Rijksuniversiteit Groningen

MSc Thesis

Examining the Internal Structure of Protein Aggregates using Two-Dimensional Infrared Spectroscopy

---

Author:
Ruben Steendam

Supervisor:
prof. dr. T.L.C. *Jansen*
Second Examiner:
prof. dr. P.C.A. *van der Wel*

A thesis submitted in fulfillment of the requirements for the Master's degree in Physics

Theory of Condensed Matter
Zernike Institute of Advanced Materials

July 21, 2020

RIJKSUNIVERSITEIT GRONINGEN

Abstract

Faculty of Science and Engineering
Zernike Institute of Advanced Materials

Master's degree in Physics

Examining the Internal Structure of Protein Aggregates using Two-Dimensional Infrared Spectroscopy

by Ruben Steendam

Two proteins are examined in order to determine the internal couplings of the PPII helices present in some forms of amyloid protein and to verify a proposed configuration of Islet Amyloid Poly peptide (IAPP). Two-dimensional infrared (2DIR) spectroscopy is used for this purpose, as it has the capability to resolve the structure of proteins within the time frame necessary to observe protein-protein interactions. Using the amide I vibration as a marker for secondary structure, several Hamiltonians are constructed from which information about internal couplings of the protein, and interactions with the environment, can be extracted. To simulate the Hamiltonian during 2DIR, the Numerical Integration of the Schrödinger Equation (NISE) method is used. *In silico* isotopic labeling is performed on IAPP to confirm *in vitro* experimental data and hypotheses about its conformation, which is done by linking results obtained from isotopically labeled absorption spectra with predictions due to configuration and results from *in vitro* experiments. Although no definitive pattern has yet been discovered for PPII helices using *in silico* 2DIR, further research into proteins containing PPII structure could yield valuable comparative information about the possible internal couplings present within such secondary structures. This can be achieved by comparing 2DIR results of additional proteins containing the PPII structure to the absorption spectra obtained from this research. The isotopically labeled absorption spectra and couplings show that the IAPP configuration proposed is indeed a viable iteration of the amylin amyloid, but subsequent experimental research will be needed to verify these claims.

Acknowledgements

It was the beginning of March and I had just started my research at the Theory of Condensed Matter physics group at the Zernike Institute of Advanced Materials. The coronavirus hit hard and fast and we all had to adept quickly to a new routine, and a way of life, that would soon become "the new normal". Luckily, I was surrounded by kindhearted individuals who understood the difficulties we all had to endure. They have helped me adept to working from home, and be as independent as possible.
I would like to thank Martin T. Zanni, Patrick C. A. van der Wel and Thomas L. C. Jansen for their scientific guidance, help and cooperation during my research. J. Looge and Maarten P. Boneschansker for their continuous support. And finally, I would like to thank my friends for their love and support during these strange times.

Contents

List of Figures

Dedicated to the people who kept me going during these strange times.

Chapter 1

Introduction

1.1 General Introduction

Proteins are ubiquitous inside every living system as they tend to nearly every biological function imaginable. The functioning of proteins is highly linked to their structural conformation. As reactive sides fold into themselves and expose other parts of their conformation, it is possible for proteins to carry out, and react to, many different chemical and thermodynamical pathways. Some functionalities that proteins can express are as enzymes, which are a form of catalysts that promote chemical reaction within cells.[9] Or, as antibodies that function as protection of the body.[10]

Sometimes, either due to damage in the cell, or by genetic mutation, the synthesis of proteins can become impaired, and the conformation of the proteins constructed can change.[11] Due to this divergence in conformation the proteins can become prone to fold themselves erroneously, leading to aggregation of clusters of defective proteins. Or, in the case of certain prions, in proteins that cause extensive harm to biological life and can directly be linked with cell-death.[12]

Amyloid is an amorphous aggregate of proteins that can appear in many organs of the human body as either a symptom or cause of many (degenerative) diseases.[13] As of yet at least fifty diseases are recognized that are characterized by the deposition of amyloid.[14,15] These diseases can range from Parkinson's disease, Huntington's disease, rheumatoid arthritis, type II diabetes, spongiform encephelopathies (a form of prion disease) etc.[16–20]

This research looks at two proteins that exhibit either some form of amyloid synthesis or show a specific configuration that will hopefully help us to understand the dynamics behind the formation of amyloid, with the intent of finding a way to disrupt the formation entirely.

1.2 Protein Structure

Proteins, and thus consequently aggregates, consist of amino acids. These building blocks are linked through a backbone of repeating nitrogen-carbon-carbon (NCC) sequence. (Figure 1.1) This NCC backbone repeats itself while the residues attached to it determine what kind of amino acid the protein contains. These backbones are not rigid but can move freely as long as the external potential energies allow it, these dihedral angles are notated as ϕ, Φ and ω in Figure 1.1.

The sequence in which these amino acids are linked, the primary structure of the protein, determines the shape of the protein and most importantly its final function. As amino acids can be either hydrophobic or hydrophilic, certain hydrogen bonding sites can cause the protein to fold into itself in order to maintain a thermodynamic minimum energy.[21] This folding of the protein causes certain secondary structures to

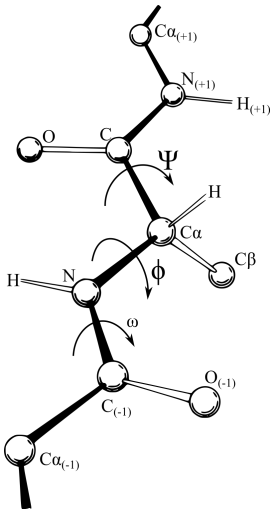


Figure 1.1. A simplified model of a protein backbone. The dihedral angles ϕ, ψ and ω are shown and represent the angles between two intersecting planes within the protein. *Image adapted from Adam Redzikowski.* [1]

appear that can be subdivided based on geometric constraints and hydrogen bonding patterns present within the protein.[21,22] Most secondary structures present in proteins can be classified as either α-helices or β-sheets. Prominent structures that cannot be categorized into either of these two groups are commonly classified as "other" or "random coil". The categorization and placement of secondary structures within proteins can be used to specify orientation, binding sites, and more structural information, and can therefore be used to illustrate the composition of the protein. (Figure 1.2)

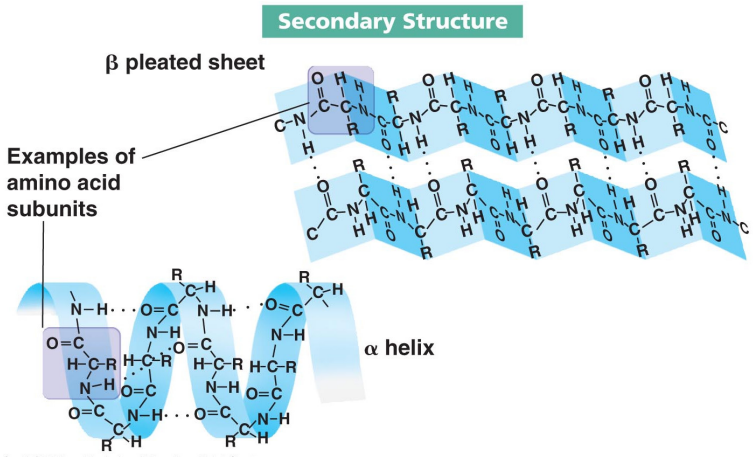


Figure 1.2. Shown are two secondary structures that occur due to interactions between amino acids. The α-helix curls up due to the close proximate interactions with neighbouring amino acids. β-sheets form between β-strands that occur when two segments of a protein overlap and form a row of hydrogen bonds.*Image adapted from Campbell et al.* [2]

One of the random coils that is of interest in this research is the Polyproline-II (PPII) helix. PPII helices are are formed when all the dihedral angles in the backbone adopt angles of -75° and 150°, causing the structure to exhibit a full rotation every three residues.† (Figure 1.3) Although the name suggest that this coil only appears

†An amino acid contained within a protein is called a residue

in proteins rich in proline, the structure also appears in other proteins.[23] Unlike other secondary structures present in proteins, PPII structures are open and show no internal hydrogen bonding.[24] PPII helices are known to be involved in transcription, self-assembly, cell motility, elasticity and immune response reactions.[25] The PPII structure has an important role in proteins that can form amyloid aggregates, but unfortunately they are not given in experimentally solved structures and are rarely used when modeling proteins. As of yet little to no research has been done on the characteristics of PPII in proteins, even though the rate of cell death has been linked to the presence of fibrils containing PPII.[4,17,26] In order to fully understand the functionality of proteins, or the formation of amyloid, it is paramount to understand the composition of such proteins, their (secondary) structure, and their dynamics.

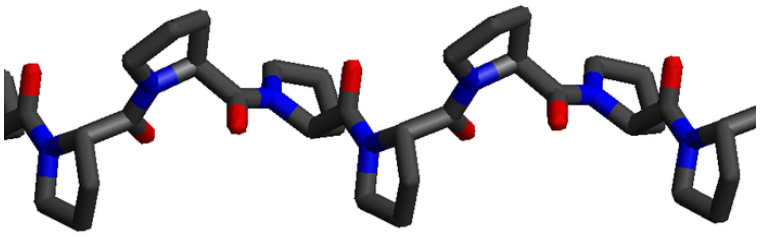


Figure 1.3. A side view of the Polyproline-II helix. The openness of the structure explains the lack of hydrogen bondings within it. The full rotation every three residues can also be observed.*Image adapted from WilowW.*[3]

1.2.1 The Formation of Protein Aggregates

An example to showcase the importance of understanding the PPII structure is the formation of specific depositions of amyloid during Huntington's disease.[17,27,28] Due to a mutation in the gene that transcribes for the formation of the Huntingtin protein it is possible that a long repeated chain of glutamine can occur within the formed protein. This so called Poly-Glutamine (Poly-Q) domain is prone to form clusters (aggregates) with other proteins that also carry this Poly-Q segment.[4,27,29] PPII structures can be found while examining these Huntingtin protein aggregates, as they occur as secondary structure in the rest of the mutated Huntingtin protein as fibrils next to Poly-Q stacks. A stacking of Poly-Q in mutagenic Huntingtin proteins can be seen in Figure 1.4. It is proposed that not necessarily the stacking of the Poly-Q, and thus the aggregation of the Poly-Q domains themselves, is the most cytotoxic. But, the fibrils that exhibit PPII structure next to it are the cause of cell death in the presence of amyloid.[4,30]

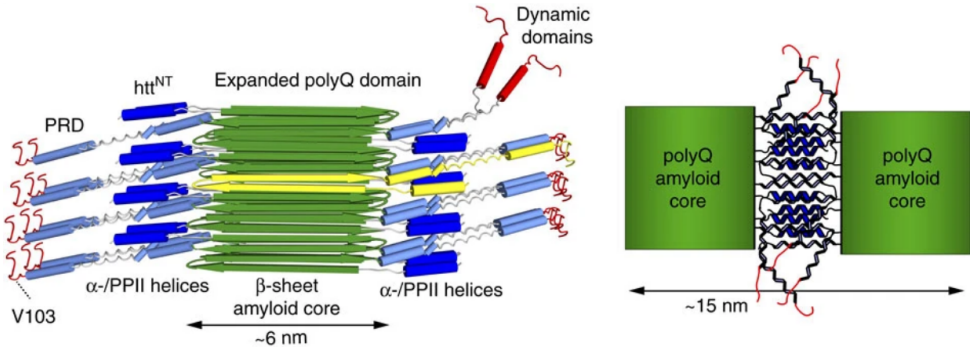
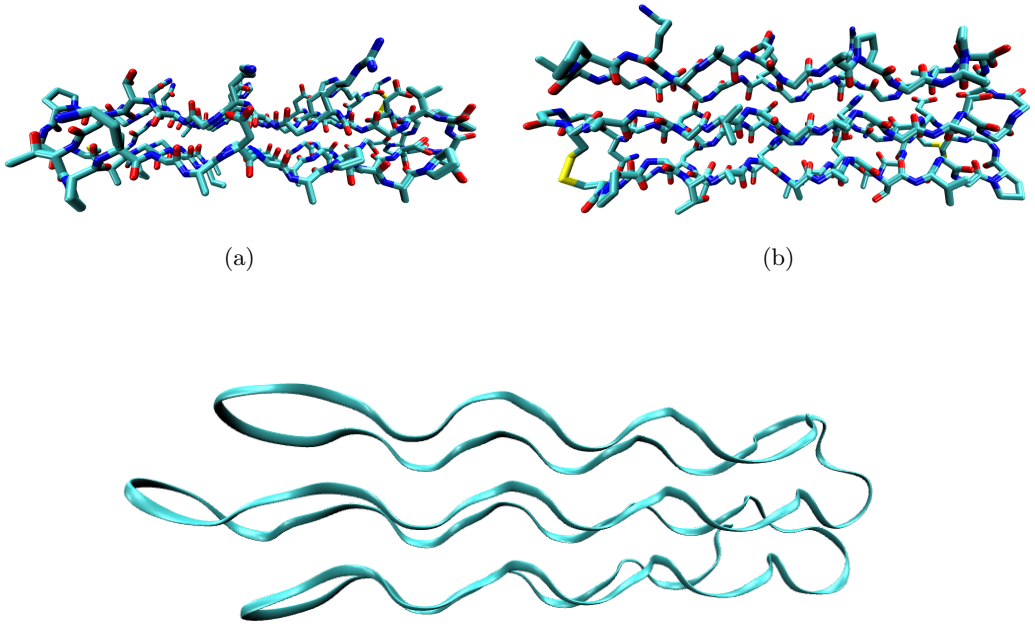


Figure 1.4. Due to stacking of the Poly-Q domains an amyloid core is formed. The Huntingtin mutagenic protein exhibits 3 structures, the Poly-Q domain (green), α helices (dark blue) and PPII helices (light blue). A single protein is shown in yellow, and shows the folding within the protein. The image right shows the interfilament flanking between two amyloid cores. It is proposed that the configurations, and angles between PPII helices, that are formed here determine cell toxicity. *Image adapted from Lin et al.*[4]

One of the reasons we look at the structural formation of such fibrils is that the exact mechanisms that cause cell death are not known, although links between its structural configuration and reactivity with the cells is highly investigated.[31] Unfortunately even processes that have been linked to the formation of amyloid and cell death are poorly understood.[32,33] And, some even speculate that it is not the final formation, but the intermediate configurations of amyloid formation that are the most toxic to living cells.[34] In addition to that it has been found that the formation of amyloid fibrils *in vitro* has been greatly accelerated by the presence of lipid membranes, although this link with *in vivo* experiments is not yet understood.[35] As some experiments have shown that one of the speculated reasons for cytotoxicity is the disruption of membranes forming due to the presence of protein aggregation.[14] One of the reasons that the exact nature of protein aggregation is still unclear is that the existing experimental setups do not posses the required structural or time resolution necessary to observe real time protein-protein interaction and aggregation kinetics.[8]

1.2.2 Antifreeze Protein 2PNE

The first protein of interest in this research is the snow flea antifreeze protein (2PNE). The protein data bank (PDB) file for 2PNE was obtained by Pentelute *et al.* Pentelute and coworkers used X-ray crystallography to structurally resolve a protein previously found in a Canadian snow flea during initial research by Davies *et al.*[36–38] PDB files contain the positions of the atoms that are known to be present within the proteins. The most important aspect of 2PNE is its secondary structure, as the protein consists of six left-handed anti parallel PPII helices, which are stacked in two sets of three and form a compact brick-shaped structure. (Figure 1.5) As one of our main point of interest is the PPII structure, 2PNE can help structurally resolve PPII. As was stated earlier, most proteins exhibiting Poly-Q characteristics are hard to resolve structurally.[39] By looking at a protein that exhibits mainly PPII structures it may be possible to discover a certain fingerprint. Which may help resolve the structures within, and around, the Poly-Q region of several protein amyloid diseases and maybe discover the link between their conformation and observed cytotoxicity.



(a) (b)

(c)

Figure 1.5. (a) Top view, (b) Side view and (c) Ribbon representation of the 2PNE protein. The protein primarily consists of six anti-parallel PPII coils in a compact brick-shaped configuration.

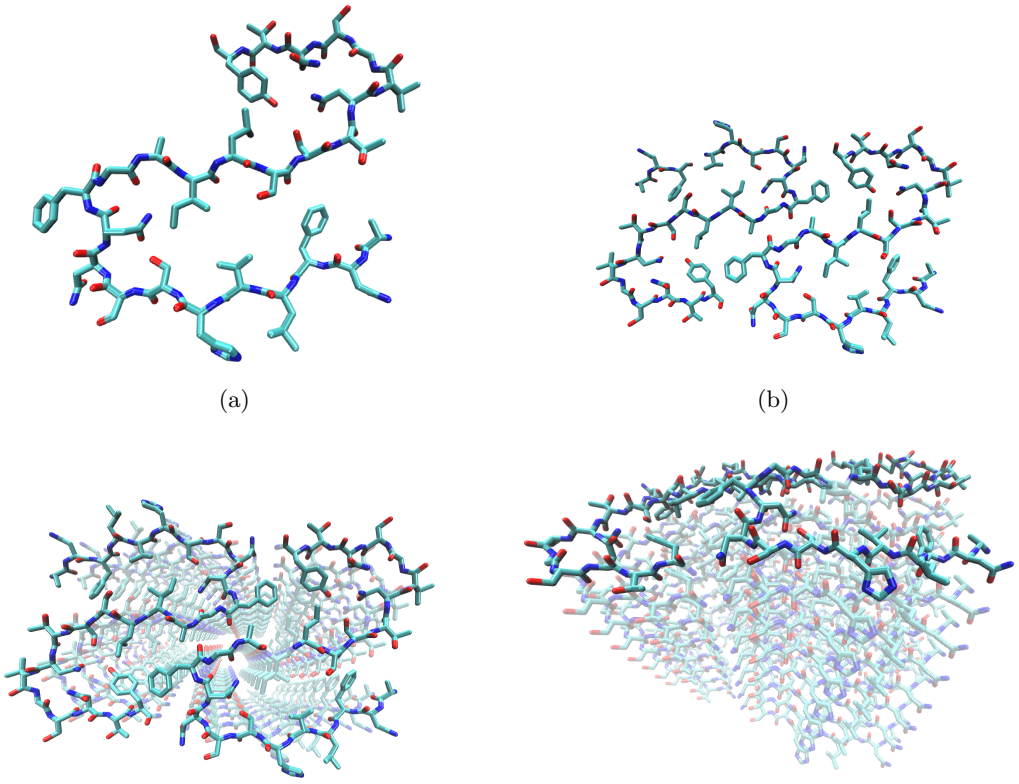
1.2.3 Amylin Fibrils 6Y1A

6Y1A is a certain configuration of the islet amyloid polypeptide (IAPP) that appears in the pancreas during diabetes mellitus type II.[40,41] It is proposed that the presence of these amyloid peptides in the pancreas causes the beta cells to die, although this is highly debated.[42] The PDB structure for 6Y1A was created by Röder *et al.* where cryo-electron microscopy was used to resolve the structure of the protein.[5,41,43] 6Y1A consist of two S-shaped interwoven filaments, that have the ability to stack lengthwise into amyloid, the same way as IAPP. (Figure 1.6)

The 6Y1A configuration is based on IAPP, but differs slightly in structure as only the residues that have been attributed to the direct formation of the amyloid have been kept from the original IAPP fibril. (figure 1.7) These residues are the Phe23-Gly24-Ala25-Ile26-Leu27 (FGAIL) sequence present in the middle of the protein, the tyrosine 37, and the amidated C-termini, which are essential for the forming or amyloid in IAPP.[43–45]

The structure of 6Y1A used during this research appears similar to Amyloid β, a protein amyloid found in the presence of Alzheimer's disease.[46,47] Similarities between the two proteins include having similar β-sheet structures, binding to the same receptor, and being degraded by the same enzyme protease.[43,48–50] Because of these similarities it is assumed that amylin fibrils may be associated with amyloid β, but the exact nature of these similarities remain unclear.[47,48]

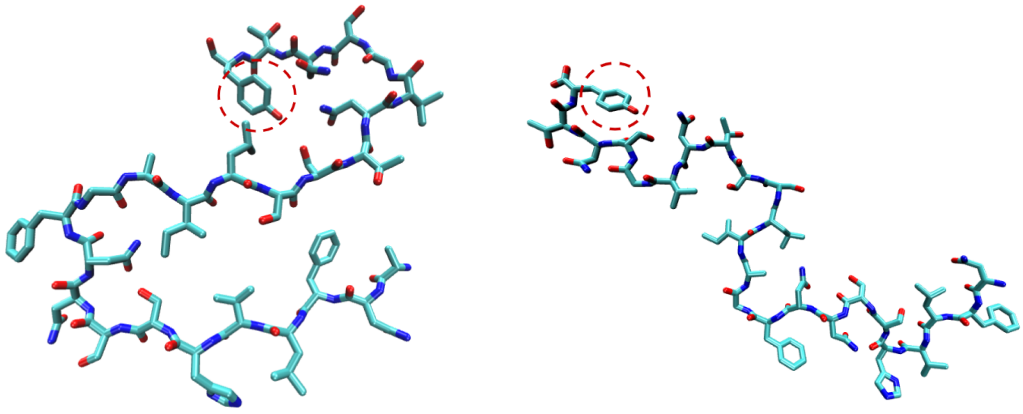
One of the reasons *in silico* experiments have been performed on 6Y1A is that experimental data has not been sufficient to conclude a definitive structure for this protein.[8] Although (infrared) absorption spectra of IAPP are known, less data of 6Y1A is available, meaning that more comparative data is necessary to verify claims about the correctness of 6Y1A as a configuration of IAPP.



(a) (b)

(c) (d)

Figure 1.6. How the 6Y1A aggregate is built up. (a) A single protein chain. (b) A single layer conformation consisting of two protein chains. (c) Top view of the complete aggregate. (d) 3/4 view of the complete amylin aggregate.



(a) 6Y1A (b) IAPP

Figure 1.7. A comparison between the structure of 6Y1A and IAPP. The residues that contribute the most to aggregation of IAPP are kept in the 6Y1A. The same tyrosine at the terminal of a single protein chain can be seen encircled in both proteins. *IAPP structure obtained from Cao et al.*[5]

1.3 Resolving the Structure of Proteins

There are several ways to determine the structure and composition of proteins, as was discussed in the origins of the proteins in section 1.2.2 and 1.2.3. X-ray crystallography is one of the known ways of examining the structure of proteins and was used to determine the structure of 2PNE.[51] By scattering x-rays on the protein it is possible to track the conformational changes that the protein undertakes, unfortunately this method lacks the dynamical information that sometimes is required to examine protein-protein interactions. This was a motivation for the team of Hamm *et al.* to develop the femtosecond non-linear-infrared spectroscopy method.[52] Infrared (IR) and two-dimensional infrared Spectroscopy (2DIR) have over the years been proven to be the ideal methods of examining the structures of proteins, especially since the time resolution of these methods is small enough to analyse protein dynamics within the femtosecond.[7,52]

1.3.1 Amide I Vibrations

Amide I vibrations are one of several distinct vibrational modes present in amino acids. They vibrate in the domain between 1600 and 1700 cm^{-1} and are caused by the stretching of the C=O bond present in the backbone of proteins. (Figure 1.8)

One of the reasons the Amide I vibrations are used as a marker during IR spectroscopy, is that it has a high excitation coefficient. This implies that when infrared pulses are used during the examination of proteins, the amide I vibrations are not only sensitive to the surrounding, but also the internal structures of the protein.[53] From this it follows that it is possible to probe the secondary structure present in proteins through the observation of the vibrational spectra of the amide I band.[54]

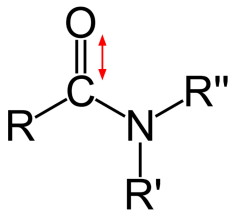


Figure 1.8. A schematic representation of the Amide I Vibration in the backbone of a protein. Exhibited as stretching of the C=O bond as shown by the red arrow.

As such this the amide I vibration has been used as a marker for secondary structures in proteins, where it it has been determined that that β-sheet secondary structure shows two peaks at 1630 cm^{-1} and 1680 cm^{-1} and the α-helices show peaks in the absorption spectrum at 1650 cm^{-1}.[55–57] As such 2DIR spectroscopy can be used to show the difference between α-helices and β-sheets within proteins.[57,58]

Unfortunately, it is not yet known if PPII has such a characteristic absorption spectrum. If it is possible to show the existence of PPII helices inside proteins through a specific absorption peak. It might be possible to resolve such secondary structures within proteins with the use of 2DIR. This would, conceivably, lead us one step closer to solving the complex dynamical processes that cause cell-death due to the presence of amyloid.

The problem with using the amide I vibrations as markers, is that the vibrations caused by excitation can spread over the entire protein due to internal amide couplings. These vibrational couplings causes the excitation to spread its energy over multiple residues throughout the protein, as vibrations between different residues are linked through electronic interaction. Fortunately, the accuracy of 2DIR can be enhanced when several residues are altered using isotopic labeling.[59]

The vibrational frequency of atomic bonds can be given by the equation of motion for a particle on a spring:

$$\omega = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \tag{1.1}$$

Where ω denotes the frequency that the particles vibrates at. k is the 'spring constant', which in the case of proteins signifies the strength of the bond between two atoms, and m is the mass of the atom in question. Isotopic labeling relies on the fact that by exchanging ^{12}C and ^{16}O by the heavier atomic isotopes ^{13}C and ^{18}O the frequency of vibration changes by 70 cm^{-1}.[60–62] Which causes the absorption wavelengths of the labeled residues to shift away from the bulk and give site specific information.[63] As we only change the residue by its isotope this replacement will have minimal effect on the protein as a whole. This makes 2DIR spectroscopy ideal for the examination of structures and dynamics within the aggregation processes of proteins.

Chapter 2

Methods

In order to accurately analyze the proteins in question we need to first simulate their behaviour and interactions with Molecular Dynamic simulations. From these simulations exciton Hamiltonians will be constructed using so-called mapping calculations. The Hamiltonians will contain the strength of the amide I vibrations and internal vibrational couplings, and will thus contain all the necessary information to simulate 2DIR spectroscopy *in silico*. This will be done using the Numerical Integration of the Schrödinger Equation (NISE) method, which has been by implemented into the NISE programme by T. L. C. Jansen and coworkers.[6,64] From the results of the 2DIR spectroscopy the internal structure of the proteins can be viewed as a function of several wavelengths. These results gives a close look into the internal couplings present in the protein and will hopefully yield a certain pattern that, when compared to other results, can help solve several question about protein conformation, interactions, and functionality.

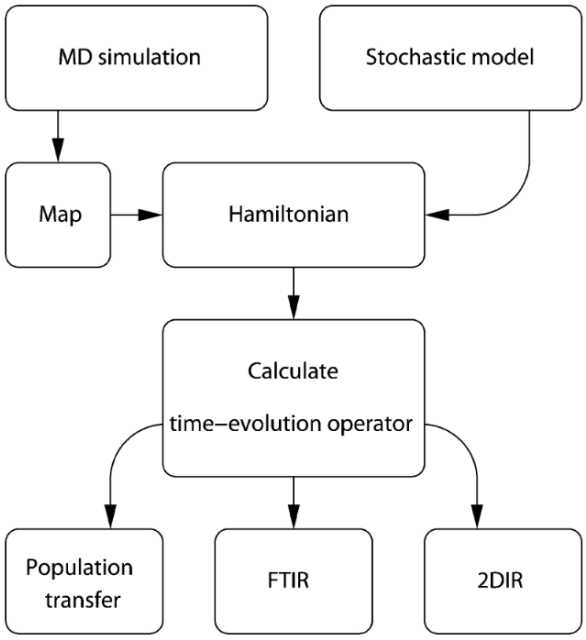


Figure 2.1. A flow diagram showing the processes preceding the NISE software *Image adapted from Jansen et al.*[6]

2.1 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations are used to study the dynamics of atoms and molecules, where they are allowed to move freely for a certain a amount of time. By giving the initial interaction energies between individual atoms, the inter-atomic forces can be calculated, and prospective movement for the proteins can be simulated for every time step of the simulation. MD simulations are performed using classical methods by setting up a certain force field. These force fields contain function that describe the potential energy of the atomic interactions that can occur between the different atoms present in the system.[65] Equations described within the force fields can specify the force that represents the bonds between the atoms, or the energetic potentials for the dihedral angles. These potential equations can be derived from experiments or by solving ab initio quantum mechanical calculations.[66] When a MD simulation is started, the interactions between every atom is calculated using their distance, and the necessary potentials given by the force field for every time step specified. It should therefore be noted that the accuracy of the MD simulation depends almost wholly on the accuracy of the force field, which will be discussed in Chapter 4. Gromacs will be used as the main software package to perform the MD simulations during this research.[67–73]

It should be stated that for both proteins discussed in this research the same initial inputs were used during MD simulations. All input files used during this research can be found on my Github repository.[74] It should first be noted that during this research an older version of Gromacs (Version 4.6.3) will be used.[‡] This is to ensure that the ensuing steps are successful, as the programme used to construct the Hamiltonians is build on this version of Gromacs. Following normal MD simulation procedures; The system is solvated and ions are added to the system to ensure that a net zero charge is maintained. The temperature inside the system is set to 300° Kelvin and allowed to energetically minimize. All inter-atomic interactions are calculated using the OPLS-AA force field, which has been proven to be the most effective at calculating the behaviour of proteins at the time of writing.[75,76]

Based on earlier observations on the vibrational states on the C=O bond in haemoglobin, it was found that relaxation of the excited states takes place in c.a. 20 picosecond.[77,78] For this reason a time step of 20 femtosecond interval was chosen for the simulations, which entails that the programme saves the coordinates and directions of the atoms within the protein every 20 femtoseconds to a trajectory file, for the entire duration of one nanosecond, yielding a total of 50,000 time steps. This trajectory file, in combination with the topology of the system, can subsequently be used to analyze the behaviour of the entire protein.

2.2 Constructing a Hamiltonian for the Amide I Vibrations of the Proteins

After initial MD simulations have been performed on the proteins in question, the necessary information needs to be extracted such that we can construct a specific Hamiltonian of the system, which shows the behaviour of the amide I vibrations within the protein.[79] This Hamiltonian will contain the necessary information for further examination of the amide I stretching as it will show not only the frequencies of the amide I vibrations within the protein, but also their couplings. In order to accurately

[‡]The newer versions of Gromacs do not differ from this version as far as research on proteins is concerned.

interpret the results of experiments several theoretical calculations need to be made on the MD simulations. Vibrational spectra of the amide I bands are obtained by treating the amide stretching quantum mechanically and the other degrees of freedom present in the protein classically.[80,81]

Because a normal MD simulation gives information on the position and trajectory of all the atoms in the protein within a certain time frame, the evolution of this Hamiltonian can be extracted, and a specific one-exciton Hamiltonian can be constructed. The diagonal terms within this Hamiltonian will describe the transitional frequencies of the amide I band and the off-diagonal terms will describe the couplings between pairs of residues.[79] Using this Hamiltonian it is possible to calculate the amide I couplings within a protein, without the use of ab initio electronic structure calculations for every time step, using nothing more than the outputs of standard MD simulations. In order to do this the information from the MD simulations needs to be mapped.

Several maps have been developed that calculate the Hamiltonian from MD simulations. First the electronic structure can be calculated for a grid of relevant dihedral angles within the protein. This is done by using a method called the Hessian matrix reconstruction approach.[82] This Hessian is a matrix of second order partial derivatives of a scalar valued function. This approach leads to the Nearest-Neighbour Frequency Shift (NNFS) maps.[83] These NNFS maps show how the amide I frequency of the residues depend on the neighbouring dihedral angles. Next to the NNFS interactions, the Hessian matrix reconstruction can be used to determine the Nearest-Neighbour Coupling (NNC) maps and more distant Transition Dipole Couplings (TDC) to predict all internal interactions that can influence amide I behaviour within a protein.[59,80]

Next to the internal interactions, the effects of the environment to the perturbation of the protein need to be taken into account. Many of these maps exists, where the most common method is by taking a molecule with only one amide I residue, namely N-Mehtylacetamide (NMA), and placing it in water.[84] The interaction that the NMA molecule has with the electric potential or electric field of the different atoms of the solvent can be mapped to accurately determine the environmental interactions.[85,86] It should be noted that the amino acids Asparagine (Asn) and Glutamine (Gln) also have amide groups in their sidechains, meaning that their presence influence internal couplings. This can be solved by taking their existence into account when creating the exciton Hamiltonian.[79] This is done by considering a different mapping, as their surroundings are different from backbone amide I vibrations. These different vibrations can be mapped using N-deuterated acetamide (ACED).[79] The Hamiltonians used in this research will be generated with the AmideIMaps programme written by S. Roy and coworkers, where the Skinner mapping will be chosen as it is the easiest way to account for the presence of amide I vibrations in the sidechains of the protein.[79,84,87–91]

2.3 Analyzing the Hamiltonian of the Proteins

2.3.1 Numerical Integration of the Schrodinger Equation

It is stated by Hamm *et al.* that the infrared pump probe spectroscopy is capable of giving us a much more information about the amide I vibrations present within proteins than by using other similar methods.[52] Next to the increased time resolution, *in vitro* 2DIR is capable of analysing proteins in solution, which nullifies the necessity of crystallizing proteins before experimentation.[52]

In order to simulate this process *in silico* we make use of the Numerical Integration of the Schrödinger Equation (NISE), which has been transferred by T.L.C. Jansen and coworkers into the NISE programme.[6,64,92–94] The NISE method first divides the system into a part that is allowed to react with the applied laser fields (The system) and the other part which does not (The Bath). The bath is obtained by the previous amide I mapping as a Hamiltonian that describes the environment.[92]. One important assumption made is that the system is affected by the applied laser field and the bath, but the system does not affect the bath. Using this bath it is now possible to simulate the trajectory of the Hamiltonian of the system. This is done quantum-mechanically, which implies that the Hamiltonian of the system is now considered a wave that changes according to the time-dependent Schrödinger.

$$\frac{d\Phi(t)}{dt} = -\frac{i}{\hbar}H(t)\Phi(t) \tag{2.1}$$

Where the wave-function Φ describes the evolution of the Hamiltonian $H(t)$. The problem is that such an equation cannot be solved as long as both the Hamiltonian and the Schrödinger equation are time-dependent. This is solved by dividing the trajectory of the Hamiltonian into very short time intervals, which can taken to be constant.[95] By solving all these time-independent frames numerically and putting them back together, the time-evolution of a Hamiltonian can be calculated and used to solve the Schrödinger equation of the system.

The Generic Hamiltonian used to describe 2DIR experiments is given by equation 2.2.[6]

$$\begin{aligned} H(t) = \sum_i \omega_i(t) B_i^+ B_i + \sum_{i\neq j} J_{ij}(t) B_i^+ B_j - \sum_j \frac{\Delta_i(t)}{2} B_i^+ B_i^+ B_i B_i \\ + \sum_i \vec{\mu_i}(t)\vec{E}(t)(B_i^+ + B_i) \end{aligned} \tag{2.2}$$

Here B^+ Denotes the Bosonic creation operator, and B denotes the Bosonic annihilation operators of the vibration i, with frequency ω_i. The coupling between a pair of vibrations is given by J_{ij}. The $\Delta_i(t)$ is the anharmonicity of the system and can be said to be the deviations of the system from being a perfect oscillator. This value has been set to 16 following earlier research by Hamm *et al.*[52]. $\vec{\nu_i}$ is the transition dipole vector and $\vec{E}(t)$ denotes the externally applied laser field (the laser pules that the NISE method emulates). Most of the necessary values for this Hamiltonian equation are obtained from initial mappings on MD simulations. With *in vitro* 2DIR, three laser pulses (ω_1,ω_2 and ω_3) are fired in sequence with a certain time delay (t_1,t_2), before detection is measured (t_3). After initial excitation by a pump pulse, the system is allowed to react before a probe pulse is used to determine how the system manipulated the prior excitation through the pump pulse. The laser field that is applied to the Hamiltonian through the NISE software emulates these inputs and simulates the response of the system. Throughout this research t_2 has been set to zero, which implies that the system has no time to disperse the initial excitation through the system.

In order to get useful results from the data, the data needs to be Fourier transformed with respect to the delay between the first and second pulse (t_2), as well as transforming them with respect to the time delay between the final pulse and emission of the system (t_3). Finally we will end up with a 1D response and absorption spectrum as well as a 2D absorption spectrum. These absorption spectra can be used to analyse the configuration and composition of the proteins. As the absorption

spectrum is nothing more than the internal couplings of the protein (The structure) and the diagonal energies present (The environment).

Next to calculating 2DIR spectra, it is possible to take the energy mapping files generated by the previous step, and generate an average Hamiltonian of the entire time frame within NISE using the Analyse functionality.[64] This will yield the average value of the Hamiltonian of the entire protein. These averages are necessary as the analysis of the Hamiltonian depends on the average dipoles during the entire simulation during excitation (Section 3.3.1), and it will allow us to verify whether isotopic labeling will have been successful by highlighting strong couplings within the protein.

2.4 Isotopic Labeling

To understand exactly which residues within the protein cause the behaviour we wish to examine, several residues are selected within the NISE programme and run through the 2DIR script. The selected residues will be the same as that discussed in section 1.2.3 and in the research by Wang *et al.* This entails that the residues that will be examined are Alanine (Ala) on positions 13 and 25, Leucine (Leu) on position 16 and 27, Serine (Ser) on position 19 and 20, Phenylalanine (Phe) on position 23 , Glycine (Gly) on position 24 and 33, Isoleucine (Ile) on position 26, Valine (Val) on position 32, and Tyrosine (Tyr) on position 37.[8] The placement of the residues within the complete 6Y1A amyloid can be found in section A.4.

The idea behind isotopic labeling is explained in section 1.3.1. Within the NISE programme it is possible to select the residues that you wish to examine without the encumbrance of doing the experiments *in vitro*. This makes *in silico* isotopic labeling a perfect tool to compare data from different experiments. By using isotopic labeling the absorption spectrum of that specific residue will move away from the main amide I vibrational bonds.[60,61] As the normal amide I vibrations are quite delocalized, such isotopic labelings can help pinpoint the behaviour of the residues within the protein. The main focus of the isotopic labeling performed is to determine the experimental validity of the observed Absorption spectra from the labeled proteins from Skinner *et al* and Serrano *et al.*[8,44]

Chapter 3

Results

3.1 Molecular Dynamics Simulations

The first step is to check the validity of the MD simulations performed. If the simulations were to give erratic or false behaviour of the proteins, then all subsequent data will be wrong. This can be done by a quick visual examination to verify that the protein did not collapse during simulation. The internal structure of the protein cannot be verified this way, but as internal couplings will become apparent in the next step the validity of the MD simulations can be re-verified thereafter.

3.2 Analyzing the Hamiltonians

In order to understand which couplings within the protein are responsible for the spectrum obtained with the NISE programme, it is important to look at the Hamiltonian of the proteins themselves and search for the most prevalent couplings. Because the diagonal entries of the Hamiltonian files show the strong amide I vibrations of the protein, these need to be filtered out for visual inspection, as the internal couplings between residues are of a much smaller scale. In order to ensure that all following visual inspections of the Hamiltonian will use the same colour scheme, the diagonal entries of the Hamiltonian will be set to zero. Two values of -11 and 15 will be placed somewhere along the Hamiltonian diagonal to secure a constant colour grading along all obtained Hamiltonians during this research.[74]

2PNE contains an Asn residue at positions 11, 35, 38 and 40. This means that four extra sidechain amide I vibrations need to be accounted for. The way the Hamiltonian is constructed, the backbone of the protein is calculated and written first and sidechains are accounted for afterwards. This can be see in Figure 3.1 where internal couplings of the protein are visualized.

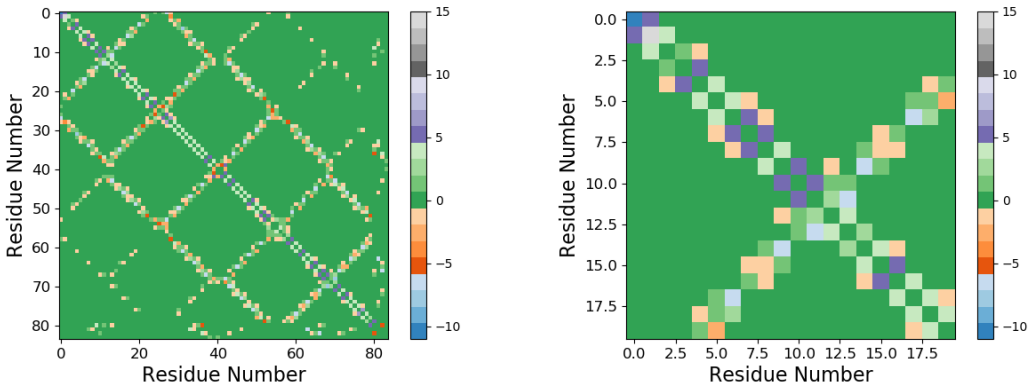


Figure 3.1. The strength of internal couplings within the entire 2PNE protein (left) and the same Hamiltonian zoomed in for the first 20 residues (right). The diagonal entries are removed and a gradient is applied.

The values presented within these Hamiltonians show the value of J_{ij} in Equation 2.2 and are given without units. The repetition of the six PPII helices can clearly be seen, as strong couplings between local residues in a turn are marked with strongly coupled, repeated, crosswise patterns. The first 80 residues are the amide I vibrations in the backbone, while the following four are the sidechain vibrations due to the presence of Asn. The relative strength of the coupling can be read from the colourbar.

Following the analysis of the Hamiltonian of 2PNE, the Hamiltonian of 6Y1A can be visualized using the same parameters. The 6Y1A amyloid consists of 16 protein chains (A-P) and each protein chain has Asn residues on positions 14, 21, 22, 31 and 35. This denotes the presence of an additional 80 sidechain bonds. The Hamiltonian for 6Y1A can be seen in Figure 3.2.

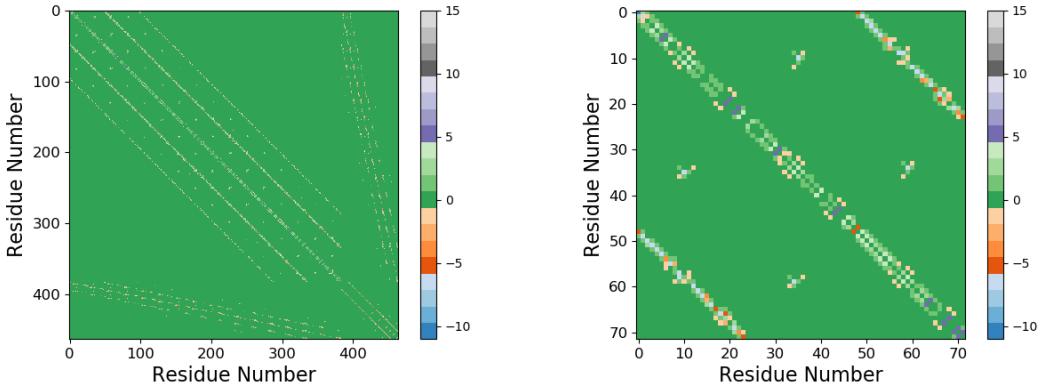


Figure 3.2. A complete Hamiltonian (left) and a Zoomed in Hamiltonian (right) of 6Y1A. Zoomed in are the first 75 residues such that the internal backbone couplings of the first three protein chains are fully visible visible.

It can be observed that the repetition in the Hamiltonian of 6Y1A matches the sixteen different protein chains that make up the amylin structure. Couplings between neighbouring residues are clearly present, showing how the structure is folded into itself. The S-shaped pattern of the amyloid fibrils can also be observed as parts where protein chains meet create small couplings between direct neighbours, these can be seen as small isolated couplings between the lines. The zoomed in image shows that residues of protein chain A couple the best with residues from protein chain C. This

is due to the way the protein is constructed from the PDB file, as these protein chains are directly on top of each other.

3.3 Two Dimensional Infrared Spectroscopy for 2PNE

After the Hamiltonians have been constructed for the proteins, we needed to go verify the results that have been generated by the NISE programme. As the main focus was to find a certain fingerprint or characteristic that could define the PPII structure within 2PNE to other secondary structures, the first step would be to check the response function to see whether the system has had enough time to relax after initial excitation.

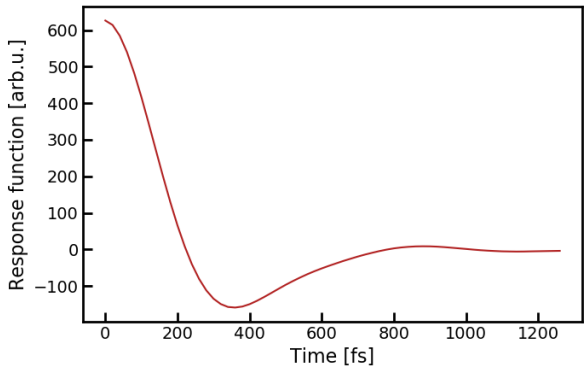


Figure 3.3. The response function of 2PNE after excitation. The graph shows that the system returns to normal within the time frame of the simulation, and thus results obtained for the (2D) absorption will not show artifacts from convolution.

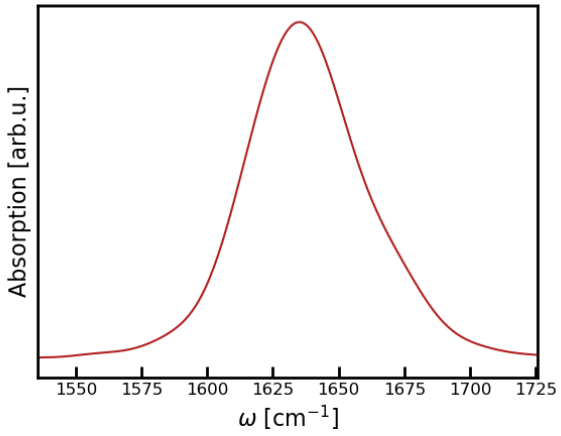


Figure 3.4. The absorption spectrum of 2PNE. The graphs shows that the system absorbs light with wavelengths within the expected values that coincide with the amide-I excitation.

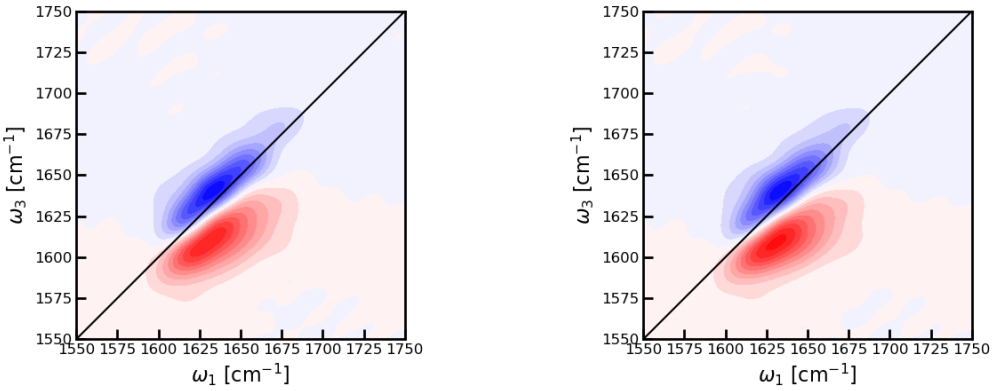


Figure 3.5. The results of 2DIR excitation of 2PNE with parallel (left) and perpendicular (right) excitation. ω_1 denotes the frequency of the pump pulse and ω_3 denotes the probe pulse.

The results of the 2DIR spectroscopy of 2PNE can be seen in Figure 3.5. The blue area indicates excited-state absorptions (increased absorption) and the red are indicates bleaching (reduces absorption) between the pump pulse(s) (ω_1,ω_2) and probe pulse (ω_3) that the system exhibits. The direction of excitation does not matter as energy is delocalized enough to pass the excitation easily across the entire protein. This causes the absence of off-diagonal peaks in the spectrum as the off-diagonal terms show interactions between the different vibrational eigenstates within the protein.

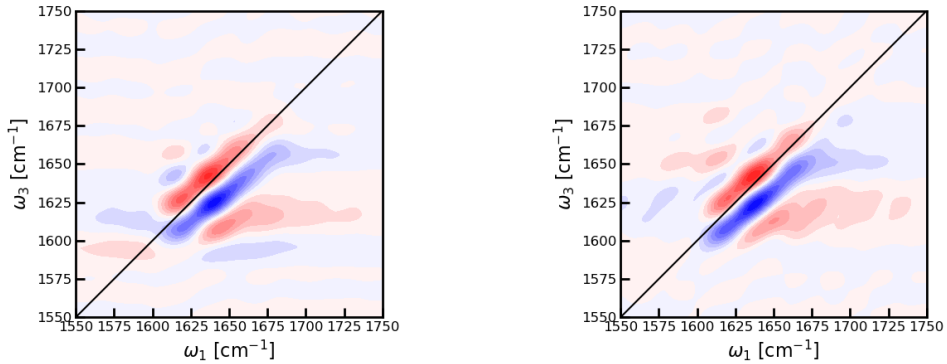


Figure 3.6. The results of 2DIR with cross polarized light on 2PNE (left) in comparison to the results of 6Y1A (right) using the same inputs.

Another way to visualize the results of 2DIR spectroscopy is with the use of cross polarized light.[96] Figure 3.6 shows a comparison between the 2DIR results from 2PNE and 6Y1A using cross polarized light. A full comparison between the 2DIR results of 2PNE and 6Y1A can be found in A.1 For 2PNE there are several distinct peaks along the ω_1 axis, namely around 1620, 1640 and 1670 cm^{-1}. Smaller peaks can be seen at different points along the diagonal axis. It is possible that these large peaks consist of multiple eigenvalues that contribute to the total absorption present. This can be verified by calculating the exciton within the system and looking at the frequencies of those excitations. One of the ways to analyze the spectra obtained with NISE is by taking the results from the parallel polarized light, and subtracting three times the value from the perpendicularly polarized light.† Doing this will give another insight into the internal couplings within the protein, as it will show different cross peaks that are not visible with normal cross-polarized light.[7]Results of the Zanni spectrum of 2PNE are given in Figure 3.7.

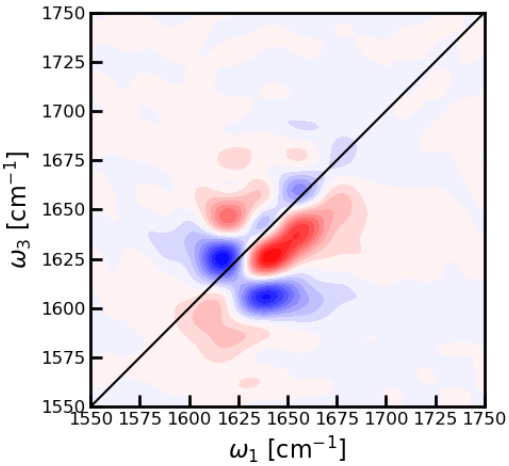


Figure 3.7. The 2DIR spectra visualized using the Zanni method.[7] cross peaks that were present in Figure 3.6 can still observed. Extra off-diagonal transitions may be indications of multiple eigenvalues contributing to the absorption spectrum at specific wavelengths.

†Subsequently coined the Zanni method, as no other name has yet been chosen for this quite common method of analyzing 2DIR data.

3.3.1 Exciton Analysis

One of ways to look at the behaviour of a protein during 2DIR is to look at how the eigenfrequencies of a protein behave during excitation. As there are certain couplings within the protein that have a maximum allowance of vibrational delocalization. By finding and visualizing these, it is possible to understand how the couplings within the protein help form and shape it. The script used to visualize the eigenfrequencies was developed by T.L.C. Jansen and can be found on Github.[74,97]

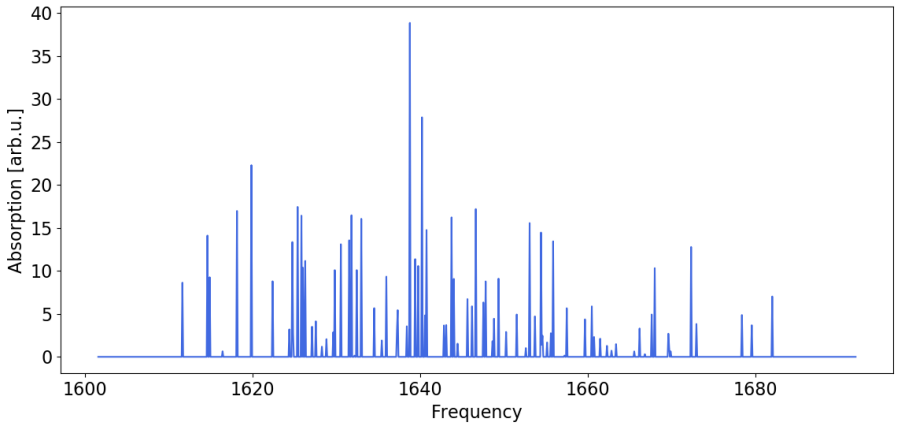


Figure 3.8. (The relative strength of absorptions present within 2PNE as a function of the frequency of light.

In order to accurately simulate the transfer of excitation across different frequencies after excitation, a Gaussian is applied to the results from Figure 3.8 to end up with Figure 3.9. This figure also shows how the initial peaks, that can be observed in Figure 3.6 and Figure 3.5 are indeed a result of multiple different eigenfrequencies interacting.

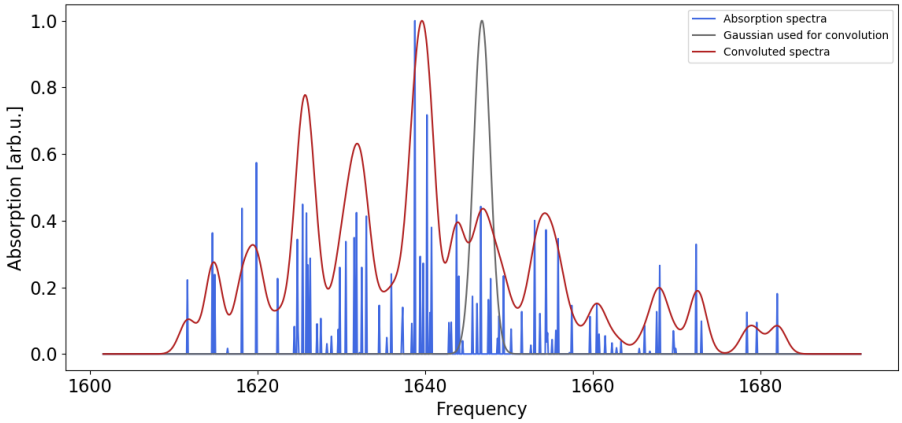


Figure 3.9. An overlap of the stick spectra and the Gaussian convolution where absorption is given as a relative value to the maximum absorption measured.

It should be noted that in order to visualize the data this way, the average Hamiltonian and dipole moments of the entire simulation are taken, this is discussed further in Chapter 4. Using a Hamiltonian that represents the average energy and dipole values of the protein during the entire simulation, it is possible to see which eigenstates contribute the most to the internal dynamics of the protein. Such an eigenstate describes the vibrational coupling of all amide I vibrations present in the protein. Following the results seen in Figure 3.8 for 2PNE we have chosen a threshold of 20 or greater for frequencies in the range of 1600 to 1650 cm^{-1}, but as there are clear peaks present in the 2DIR results in Figure 3.6 and Figure 3.7 at higher frequencies, the strongest absorptions between 1660 and 1700 cm^{-1} are also examined. The results can be seen in Table 3.1 and a visual representation of the direction of amide I couplings and oscillations due to excitation can be found in Figure A.4.

Exciton Eigenstates with value and transition dipole direction					
Eigenstate number	Oscillator strength	Frequency (cm^{-1})	X direction	Y direction	Z direction
5	22.31	1619.84	-0.56	-0.10	4.69
32	38.87	1638.78	-3.35	-3.22	-4.15
35	27.88	1640.17	3.68	2.82	-2.50
75	10.32	1667.91	-2.17	-1.22	2.03
79	12.79	1672.31	-2.97	-1.97	-0.32

Table 3.1. Oscillator strength, Frequency of oscillation, and Cartesian directions of the largest absorptions measured in the 2DIR spectrum of 2PNE.

From the data it can be seen that there are five prominent eigenstates that contribute the most to the behaviour of the protein. Eigenstate 5 shows that the oscillation is predominantly in the Z direction, which helps orientate the protein in Figure A.4. Other eigenfrequencies do not show a clear direction, which can be attributed to the lack of symmetry in the 2PNE protein. If we look at Figure 1.3 the full rotation every three residues, that was expected from PPII, can indeed be seen in the directions of the couplings within 2PNE in Figure A.4. A section of nine residues has been selected to highlight this three repeat in Figure A.3. If all the strong eigenstates were indeed the results of strong couplings within the PPII structure, the directions of oscillation would be more uniform, this is unfortunately not the case for 2PNE.

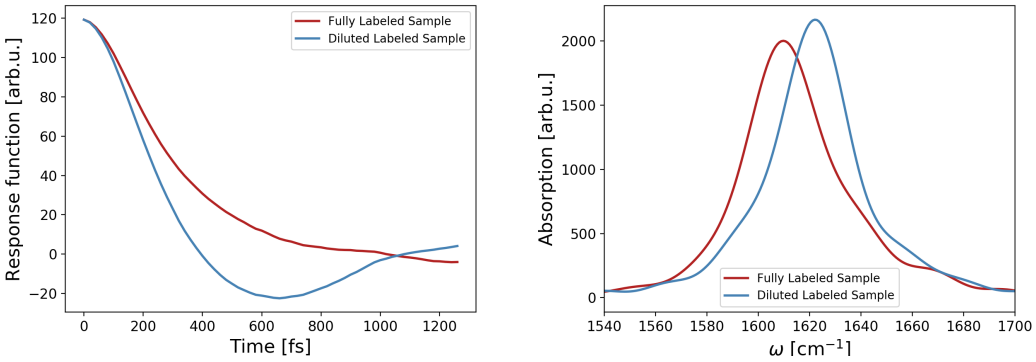
Although differences between the 2DIR spectra of 2PNE and 6Y1A can be found, it is yet unclear if the results obtained from 2PNE are purely due to the PPII structure. As such, unfortunately no characteristics could be determined for PPII yet. The structure is indeed organized and shows some clear peaks and off diagonal elements, yet there are no strong couplings visible that could be used to predict the presence of PPII in other proteins as is the case for α-helices and β-sheets. This could be do to the fact that there are no strong internal hydrogen bondings inside PPII helices, but further research is necessary to confidently establish this.

3.4 Isotopic Labeling for 6Y1A

When looking at the data from the isotopic labeling it is important to verify that the isotopic labeling exhibits couplings that could be predicted from the complete Hamiltonian. To check the validity of the obtained NISE spectra it is important to check whether the individual couplings, that have been observed in the complete Hamiltonian of the protein, are visible in the smaller Hamiltonian as well. (Figure 3.11) As isotopic labeling within NISE is simply done by isolating the necessary residues from the complete Hamiltonian. All Hamiltonians obtained for the isotopic couplings can be found in section A.3 where the couplings between residues from different protein chains can clearly be seen.

When simulating the behaviour of the protein during 2DIR it is possible to increase the so-called coupling cut. Which is the distance at which the couplings between residues is neglected. By increasing the coupling cut we can simulate the behaviour of the model as if the protein is diluted. By diluting the sample the couplings between neighbours are less likely to occur and assumed to be localized, as the isotopically labeled residues are more likely to vibrate independently. This difference between a "normal" state and the labeled state can show how the individual residues contribute to the overall absorption spectrum. The response function and absorption spectrum for isotopically labeled Ala13, showing the shift in absorption between a diluted and non-diluted sample, can be see in Figure 3.10.

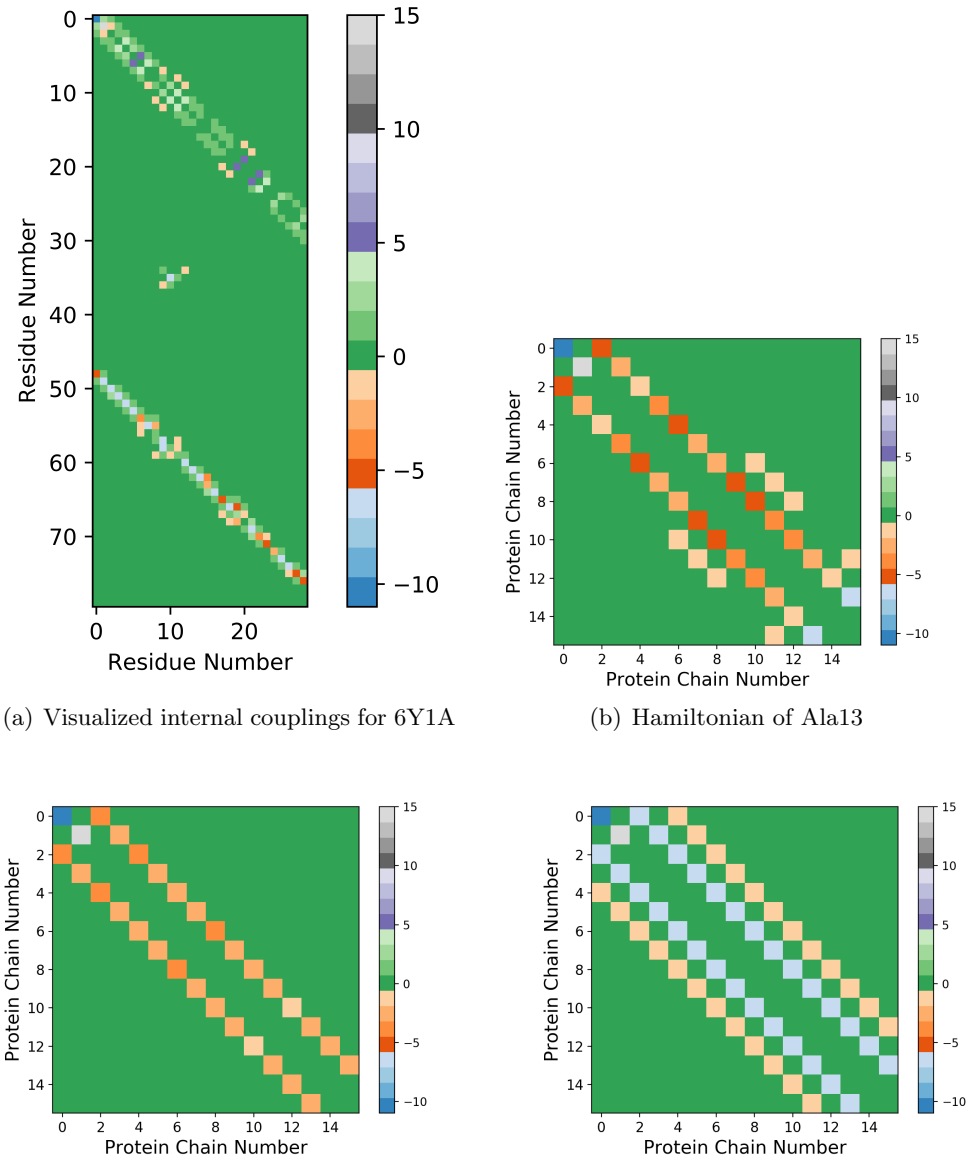
Unlike the research done by Skinner *et al.* we are not confined by the selection of residues, as we are doing experiments *in silico*, and as such it is possible to label virtually all residues within the protein.[79] For this reason all the residues that have been discussed in section 1.2.3 have been isotopically labelled to see the individual behaviour of the residues within the protein.



(a) Response Function for Ala13

(b) Absorption Function for Ala13

Figure 3.10. The Response function and the Absorption spectrum obtained for the isotopic labeling of Ala13 using the NISE programme. The absorption is shifted 12 cm^{-1} which is in slightly larger than the 10 cm^{-1} shift found by Wang *et al.*[8]



(a) Visualized internal couplings for 6Y1A

(b) Hamiltonian of Ala13

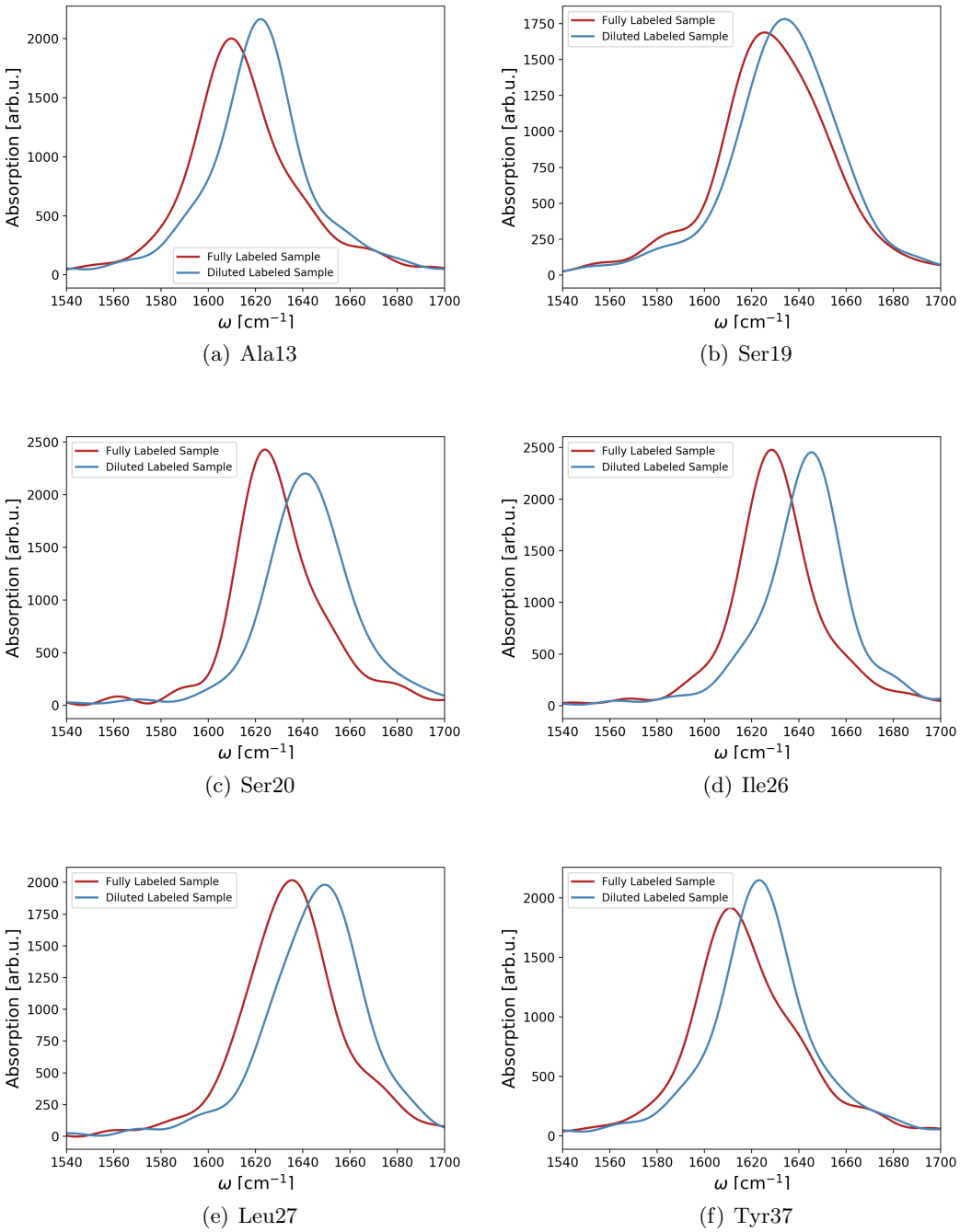
(c) Hamiltonian of Ser19

(d) Hamiltonian of Ala25

Figure 3.11. A selection of the coloured Hamiltonians, Ala13, Ser19 and Ala25, which show the internal couplings between the residues present in the 6Y1A protein. Shown is how the couplings, between a protein chain and its second neighbour, as seen in Figure 3.2 are also present in the Hamiltonians of the isotopically labeled proteins.

3.4.1 Absorption Spectra of the Isotopically Labeled Residues

After the calculations were performed within the NISE software the spectra of the diluted and undiluted isotopically labeled residues could be compared. First the residues that exhibited a redshift in the isotopically labeled spectrum when the samples are coupled together are examined. (Figure 3.12)



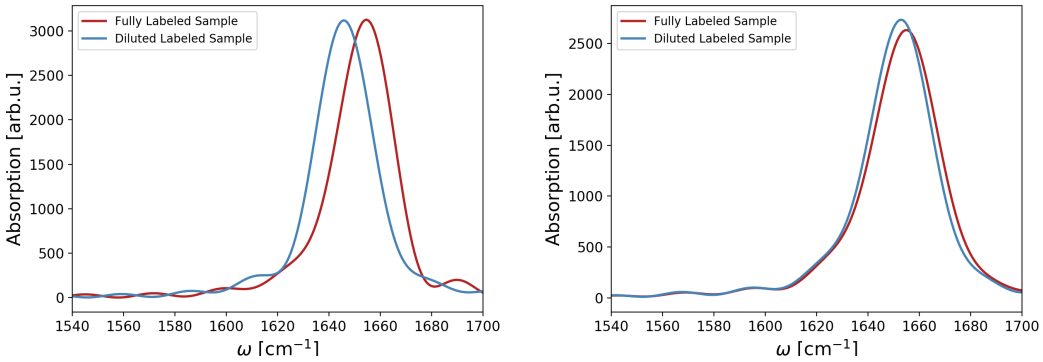
(a) Ala13 (b) Ser19

(c) Ser20 (d) Ile26

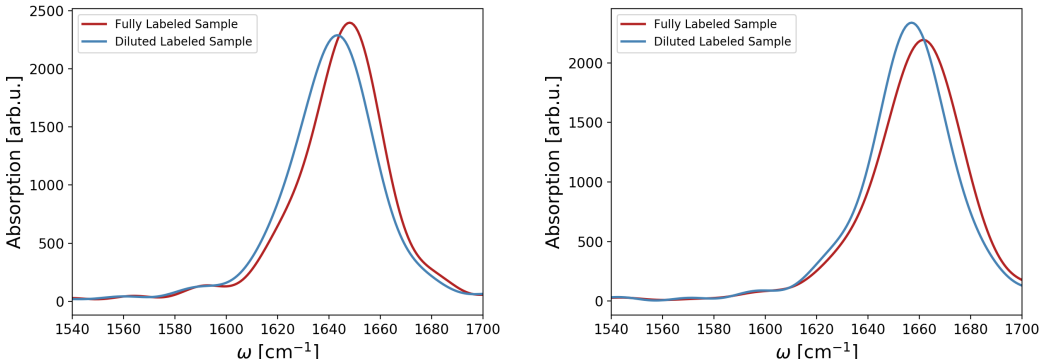
(e) Leu27 (f) Tyr37

Figure 3.12. Absorption spectra of several isotopically labeled samples. The graphs show minimal difference except the expected redshift of the spectrum. All the residues shown here have a negative coupling, which coincides with the coupling J_{ij} in Equation 2.2

The following residues show a blueshift in fully labeled samples. Meaning that the dilution caused the spectra to absorb at a lower wavelength than when the residues are coupled. (Figure 3.13)

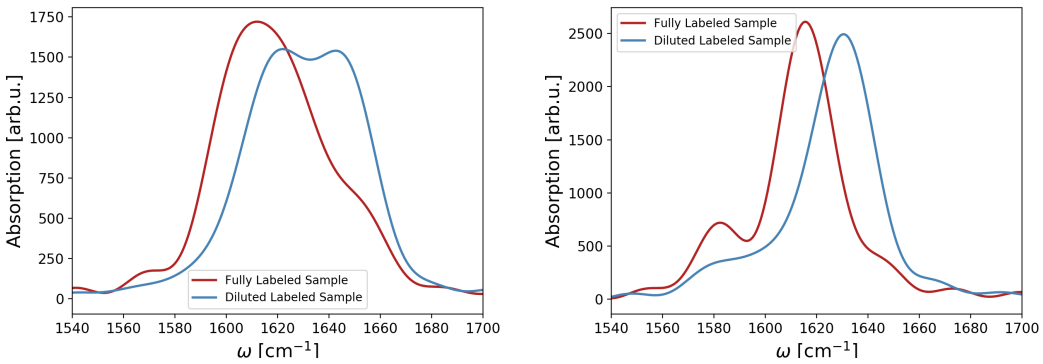


(a) Phe23 (b) Gly24



(c) Val32 (d) Gly33

Figure 3.13. The absorption spectra of isotopically labeled where the graphs show a blueshift. Which corresponds with a positive coupling J_{ij}.



(a) Leu16 (b) Ala25

Figure 3.14. Absorption spectra of isotopically labeled samples that show a peculiarity next to the expected shift in peak position.

Some absorption spectra showed extra peculiarities between the normal and diluted samples. (Figure 3.14) Leu16 showed two peaks in the diluted case, meaning that the residue has two frequencies where it can dissipate the excitation across the protein. Ala25 shows the reverse results, and exhibits two absorption peaks when the system is coupled together. The results of the absorption spectrum shifting, as well as the strength of the coupling between the residues can be found in Table 3.2

3.4.2 Overall Results

If we compare the data obtained from the isotopic labeling with the measured couplings and the shift in absorption, the link between the direction of the coupling and the direction of the absorption shifting, can immediately be connected. The narrowing of the absorption spectrum in the fully labeled sample could be attributed to exchange narrowing, intensity borrowing, or the contribution of multiple eigenstates to the new absorption spectrum.[98,99]

Absorption Shift and Coupling Strength				
Residue	Peak fully labeled sample	Peak diluted labeled sample	Relative Shift (cm^{-1})	Coupling (cm^{-1})
Ala13	1609.77	1621.99	12.22	-3.832
Leu16	1612.22	1621.98†	9.76	-5.906
Ser19	1625.25	1634.20	8.95	-2.640
Ser20	1623.62	1640.72	17.10	-6.511
Phe23	1654.56	1645.60	-8.96	2.214
Gly24	1654.56	1652.93	-1.63	1.121
Ala25	1615.47	1630.13	14.66	-6.446
Ile26	1628.50	1645.60	17.40	-6.370
Leu27	1635.83	1649.68	14.03	-5.744
Val32	1648.05	1643.16	-4.89	1.961
Gly33	1661.89	1657.01	-4.88	1.718
Tyr37	1611.40	1622.80	11.40	-3.435

Table 3.2. Peak positions of fully labeled and diluted labeled samples, and their relative shift in the position of their maximum absorption peak. † For Leu16 Two peaks are observed, the peak of the strongest coupling is used as a marker

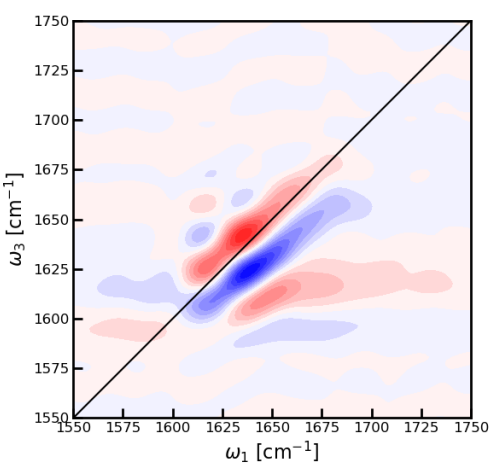
It can be seen that if the coupling present in the residue is of a positive value, then the labeled graph moves towards a higher range in the wave function spectrum. And it is reverse if the coupling is negative. Most residues follow the trend that the relative shift is between negative three and negative five times the coupling strength, with some notable exceptions. Skinner *et al.* found that the couplings in the fully labeled samples could result in the absorption spectrum moving by as much as 10 cm^{-1}, and this is also found by the isotopic labeling performed *in silico*.[8] Yet, some labels show an even greater alteration. We can state that initial results show a greater and more varying degree of absorption shift. It is quite possible that the increases magnitude of observed shifts after isotopic labeling *in silico* is due to relative scaling present in the system. The FGAIL sequence behaves similarly to the earlier research on the amylin structure with relatively strong coupling between Ila25, Ile26 and Leu27. (Table A.1) Initial results show that the 6Y1A is indeed a probable configuration of the amylin amyloid, but *in vitro* experimentation will need to be performed to verify these claims.[8,44]

Chapter 4

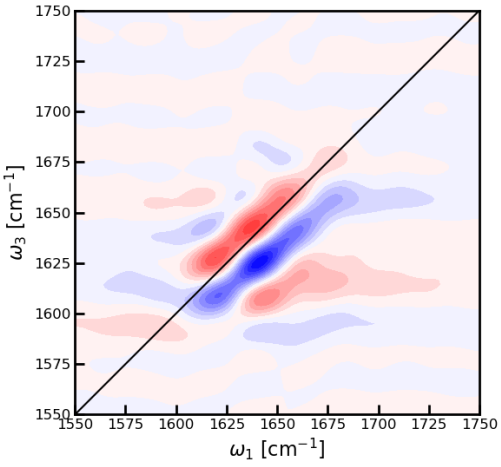
Discussion

One point of discussion is whether the usage of *in silico* simulations to predict the behaviour of, or complex process involving, protein dynamics is valid as a comparison to *in vitro* experiments. *In silico* experiments have long been used to prove or enhance the results obtained from *in vitro* experiments.[100] And, it has been found that some *in silico* experiments can be more efficient at predicting the behaviour of molecules, than similar experiments performed on animals.[101] Unfortunately, in some prominent fields of research, *in silico* experiments lack the accuracy needed to accurately predict results.[102] The most important factor that determines the accuracy of *in silico* experimentation are the specific computational approaches that have been used to obtain the results.[103] As was specified in chapter 2 the Amide I mapping software and NISE package have been proven to be relatively accurate as they do nothing more than apply computationally expensive calculations on the trajectories of the atoms provided by MD simulations. The most important aspect, in regards to the accuracy of our data, is the correctness of the MD simulations performed. Even though the version of Gromacs used for MD simulations is not up to date, the accuracy of MD simulations depends almost completely on the completeness of the chosen Force Field.[104] In our case that would be the OPLS-AA force field, which is based upon ab initio calculations and experimental data to maximize its validity.[75,76] As *in silico* experiments will become more and more common for researchers as an alternative to expensive and time consuming *in vitro* experimentation it is important to constantly re-asses the validity of the systems and models used.[105] Most, if not all, of the models used during *in silico* experimentation are derived from previously acquired data, meaning that as long as the ways in which the models are build, be they ab initio, experimental, or quantum mechanically. If it is discovered that one of the assumptions are faulty, then all subsequent usage of this data is also wrong. Fortunately, the accuracy and validity of the models used in this research has been proven to be sufficient enough to predict, or accompany, *in vitro* experimentation.[93,106]

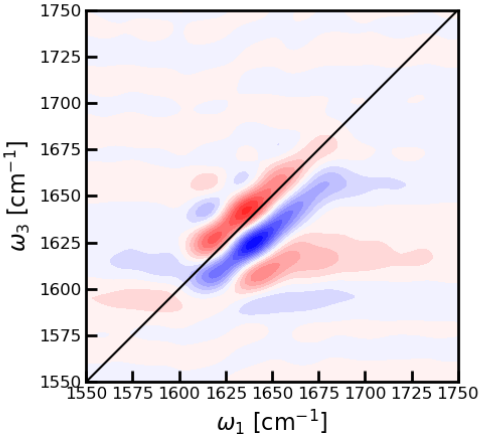
Another point of discussion could be the usage of the averages for the Hamiltonian and dipoles for calculations necessary to get the 2DIR results. As the proteins are started from man-made static configurations in the form of a PDB file, it is quite possible that sudden shifts during initial MD simulations can cause alterations in the Hamiltonian or dipole files obtained with the amide-I mapping. One way to verify the integrity of the data is by checking whether the system does indeed change spontaneously during simulation. This is done by separating the total duration of the simulation into two equal parts. This can easily be done within the NISE programme by stating that the numerical calculations start and stop at different points. If the first half of the MD simulation is compared against the second half, minimal differences should be observed. Differences between these halves would indicate that the protein needed to find an thermodynamic minimum before necessary couplings be established. The cross-related 2DIR results of the first and second half of the total run time of 2PNE and 6Y1A are plotted in Figure 4.1 and Figure 4.2.



(a) 2PNE First half

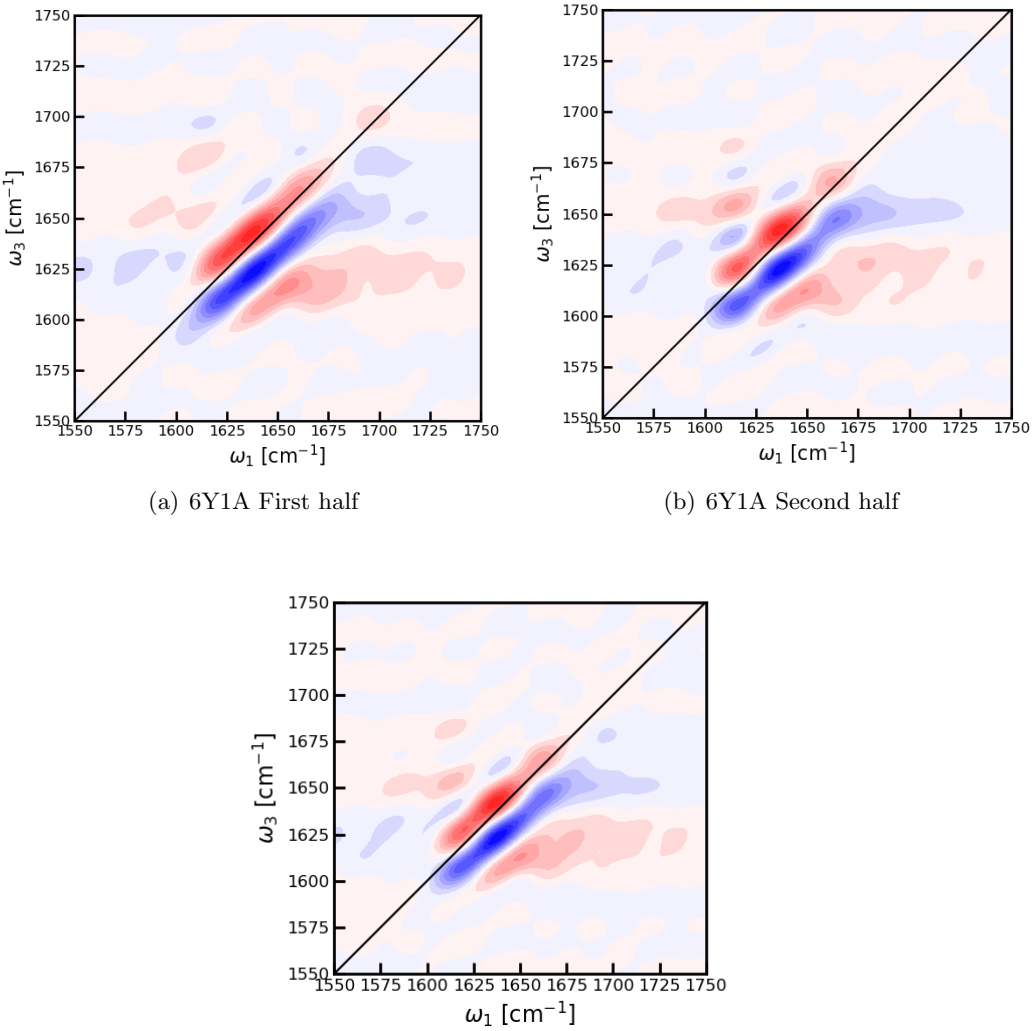


(b) 2PNE Second half



(c) 2PNE complete

Figure 4.1. The 2DIR results first and Second half of the 2PNE simulation in comparison to results obtained from the average Hamiltonian. All the images are constructed with cross polarized light.



(a) 6Y1A First half

(b) 6Y1A Second half

(c) 6Y1A complete

Figure 4.2. The 2DIR results first and Second half of the 6Y1A simulation in comparison to results obtained from the average Hamiltonian

As can be seen in Figure 4.1 and Figure 4.2; 2PNE shows little to no difference between the visible peaks in the spectrum. What can be observed is how it takes time for the internal couplings in 6Y1A to manifest themselves clearly. This can be seen as by the appearances of off-diagonal couplings in the second half, where excitations can traverse the protein better as couplings between residues are better. As the MD simulation progresses the residues within the protein will more more closely as they slowly move to their final energetically favourable position. Luckily it can be stated that as the protein itself does not change during simulation, and the wavelengths of the absorption peaks stay the same, that the transition dipoles used for calculations, where the average value of the entire run are necessary, are still accurate enough to predict *in vitro* behaviour of the proteins.

Chapter 5

Conclusion

Protein aggregates are found as a result of several degenerative genetic diseases. It is unclear whether these clustering of faulty proteins are a symptom of the disease, or the source. Two-Dimensional Infrared Spectroscopy has been chosen to examine such protein conformations, as it is one of the few methods available that can resolve structures within the time frame that is necessary for protein-protein interactions to be examined, as well as the fact it can be applied to proteins in solution. Hamiltonians containing the information about internal couplings of the amide I vibrational mode have been constructed from trajectories obtained by MD simulations, which have subsequently be analyzed using the Numerical Integration of the Schrödinger Equation (NISE).

The snow flea antifreeze protein (2PNE) consists of three pairs of anti-parallel PPII helices. These PPII helices have been found to be involved in the cytotoxicity of Poly-Q amyloid clusters, although the exact nature of this is still under investigation. 6Y1A is a proposed configuration of the amylin amyloid found in the pancreas during diabetes mellitus type 2. In order to verify this configuration, isotopic labeling has been applied to 6Y1A to investigate whether the obtained absorption spectra and couplings agree with proposed hypothesis and experimental results.

No direct identifiable fingerprint could be discovered for the PPII structure. As is common with the other more prevalent secondary structures, the α-helix and the β-sheet, internal hydrogen bondings cause the transfer of excitation to show in the absorption spectrum of 2DIR. It is possible that characteristics absorption spectra the PPII structure are present within the protein, but they are not clearly visible within 2PNE using current settings. Isotopic labeling on the 6Y1A configuration was performed in order to verify speculations about the possible internal structures. This was done by repeating isotopic labeling performed by Wang *et al.*[8] and examining the relationship between observed absorption spectra and predicted behaviour due to residue proximity. 6Y1A was indeed deemed a possible configuration of the amylin amyloid, but further *in vitro* experimentation will be necessary to definitively conclude the validity of the 6Y1A protein.

The validity of using *in silico* experiments to verify *in vitro* results is discussed, and it was concluded that the methods involved in this research have al been proven accurate enough to simulate real physical processes. *In silico* experimentation could be used to verify or further examine *in vitro* experiments, as the lower costs and extra variability of the method allow for different aspects of the protein interactions to be examined. Although no clear results have been obtained, it is my hope that further research is possible from here, where more knowledge about the formation of protein aggregates will lead to a way to stop them from forming altogether.

Bibliography

[1] Wikipedia, the free encyclopedia, "Dihedral angle," Dec 2018. [Online; accessed May 22, 2020].

[2] N. A. Campbell and J. B. Reece, *Biology*. Pearson Education India, 2008.

[3] "Polyproline helix," May 2020.

[4] H.-K. Lin, J. C. Boatz, I. E. Krabbendam, R. Kodali, Z. Hou, R. Wetzel, A. M. Dolga, M. A. Poirier, and P. C. van der Wel, "Fibril polymorphism affects immobilized non-amyloid flanking domains of huntingtin exon1 rather than its polyglutamine core," *Nature communications*, vol. 8, no. 1, pp. 1–12, 2017.

[5] Q. Cao, D. R. Boyer, M. R. Sawaya, P. Ge, and D. S. Eisenberg, "Cryo-em structure and inhibitor design of human iapp (amylin) fibrils," *bioRxiv*, 2020.

[6] T. l. C. Jansen and J. Knoester, "Waiting time dynamics in two-dimensional infrared spectroscopy," *Accounts of chemical research*, vol. 42, no. 9, pp. 1405–1411, 2009.

[7] P. Hamm and M. Zanni, *Concepts and Methods of 2D Infrared Spectroscopy*. Cambridge University Press, 2011.

[8] L. Wang, C. T. Middleton, S. Singh, A. S. Reddy, A. M. Woys, D. B. Strasfeld, P. Marek, D. P. Raleigh, J. J. de Pablo, M. T. Zanni, and J. L. Skinner, "2dir spectroscopy of human amylin fibrils reflects stable β-sheet structure," *Journal of the American Chemical Society*, vol. 133, no. 40, pp. 16062–16071, 2011. PMID: 21916515.

[9] S. C. L. Kamerlin and A. Warshel, "At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis?," *Proteins: Structure, Function, and Bioinformatics*, vol. 78, no. 6, pp. 1339–1375, 2010.

[10] D. R. Davies and G. H. Cohen, "Interactions of protein antigens with antibodies," *Proceedings of the National Academy of Sciences*, vol. 93, no. 1, pp. 7–12, 1996.

[11] M. Ramírez-Alvarado, J. S. Merkel, and L. Regan, "A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro," *Proceedings of the National Academy of Sciences*, vol. 97, no. 16, pp. 8979–8984, 2000.

[12] S. B. Prusiner, M. R. Scott, S. J. DeArmond, and F. E. Cohen, "Prion protein biology," *Cell*, vol. 93, no. 3, pp. 337–348, 1998.

[13] F. Chiti and C. M. Dobson, "Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade," *Annual review of biochemistry*, vol. 86, pp. 27–68, 2017.

[14] B. Kagan, R. Azimov, and R. Azimova, "Amyloid peptide channels," *The Journal of membrane biology*, vol. 202, no. 1, pp. 1–10, 2004.

[15] M. D. Benson, J. N. Buxbaum, D. S. Eisenberg, G. Merlini, M. J. M. Saraiva, Y. Sekijima, J. D. Sipe, and P. Westermark, "Amyloid nomenclature 2018: recommendations by the international society of amyloidosis (isa) nomenclature committee," *Amyloid*, vol. 25, no. 4, pp. 215–219, 2018. PMID: 30614283.

[16] K. A. Conway, J. D. Harper, and P. T. Lansbury, "Fibrils formed in vitro from α-synuclein and two mutant forms linked to parkinson's disease are typical amyloid," *Biochemistry*, vol. 39, no. 10, pp. 2552–2563, 2000.

[17] E. Scherzinger, A. Sittler, K. Schweiger, V. Heiser, R. Lurz, R. Hasenbank, G. P. Bates, H. Lehrach, and E. E. Wanker, "Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for huntington's disease pathology," *Proceedings of the National Academy of Sciences*, vol. 96, no. 8, pp. 4604–4609, 1999.

[18] R. O'Hara, E. P. Murphy, A. S. Whitehead, O. FitzGerald, and B. Bresnihan, "Acute-phase serum amyloid a production by rheumatoid arthritis synovial tissue," *Arthritis Research & Therapy*, vol. 2, no. 2, p. 142, 2000.

[19] P. Westermark, A. Andersson, and G. T. Westermark, "Islet amyloid polypeptide, islet amyloid, and diabetes mellitus," *Physiological reviews*, vol. 91, no. 3, pp. 795–826, 2011.

[20] R. N. Rambaran and L. C. Serpell, "Amyloid fibrils: abnormal protein assembly," *Prion*, vol. 2, no. 3, pp. 112–117, 2008.

[21] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers: Original Research on Biomolecules*, vol. 22, no. 12, pp. 2577–2637, 1983.

[22] D. Eisenberg, "The discovery of the α-helix and β-sheet, the principal structural features of proteins," *Proceedings of the National Academy of Sciences*, vol. 100, no. 20, pp. 11207–11210, 2003.

[23] A. A. Adzhubei and M. J. Sternberg, "Left-handed polyproline ii helices commonly occur in globular proteins," *Journal of molecular biology*, vol. 229, no. 2, pp. 472–493, 1993.

[24] M. Cubellis, F. Caillez, T. Blundell, and S. Lovell, "Properties of polyproline ii, a secondary structure element implicated in protein–protein interactions," *Proteins: Structure, Function, and Bioinformatics*, vol. 58, no. 4, pp. 880–892, 2005.

[25] A. A. Adzhubei, M. J. Sternberg, and A. A. Makarov, "Polyproline-ii helix in proteins: Structure and function," *Journal of Molecular Biology*, vol. 425, no. 12, pp. 2100 – 2132, 2013.

[26] V. V. Lakhani, F. Ding, and N. V. Dokholyan, "Polyglutamine induced misfolding of huntingtin exon1 is modulated by the flanking sequences," *PLoS computational biology*, vol. 6, no. 4, 2010.

[27] S. Treusch, D. M. Cyr, and S. Lindquist, "Amyloid deposits: protection against toxic protein species?," *Cell cycle*, vol. 8, no. 11, pp. 1668–1674, 2009.

[28] 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, "Structural motif of polyglutamine amyloid fibrils discerned with mixed-isotope infrared spectroscopy," *Proceedings of the National Academy of Sciences*, vol. 111, no. 16, pp. 5796–5801, 2014.

[29] D. Cohen-Carmon and E. Meshorer, "Polyglutamine (polyq) disorders: the chromatin connection," *Nucleus*, vol. 3, no. 5, pp. 433–441, 2012.

[30] E. J. Calabrese, T. N. Bhatia, V. Calabrese, G. Dhawan, J. Giordano, Y. N. Hanekamp, R. Kapoor, W. J. Kozumbo, and R. K. Leak, "Cytotoxicity models of huntington's disease and relevance of hormetic mechanisms: A critical assessment of experimental approaches and strategies," *Pharmacological research*, 2019.

[31] H. Kadowaki, H. Nishitoh, F. Urano, C. Sadamitsu, A. Matsuzawa, K. Takeda, H. Masutani, J. Yodoi, Y. Urano, T. Nagano, and et al., "Amyloid β induces neuronal cell death through ros-mediated ask1 activation," *Cell Death & Differentiation*, vol. 12, p. 19–24, Sep 2004.

[32] B. Zott, M. M. Simon, W. Hong, F. Unger, H.-J. Chen-Engerer, M. P. Frosch, B. Sakmann, D. M. Walsh, and A. Konnerth, "A vicious cycle of β amyloid–dependent neuronal hyperactivation," *Science*, vol. 365, no. 6453, pp. 559–565, 2019.

[33] H. Choi, C. Kim, H. Song, M.-Y. Cha, H. J. Cho, S. M. Son, H. J. Kim, and I. Mook-Jung, "Amyloid β-induced elevation of o-glcnacylated c-fos promotes neuronal cell death," *Aging Cell*, vol. 18, no. 1, p. e12872, 2019.

[34] M. F. M. Engel, L. Khemtémourian, C. C. Kleijer, H. J. D. Meeldijk, J. Jacobs, A. J. Verkleij, B. de Kruijff, J. A. Killian, and J. W. M. Höppener, "Membrane damage by human islet amyloid polypeptide through fibril growth at the membrane," *Proceedings of the National Academy of Sciences*, vol. 105, no. 16, pp. 6033–6038, 2008.

[35] S. A. Jayasinghe and R. Langen, "Membrane interaction of islet amyloid polypeptide," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1768, no. 8, pp. 2002 – 2009, 2007. Amyloidogenic Protein–Membrane Interaction.

[36] B. L. Pentelute, Z. P. Gates, V. Tereshko, J. L. Dashnau, J. M. Vanderkooi, A. A. Kossiakoff, and S. B. Kent, "X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers," *Journal of the American Chemical Society*, vol. 130, no. 30, pp. 9695–9701, 2008.

[37] F.-H. Lin, L. A. Graham, R. L. Campbell, and P. L. Davies, "Structural modeling of snow flea antifreeze protein," *Biophysical journal*, vol. 92, no. 5, pp. 1717–1723, 2007.

[38] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, pp. 235–242, 01 2000.

[39] F. Totzeck, M. A. Andrade-Navarro, and P. Mier, "The protein structure context of polyq regions," *PloS one*, vol. 12, no. 1, 2017.

[40] J. R. Brender, S. Salamekh, and A. Ramamoorthy, "Membrane disruption and early events in the aggregation of the diabetes related peptide iapp from a molecular perspective," *Accounts of chemical research*, vol. 45, no. 3, pp. 454–462, 2012.

[41] M. R. Hayden and S. C. Tyagi, "A" is for amylin and amyloid in type 2 diabetes mellitus," *Jop*, vol. 2, no. 4, pp. 124–139, 2001.

[42] L. Marzban, K. Park, and C. B. Verchere, "Islet amyloid polypeptide and type 2 diabetes," *Experimental gerontology*, vol. 38, no. 4, pp. 347–351, 2003.

[43] C. Röder, T. Kupreichyk, L. Gremer, L. U. Schäfer, K. R. Pothula, R. B. Ravelli, D. Willbold, W. Hoyer, and G. F. Schröder, "Amyloid fibril structure of islet amyloid polypeptide by cryo-electron microscopy reveals similarities with amyloid beta," *bioRxiv*, 2020.

[44] A. L. Serrano, J. P. Lomont, L.-H. Tu, D. P. Raleigh, and M. T. Zanni, "A free energy barrier caused by the refolding of an oligomeric intermediate controls the lag time of amyloid formation by hiapp," *Journal of the American Chemical Society*, vol. 139, no. 46, pp. 16748–16758, 2017.

[45] G. G. Tartaglia, A. Cavalli, R. Pellarin, and A. Caflisch, "Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences," *Protein Science*, vol. 14, no. 10, pp. 2723–2734, 2005.

[46] J. Janson, T. Laedtke, J. E. Parisi, P. O'Brien, R. C. Petersen, and P. C. Butler, "Increased risk of type 2 diabetes in alzheimer disease," *Diabetes*, vol. 53, no. 2, pp. 474–481, 2004.

[47] S. Rosales-Corral, D.-X. Tan, L. Manchester, and R. J. Reiter, "Diabetes and alzheimer disease, two overlapping pathologies with the same background: oxidative stress," *Oxidative medicine and cellular longevity*, vol. 2015, 2015.

[48] W. Q. Qiu, M. Wallack, M. Dean, E. Liebson, M. Mwamburi, and H. Zhu, "Association between amylin and amyloid-β peptides in plasma in the context of apolipoprotein e4 allele," *PloS one*, vol. 9, no. 2, 2014.

[49] J. Gotz, Y.-A. Lim, and A. Eckert, "Lessons from two prevalent amyloidoses—what amylin and aβ have in common," *Frontiers in aging neuroscience*, vol. 5, p. 38, 2013.

[50] L. M. Luheshi, G. G. Tartaglia, A.-C. Brorsson, A. P. Pawar, I. E. Watson, F. Chiti, M. Vendruscolo, D. A. Lomas, C. M. Dobson, and D. C. Crowther, "Systematic in vivo analysis of the intrinsic determinants of amyloid β pathogenicity," *PLoS biology*, vol. 5, no. 11, 2007.

[51] G. E. Schulz, "Jan drenth: Principles of protein x-ray crystallography, second edition," *Crystal Research and Technology*, vol. 34, no. 9, pp. 1120–1120, 1999.

[52] P. Hamm, M. Lim, and R. M. Hochstrasser, "Structure of the amide i band of peptides measured by femtosecond nonlinear-infrared spectroscopy," *The Journal of Physical Chemistry B*, vol. 102, no. 31, pp. 6123–6138, 1998.

[53] H. Kim and M. Cho, "Infrared probes for studying the structure and dynamics of biomolecules," *Chemical reviews*, vol. 113, no. 8, pp. 5817–5847, 2013.

[54] S. Krimm and J. Bandekar, "Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins," in *Advances in protein chemistry*, vol. 38, pp. 181–364, Elsevier, 1986.

[55] L. P. DeFlores, Z. Ganim, R. A. Nicodemus, and A. Tokmakoff, "Amide i- ii 2d ir spectroscopy provides enhanced protein secondary structural sensitivity," *Journal of the American Chemical Society*, vol. 131, no. 9, pp. 3385–3391, 2009.

[56] Z. Ganim, H. S. Chung, A. W. Smith, L. P. DeFlores, K. C. Jones, and A. Tokmakoff, "Amide i two-dimensional infrared spectroscopy of proteins," *Accounts of chemical research*, vol. 41, no. 3, pp. 432–441, 2008.

[57] C. R. Baiz, M. Reppert, and A. Tokmakoff, "Amide i two-dimensional infrared spectroscopy: Methods for visualizing the vibrational structure of large proteins," *The Journal of Physical Chemistry A*, vol. 117, no. 29, pp. 5955–5961, 2013.

[58] C. R. Baiz, C. S. Peng, M. E. Reppert, K. C. Jones, and A. Tokmakoff, "Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution," *Analyst*, vol. 137, no. 8, pp. 1793–1799, 2012.

[59] A. Ghosh, J. S. Ostrander, and M. T. Zanni, "Watching proteins wiggle: Mapping structures with two-dimensional infrared spectroscopy," *Chemical reviews*, vol. 117, no. 16, pp. 10726–10759, 2017.

[60] A. M. Woys, Y.-S. Lin, A. S. Reddy, W. Xiong, J. J. de Pablo, J. L. Skinner, and M. T. Zanni, "2d ir line shapes probe ovispirin peptide conformation and depth in lipid bilayers," *Journal of the American Chemical Society*, vol. 132, no. 8, pp. 2832–2838, 2010.

[61] C. Falvo, W. Zhuang, Y. S. Kim, P. H. Axelsen, R. M. Hochstrasser, and S. Mukamel, "Frequency distribution of the amide-i vibration sorted by residues in amyloid fibrils revealed by 2d-ir measurements and simulations," *The Journal of Physical Chemistry B*, vol. 116, no. 10, pp. 3322–3330, 2012. PMID: 22338639.

[62] S.-H. Shim, R. Gupta, Y. L. Ling, D. B. Strasfeld, D. P. Raleigh, and M. T. Zanni, "Two-dimensional ir spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution," *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6614–6619, 2009.

[63] B. Ding, J. E. Laaser, Y. Liu, P. Wang, M. T. Zanni, and Z. Chen, "Site-specific orientation of an α-helical peptide ovispirin-1 from isotope-labeled sfg spectroscopy," *The Journal of Physical Chemistry B*, vol. 117, no. 47, pp. 14625–14634, 2013.

[64] T. L. C. Jansen, "Nise." https://github.com/GHlacour/NISE_2017, 2017.

[65] A. R. Leach, *Molecular modelling: principles and applications*. Pearson/Prentice Hall, 2009.

[66] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, "A new force field for molecular mechanical simulation of nucleic acids and proteins," *Journal of the American Chemical Society*, vol. 106, no. 3, pp. 765–784, 1984.

[67] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, "GROMACS: A message-passing parallel molecular dynamics implementation," *Computer Physics Communications*, vol. 91, pp. 43–56, Sept. 1995.

[68] E. Lindahl, B. Hess, and D. Van Der Spoel, "Gromacs 3.0: a package for molecular simulation and trajectory analysis," *Molecular modeling annual*, vol. 7, no. 8, pp. 306–317, 2001.

[69] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, "Gromacs: fast, flexible, and free," *Journal of computational chemistry*, vol. 26, no. 16, pp. 1701–1718, 2005.

[70] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, "Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of chemical theory and computation*, vol. 4, no. 3, pp. 435–447, 2008.

[71] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, *et al.*, "Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, no. 7, pp. 845–854, 2013.

[72] S. Páll, M. J. Abraham, C. Kutzner, B. Hess, and E. Lindahl, "Tackling exascale software challenges in molecular dynamics simulations with gromacs," in *International Conference on Exascale Applications and Software*, pp. 3–27, Springer, 2014.

[73] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1-2, pp. 19 – 25, 2015.

[74] R. Steendam, "Msc-thesis." https://github.com/R-Steendam/MSc-Thesis, 2020.

[75] A. J. Sweere and J. G. Fraaije, "Accuracy test of the opls-aa force field for calculating free energies of mixing and comparison with pac-mac," *Journal of chemical theory and computation*, vol. 13, no. 5, pp. 1911–1923, 2017.

[76] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids," *Journal of the American Chemical Society*, vol. 118, no. 45, pp. 11225–11236, 1996.

[77] J. R. Hill, A. Tokmakoff, K. A. Peterson, B. Sauter, D. Zimdars, D. D. Dlott, and M. Fayer, "Vibrational dynamics of carbon monoxide at the active site of myoglobin: Picosecond infrared free-electron laser pump-probe experiments," *The Journal of Physical Chemistry*, vol. 98, no. 43, pp. 11213–11219, 1994.

[78] J. Owrutsky, M. Li, B. Locke, and R. Hochstrasser, "Vibrational relaxation of the co stretch vibration in hemoglobin-co, myoglobin-co, and protoheme-co," *The Journal of Physical Chemistry*, vol. 99, no. 13, pp. 4842–4846, 1995.

[79] L. Wang, C. T. Middleton, M. T. Zanni, and J. L. Skinner, "Development and validation of transferable amide i vibrational frequency maps for peptides," *The Journal of Physical Chemistry B*, vol. 115, no. 13, pp. 3713–3724, 2011.

[80] T. Zeko, S. F. Hannigan, T. Jacisin, M. J. Guberman-Pfeffer, E. R. Falcone, M. J. Guildford, C. Szabo, K. E. Cole, J. Placido, E. Daly, *et al.*, "Ft-ir spectroscopy and density functional theory calculations of 13c isotopologues of the helical peptide z-aib6-otbu," *The Journal of Physical Chemistry B*, vol. 118, no. 1, pp. 58–68, 2014.

[81] M. Gerace and R. F. Loring, "Two-dimensional spectroscopy of coupled vibrations with the optimized mean-trajectory approximation," *The Journal of Physical Chemistry B*, vol. 117, no. 49, pp. 15452–15461, 2013.

[82] J. Liu, T. Zhu, X. He, and J. Z. Zhang, "Mfcc-based fragmentation methods for biomolecules," *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, pp. 323–348, 2017.

[83] K. Kwac and M. Cho, "Differential evolution algorithm approach for describing vibrational solvatochromism," *The Journal of chemical physics*, vol. 151, no. 13, p. 134112, 2019.

[84] T. L. C. Jansen and J. Knoester, "A transferable electrostatic map for solvation effects on amide i vibrations and its application to linear and two-dimensional spectroscopy," *The Journal of Chemical Physics*, vol. 124, no. 4, p. 044502, 2006.

[85] S. H. Brewer, Y. Tang, D. M. Vu, S. Gnanakaran, D. P. Raleigh, and R. B. Dyer, "Temperature dependence of water interactions with the amide carbonyls of α-helices," *Biochemistry*, vol. 51, no. 26, pp. 5293–5299, 2012.

[86] E.-L. Karjalainen, T. Ersmark, and A. Barth, "Optimization of model parameters for describing the amide i spectrum of a large set of proteins," *The Journal of Physical Chemistry B*, vol. 116, no. 16, pp. 4831–4842, 2012.

[87] T. l. C. Jansen, A. G. Dijkstra, T. M. Watson, J. D. Hirst, and J. Knoester, "Erratum:"modeling the amide i bands of small peptides"[j. chem. phys. 125, 044312 (2006)]," *The Journal of Chemical Physics*, vol. 136, no. 20, p. 209901, 2012.

[88] S. Roy, J. Lessing, G. Meisl, Z. Ganim, A. Tokmakoff, J. Knoester, and T. L. Jansen, "Solvent and conformation dependence of amide i vibrations in peptides and proteins containing proline," *The Journal of chemical physics*, vol. 135, no. 23, p. 234507, 2011.

[89] H. Torii and M. Tasumi, "Ab initio molecular orbital study of the amide i vibrational interactions between the peptide groups in di-and tripeptides and considerations on the conformation of the extended helix," *Journal of Raman Spectroscopy*, vol. 29, no. 1, pp. 81–86, 1998.

[90] M. Reppert and A. Tokmakoff, "Electrostatic frequency shifts in amide i vibrational spectra: Direct parameterization against experiment," *The Journal of chemical physics*, vol. 138, no. 13, p. 04B603, 2013.

[91] T. L. C. Jansen and S. Roy, "Amide-i maps." https://github.com/GHlacour/AmideImaps, 2018.

[92] T. l. C. Jansen and J. Knoester, "Nonadiabatic effects in the two-dimensional infrared spectra of peptides: application to alanine dipeptide," *The Journal of Physical Chemistry B*, vol. 110, no. 45, pp. 22910–22916, 2006.

[93] A. V. Cunha, A. S. Bondarenko, and T. L. C. Jansen, "Assessing spectral simulation protocols for the amide i band of proteins," *Journal of Chemical Theory and Computation*, vol. 12, no. 8, pp. 3982–3992, 2016. PMID: 27348022.

[94] T. l. C. Jansen, B. Auer, M. Yang, and J. Skinner, "Two-dimensional infrared spectroscopy and ultrafast anisotropy decay of water," *The Journal of chemical physics*, vol. 132, no. 22, p. 224503, 2010.

[95] C. Leforestier, R. Bisseling, C. Cerjan, M. Feit, R. Friesner, A. Guldberg, A. Hammerich, G. Jolicard, W. Karrlein, H.-D. Meyer, *et al.*, "A comparison of different propagation schemes for the time dependent schrödinger equation," *Journal of Computational Physics*, vol. 94, no. 1, pp. 59–80, 1991.

[96] C. R. Baiz, D. Schach, and A. Tokmakoff, "Ultrafast 2d ir microscopy," *Optics Express*, vol. 22, no. 15, pp. 18724–18735, 2014.

[97] T. L. C. Jansen, "Hamiltonian spectra visualizer." https://github.com/lacourjansenlab/Hamiltonians/tree/master/Spectra, 2020.

[98] J. Han, H. Zhang, and D. Abramavicius, "Exchange narrowing and exciton delocalization in disordered j aggregates: Simulated peak shapes in the two dimensional spectra," *The Journal of Chemical Physics*, vol. 139, no. 3, p. 034313, 2013.

[99] G. Orlandi and W. Siebrand, "Theory of vibronic intensity borrowing. comparison of herzberg-teller and born-oppenheimer coupling," *The Journal of Chemical Physics*, vol. 58, no. 10, pp. 4513–4523, 1973.

[100] I. Sushko, S. Novotarskyi, R. Körner, A. K. Pandey, V. V. Kovalishyn, V. V. Prokopenko, and I. V. Tetko, "Applicability domain for in silico models to achieve accuracy of experimental measurements," *Journal of Chemometrics*, vol. 24, no. 3-4, pp. 202–208, 2010.

[101] E. Passini, O. J. Britton, H. R. Lu, J. Rohrbacher, A. N. Hermans, D. J. Gallacher, R. J. Greig, A. Bueno-Orovio, and B. Rodriguez, "Human in silico drug trials demonstrate higher accuracy than animal models in predicting clinical pro-arrhythmic cardiotoxicity," *Frontiers in physiology*, vol. 8, p. 668, 2017.

[102] X. Ma, Y. Shao, L. Tian, D. A. Flasch, H. L. Mulder, M. N. Edmonson, Y. Liu, X. Chen, S. Newman, J. Nakitandwe, *et al.*, "Analysis of error profiles in deep next-generation sequencing data," *Genome biology*, vol. 20, no. 1, p. 50, 2019.

[103] E. Luiz Folador, A. Fernandes de Oliveira Junior, S. Tiwari, S. Babar Jamal, R. Salgado Ferreira, D. Barh, P. Ghosh, A. Silva, and V. Azevedo, "In silico protein-protein interactions: avoiding data and method biases over sensitivity and specificity," *Current Protein and Peptide Science*, vol. 16, no. 8, pp. 689–700, 2015.

[104] Y. Mu, D. S. Kosov, and G. Stock, "Conformational dynamics of trialanine in water. 2. comparison of amber, charmm, gromos, and opls force fields to nmr and infrared experiments," *The Journal of Physical Chemistry B*, vol. 107, no. 21, pp. 5064–5073, 2003.

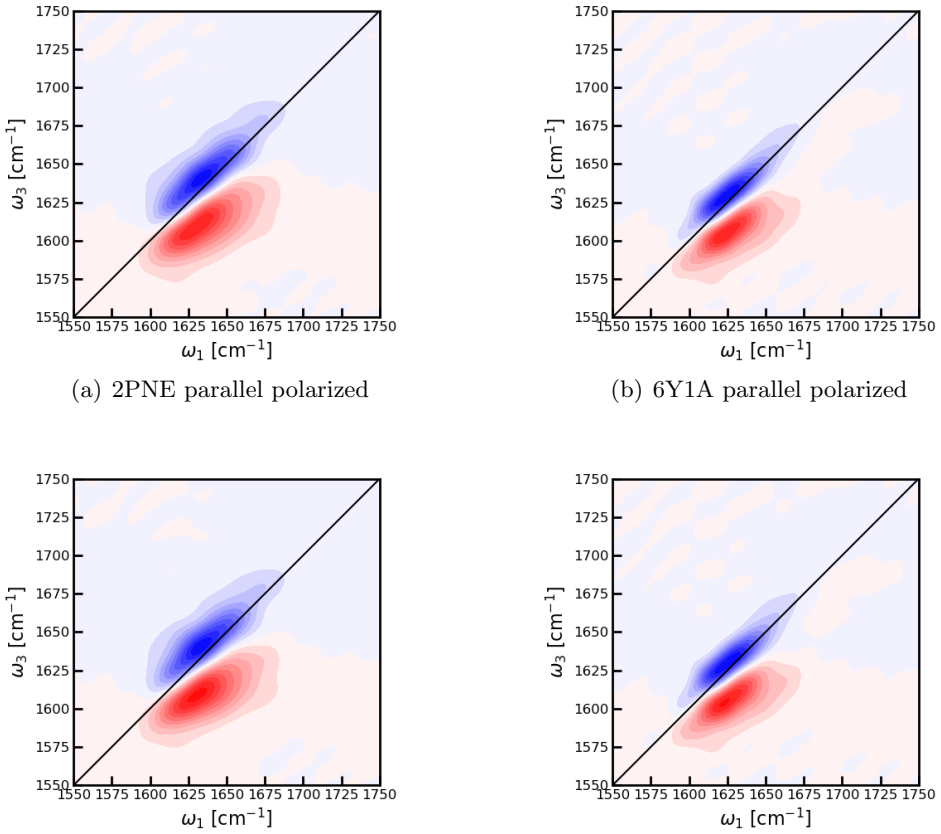
[105] M. Viceconti, F. Pappalardo, B. Rodriguez, M. Horner, J. Bischoff, and F. M. Tshinanu, "In silico trials: Verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products," *Methods*, 2020.

[106] K. A. Beauchamp, Y.-S. Lin, R. Das, and V. S. Pande, "Are protein force fields getting better? a systematic benchmark on 524 diverse nmr measurements," *Journal of chemical theory and computation*, vol. 8, no. 4, pp. 1409–1414, 2012.

Appendix A

Appendix

A.1 2D infrared results 2PNE and 6Y1A Comparison



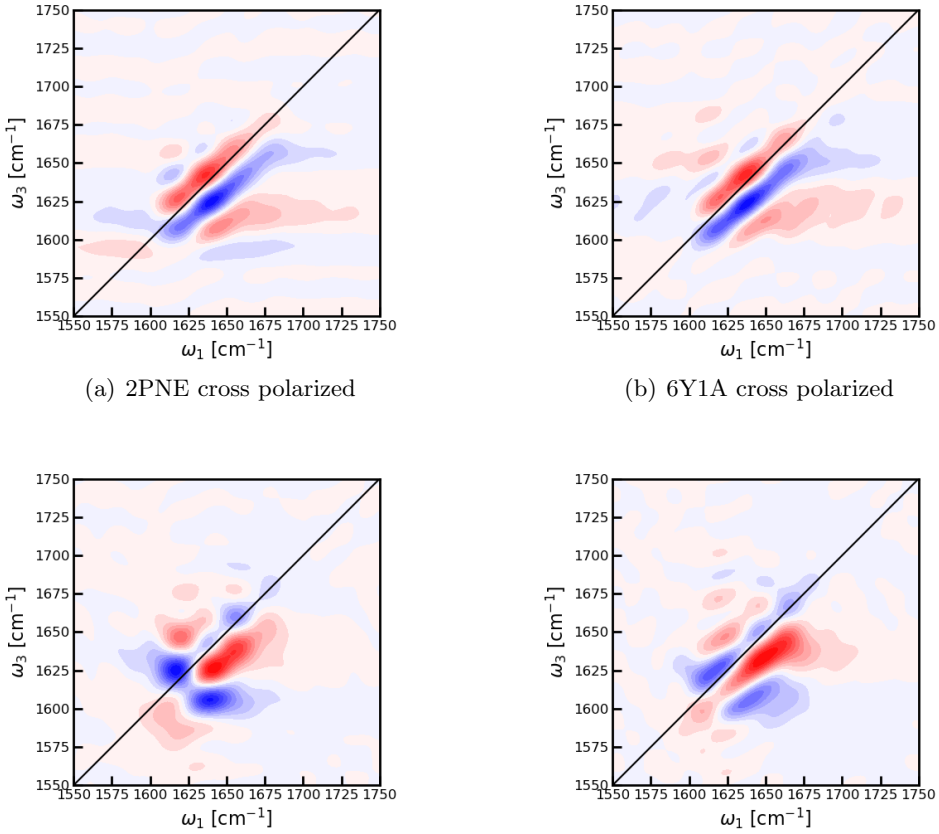
(a) 2PNE parallel polarized

(b) 6Y1A parallel polarized

(c) 2PNE perpendicular polarized

(d) 6Y1A perpendicular polarized

Figure A.1. The comparison between 2D IR spectra obtained from 2PNE and 6Y1A 1



(a) 2PNE cross polarized

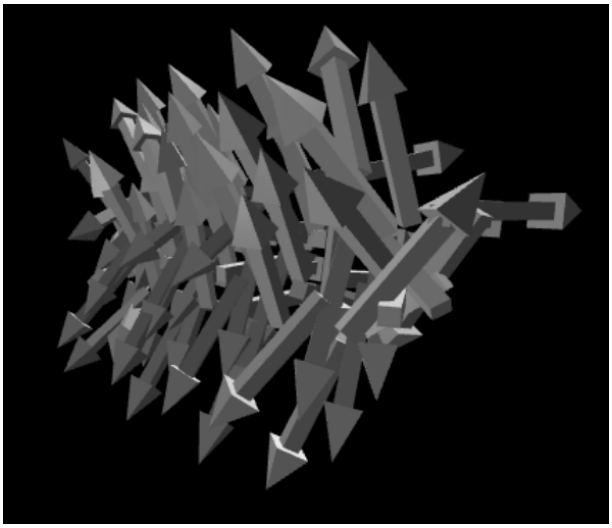
(b) 6Y1A cross polarized

(c) 2PNE Zanni method

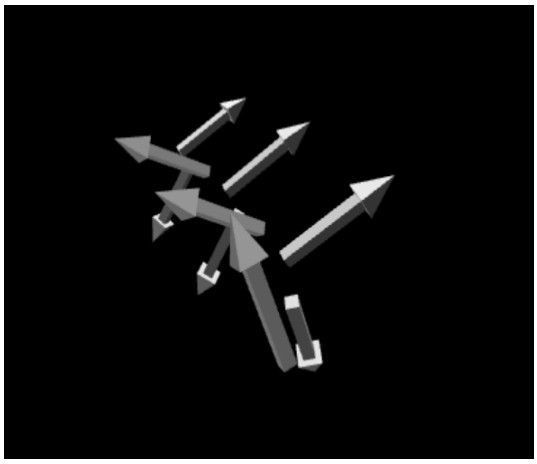
(d) 6Y1A Zanni method

Figure A.2. The comparison between 2D IR spectra obtained from 2PNE and 6Y1A 2

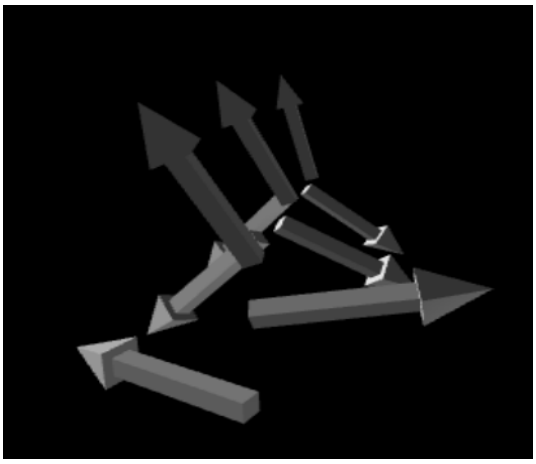
A.2 2PNE Exciton magnitude and direction



(a)

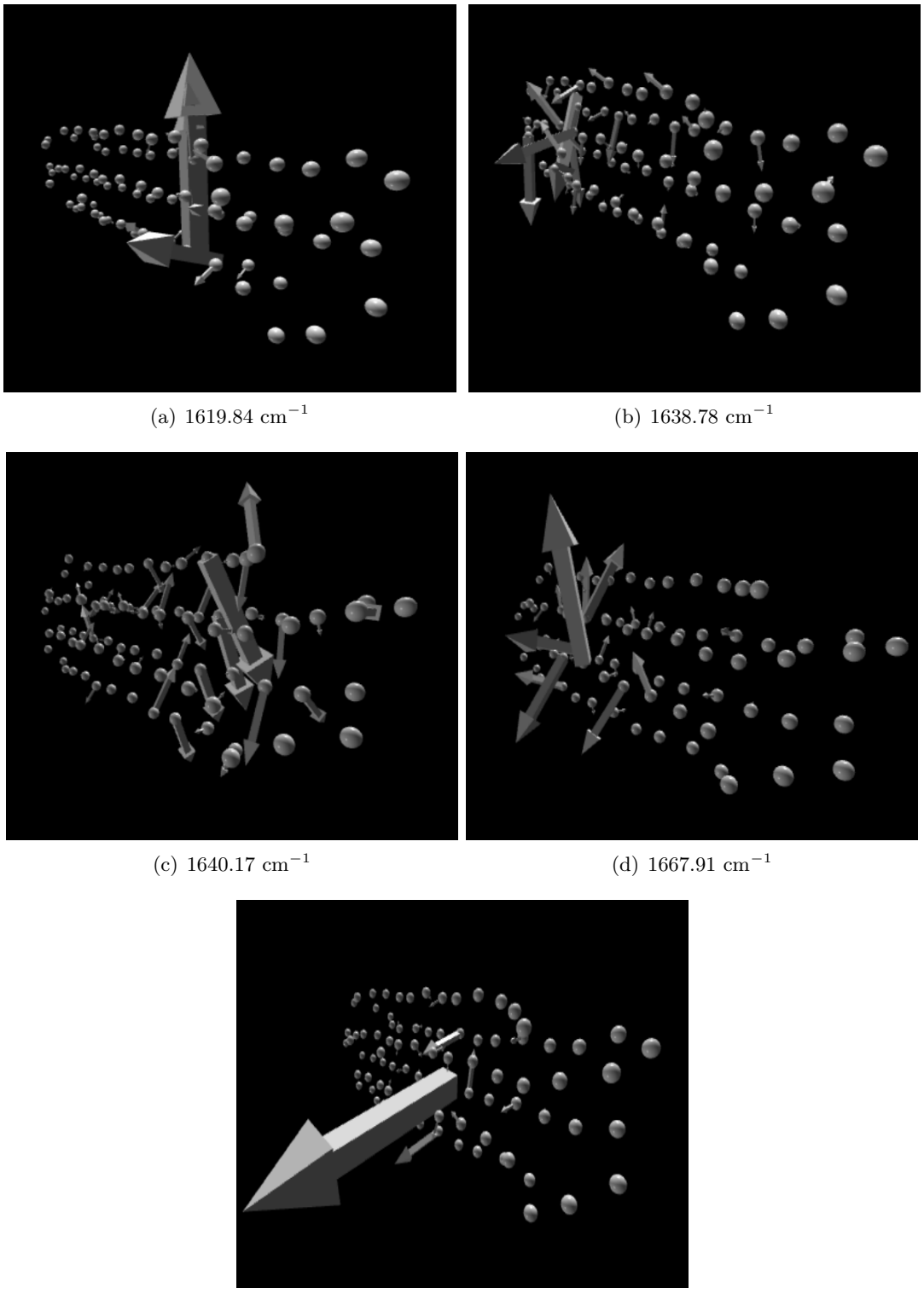


(b)



(c)

Figure A.3. (a) All the transition dipole moments within the 2PNE protein. (b) and (c) show a segment of nine residues in the beginning of protein which exhibit the full rotation every three residues, as discussed in section 1.2.2



(a) 1619.84 cm^{-1}

(b) 1638.78 cm^{-1}

(c) 1640.17 cm^{-1}

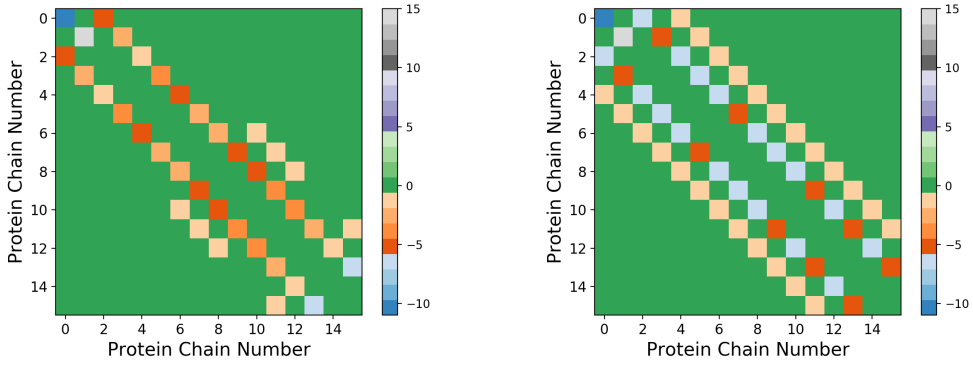
(d) 1667.91 cm^{-1}

(e) 1672.31 cm^{-1}

Figure A.4. The exciton weighted value of excitation for the eigenvalues calculated in section 3.3.1. The couplings within the protein are shown as beads and the arrows show the direction and magnitude of the oscillation of the 2PNE protein for different eigenfrequencies.

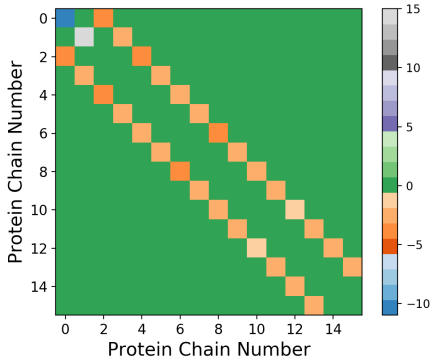
A.3 Hamiltonians obtained with isotopic labeling

These Hamiltonians in Figure A.5 and Figure A.6 show the couplings that the isotopically labeled residues have with their neighbours of a different protein chain

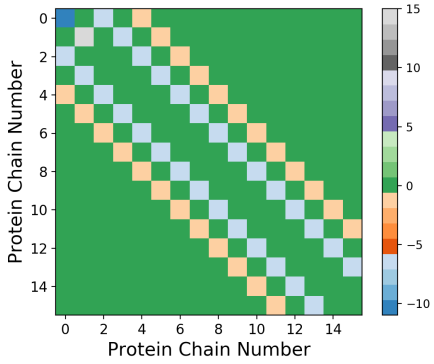


(a) Ala13

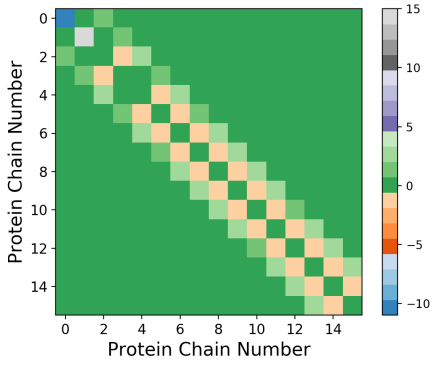
(b) Leu16



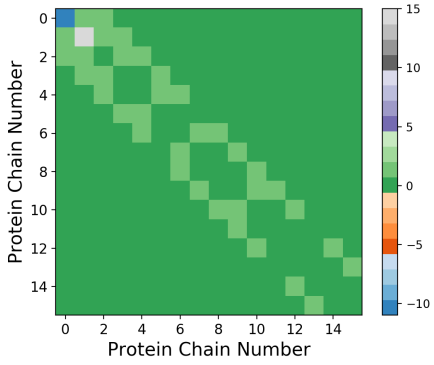
(c) Ser19



(d) Ser20

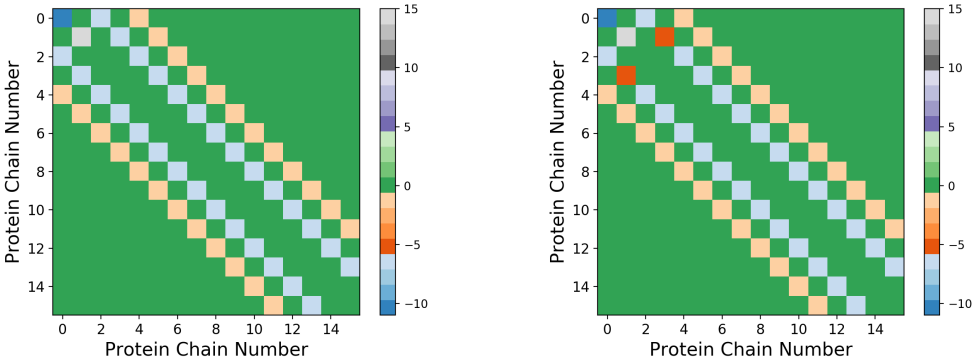


(e) Phe23

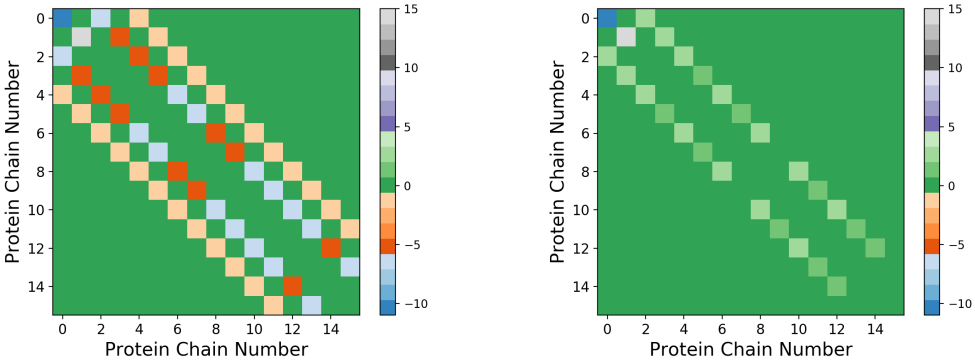


(f) Gly24

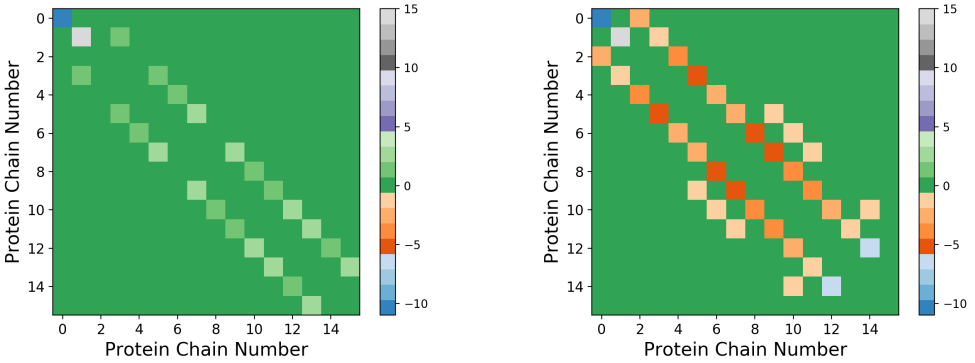
Figure A.5. Hamiltonians showing the internal couplings between residues. (1/2)



(a) Ala25 (b) Ile26



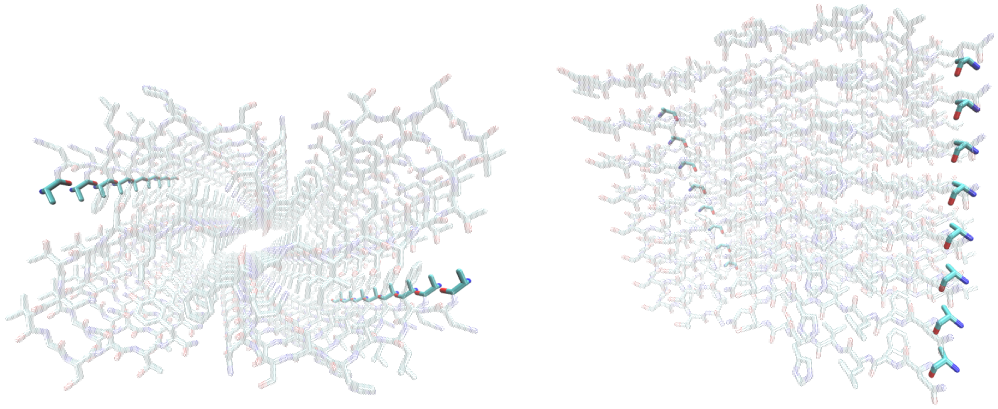
(c) Leu27 (d) Val32



(e) Gly33 (f) Tyr37

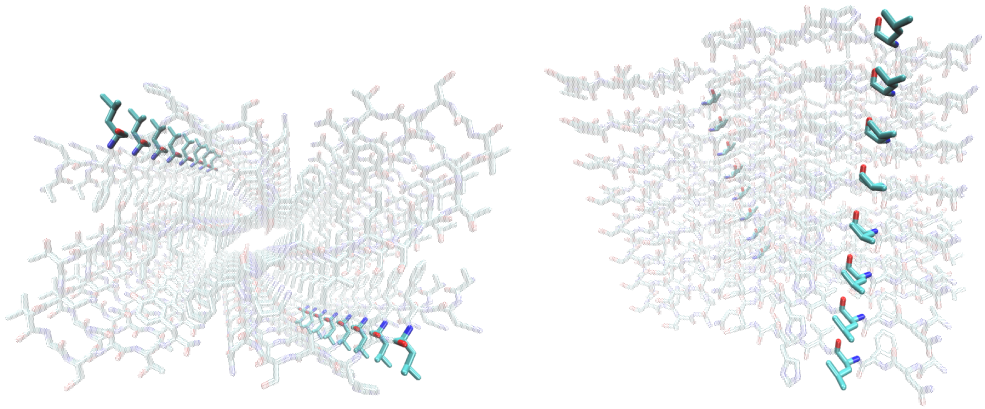
Figure A.6. Hamiltonians showing the internal couplings between residues. (2/2)

A.4 Structural placement of Residues in isotopic labeling



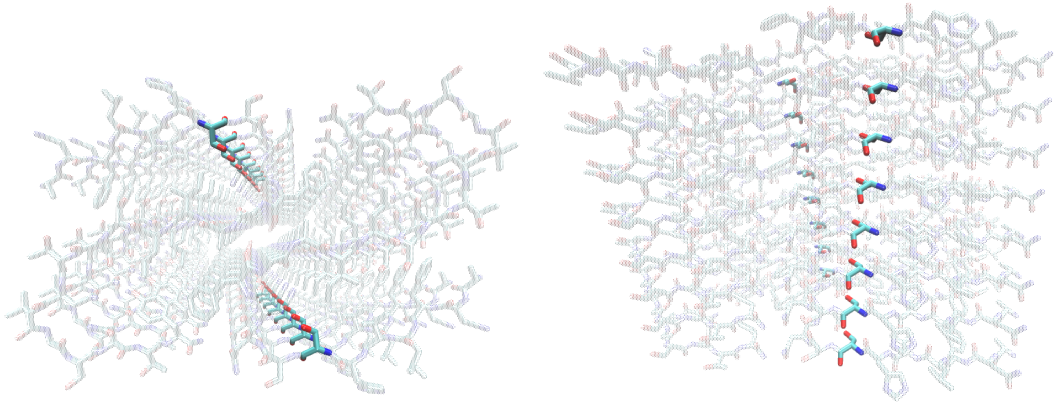
(a) Top view (b) Side view

Figure A.7. Ala13 residues in the 6Y1A aggregate



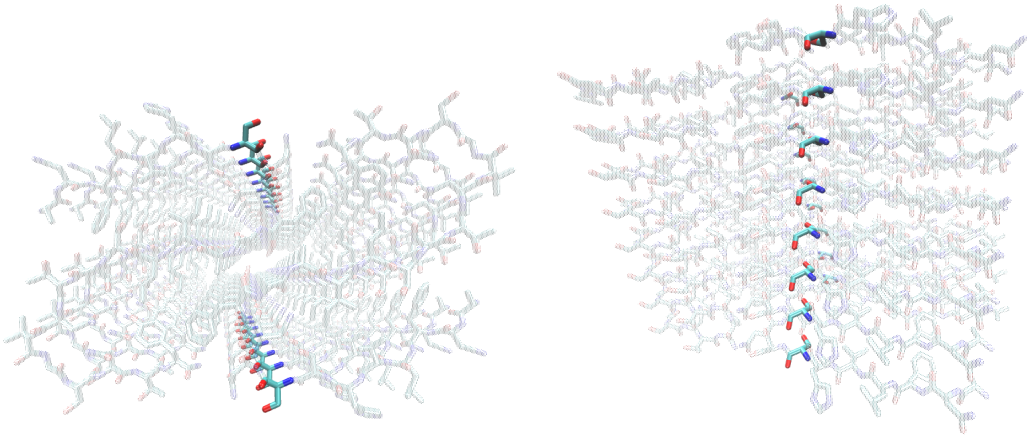
(a) Top view (b) Side view

Figure A.8. Leu16 residues in the 6Y1A aggregate



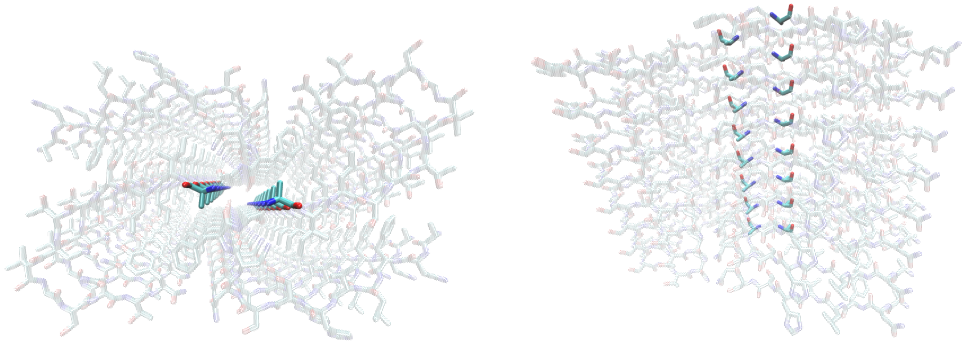
(a) Top view (b) Side view

Figure A.9. Ser19 residues in the 6Y1A aggregate



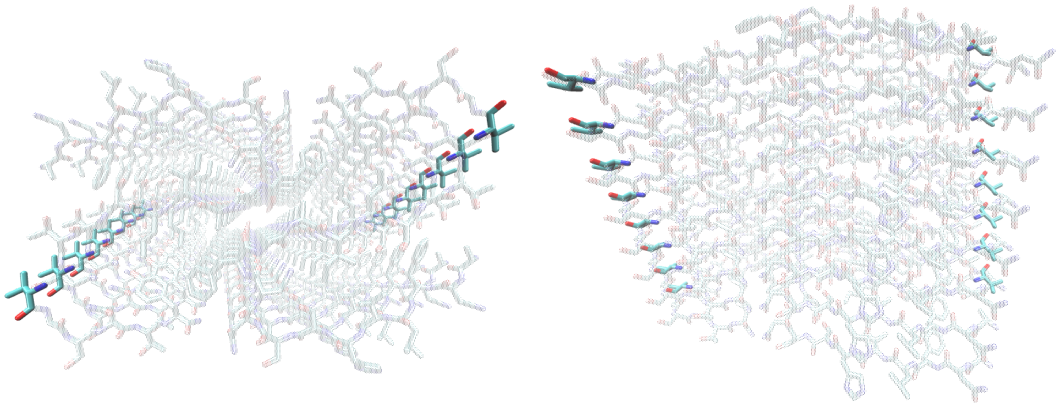
(a) Top view (b) Side view

Figure A.10. Ser20 residues in the 6Y1A aggregate



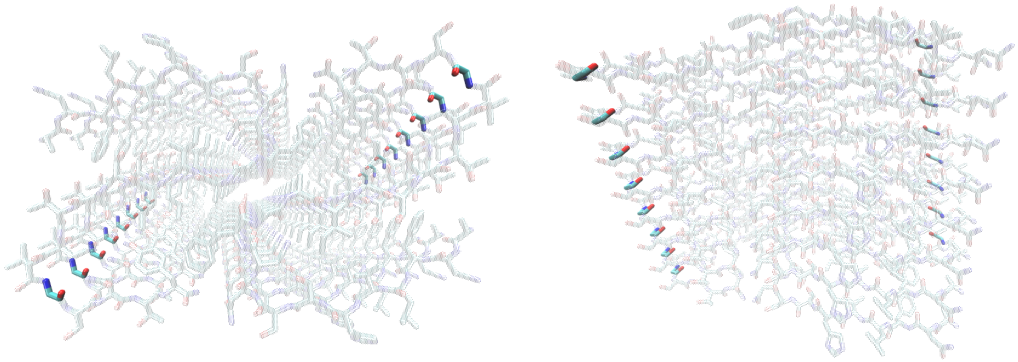
(a) Top view (b) Side view

Figure A.11. Ala25 residues in the 6Y1A aggregate



(a) Top view (b) Side view

Figure A.12. Val32 residues in the 6Y1A aggregate



(a) Top view (b) Side view

Figure A.13. Gly33 residues in the 6Y1A aggregate

A.5 Internal Couplings per chain for the FGAIL sequence

Couplings between FGAIL sequences					
Initial Protein Chain	F-F	G-G	A-A	I-I	L-L
A	1.850	1.374	-6.702	-6.123	-6.348
B	1.785	1.031	-6.513	-5.794	-4.539
C	2.081	1.607	-6.579	-6.222	-5.529
D	1.908	0.898	-6.359	-6.275	-5.611
E	2.231	1.351	-6.382	-6.459	-6.082
F	1.908	0.617	-6.395	-6.273	-5.858
G	2.155	1.440	-6.573	-6.358	-5.512
H	2.209	0.733	-6.457	-6.612	-5.323
I	2.049	1.327	-6.482	-6.478	-5.961
J	2.735	0.745	-6.508	-6.544	-6.465
K	1.859	0.804	-6.316	-6.707	-6.132
L	2.459	0.686	-6.498	-6.460	-6.179
M	2.639	1.965	-6.549	-6.474	-4.875
N	3.128	1.118	-5.926	-6.400	-6.006

Table A.1. Internal Couplings between the FGAIL sequence of a protein chain and the protein chain directly beneath it. Data is within expected margins from Serrano *et al.*[44]