G. Cheeseman, J. R.; Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, K. Hada, M.; Ehara, M.; Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. E. Montgomery, J. A., Jr.; Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, R. Staroverov, V. N.; Kobayashi, K. Normand, J.; Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, S. Dannenberg, J. J.; Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian, Inc. Wallingford, CT*, 2009.
- 105 P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864–B871.
- 106 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.
- 107 R. G. Parr, *Annu. Rev. Phys. Chem.*, 1983, **34**, 631–656.
- 108 E. Wimmer, in *New Trends in Materials Chemistry*, Springer Netherlands, Dordrecht, 1997, pp. 195–238.
- 109 N. Mardirossian and M. Head-Gordon, *Mol. Phys.*, 2017, **115**, 2315–2372.
- 110 J. P. Perdew and W. Yue, *Phys. Rev. B*, 1986, **33**, 8800–8802.
- 111 J. P. Perdew and Y. Wang, *Phys. Rev. B*, 1992, **45**, 13244–13249.
- 112 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
- 113 A. D. Becke, D.R. Yarkony (Ed.), Singapore, 1995, pp. 1022–1046.
- 114 J. Kubelka, P. Bour, R. A. G. D. Silva, S. M. Decatur and T. A. Keiderling, *ChemInform*, 2010, **33**, 296–296.

- 115 P. Bouř, J. McCann and H. Wieser, *J. Phys. Chem. A*, 1998, **102**, 102–110.
- 116 J. Kim, J. Kapítán, A. Lakhani, P. Bouř and T. A. Keiderling, *Theor. Chem. Acc.*, 2008, **119**, 81–97.
- 117 R. Huang, L. Wu, D. McElheny, P. Bouř, A. Roy and T. A. Keiderling, *J. Phys. Chem. B*, 2009, **113**, 5661–5674.
- 118 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 119 L. N. Zhao, T. Zhang, C. Zhang, C. Wang, L. A. Morozova-Roche, L. Y. Chew and Y. Mu, *RSC Adv.*, 2013, **3**, 24081.
- 120 C. Wang, A. G. Klechikov, A. L. Gharibyan, S. K. T. S. Wärmländer, J. Jarvet, L. Zhao, X. Jia, S. K. Shankar, A. Olofsson, T. Brännström, Y. Mu, A. Gräslund and L. A. Morozova-Roche, *Acta Neuropathol.*, 2014, **127**, 507–522.
- 121 J. O. Matos, G. Goldblatt, J. Jeon, B. Chen and S. A. Tatulian, *J. Phys. Chem. B*, 2014, **118**, 5637–5643.
- 122 S. D. Moran, A. M. Woys, L. E. Buchanan, E. Bixby, S. M. Decatur and M. T. Zanni, *Proc. Natl. Acad. Sci.*, 2012, **109**, 3329–3334.
- 123 J. Hilario, J. Kubelka and T. A. Keiderling, *J. Am. Chem. Soc.*, 2003, **125**, 7562–7574.
- 124 D. Scheerer, H. Chi, D. McElheny, T. A. Keiderling and K. Hauser, *J. Phys. Chem. B*, 2018, **122**, 10445–10454.
- 125 Y. Xu, P. Purkayastha and F. Gai, *J. Am. Chem. Soc.*, 2006, **128**, 15836–15842.
- 126 C. Lendel, M. Bjerring, A. Dubnovitsky, R. T. Kelly, A. Filippov, O. N. Antzutkin, N. C. Nielsen and T. Härd, *Angew. Chemie Int. Ed.*, 2014, **53**, 12756–12760.
- 127 A. G. Kreutzer, I. L. Hamza, R. K. Spencer and J. S. Nowick, *J. Am. Chem. Soc.*, 2016, **138**, 4634–4642.
- 128 A. Sandberg, L. M. Luheshi, S. Söllvander, T. de Barros, B. Macao, T. P. J. Knowles, H. Biverstål, C. Lendel, F. Ekholm-Petterson, A. Dubnovitsky, L. Lannfelt, C. M. Dobson, T. Härd, H. Biverstål, C.

- Lendel, F. Ekholm-Petterson, A. Dubnovitsky, L. Lannfelt, C. M. Dobson and T. Hård, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 15595–15600.
- 129 L. Yu, R. Edalji, J. E. Harlan, T. F. Holzman, A. P. Lopez, B. Labkovsky, H. Hillen, S. Barghorn, U. Ebert, P. L. Richardson, L. Miesbauer, L. Solomon, D. Bartley, K. Walter, R. W. Johnson, P. J. Hajduk and E. T. Olejniczak, *Biochemistry*, 2009, **48**, 1870–1877.
- 130 W. M. Tay, D. Huang, T. L. Rosenberry and A. K. Paravastu, *J. Mol. Biol.*, 2013, **425**, 2494–2508.
- 131 J. Wang, *Phys. Chem. Chem. Phys.*, 2009, **11**, 5310–5322.
- 132 T. C. Cheam and S. Krimm, *J. Chem. Phys.*, 1985, **82**, 1631–1641.
- 133 L. Ackels, P. Stawski, K. E. Amunson and J. Kubelka, *Vib. Spectrosc.*, 2009, **50**, 2–9.
- 134 E. M. Bradbury and A. Elliott, *Spectrochim. Acta*, 1963, **19**, 995–1012.
- 135 E. Suzuki, *Spectrochim. Acta Part A Mol. Spectrosc.*, 1967, **23**, 2303–2308.
- 136 A. Ghosh and R. M. Hochstrasser, *Chem. Phys.*, 2011, **390**, 1–13.
- 137 F. Zehender, A. Labahn and T. Koslowski, *J. Phys. Chem. B*, 2005, **109**, 23674–23678.
- 138 R. Huang, J. Kubelka, W. Barber-Armstrong, R. A. G. D. Silva, S. M. Decatur and T. A. Keiderling, *J. Am. Chem. Soc.*, 2004, **126**, 2346–2354.
- 139 Y. N. Chirgadze, B. V. Shestopalov and S. Y. Venyaminov, *Biopolymers*, 1973, **12**, 1337–1351.
- 140 J. Kubelka and T. A. Keiderling, *J. Am. Chem. Soc.*, 2001, **123**, 12048–12058.
- 141 R. A. G. D. Silva, W. Barber-Armstrong and S. M. Decatur, *J. Am. Chem. Soc.*, 2003, **125**, 13674–13675.

- 142 T. C. T. Michaels, A. Šarić, J. Habchi, S. Chia, G. Meisl, M. Vendruscolo, C. M. Dobson and T. P. J. Knowles, *Annu. Rev. Phys. Chem.*, 2018, **69**, 273–298.
- 143 H. Susi and D. M. Byler, *Arch. Biochem. Biophys.*, 1987, **258**, 465–469.
- 144 R. Khurana and A. L. Fink, *Biophys. J.*, 2000, **78**, 994–1000.
- 145 A. M. Woys, A. M. Almeida, L. Wang, C.-C. Chiu, M. McGovern, J. J. de Pablo, J. L. Skinner, S. H. Gellman and M. T. Zanni, *J. Am. Chem. Soc.*, 2012, **134**, 19118–19128.
- 146 H. Torii, *J. Phys. Chem. Lett.*, 2015, **6**, 727–733.
- 147 H. Torii, *J. Phys. Chem. B*, 2018, **122**, 154–164.
- 148 J. P. Lomont, K. L. Rich, M. Maj, J.-J. Ho, J. S. Ostrander and M. T. Zanni, *J. Phys. Chem. B*, 2018, **122**, 144–153.
- 149 Z. Ganim and A. Tokmakoff, *Biophys. J.*, 2006, **91**, 2636–2646.