Computationally Distinguishing Various α-Synuclein Polymorphs with Infrared Absorption Spectra

Tim Zuidema
s3905098

Faculty of Science and Engineering
University of Groningen
Theory of Condensed Matter group
2023 1a

Made under the supervision of
Thomas L.C. Jansen
Kim E. van Adrichem

rijksuniversiteit
groningen

Abstract

This thesis aims to provide insight into the infrared absorption spectrum of three different polymorphs of α-synuclein, which all might provide for a different pathway of Parkinson's disease. Parkinson's disease is a neurodegenerative disorder characterized by Lewy bodies [1]. The Hamiltonians of the polymorphs are investigated to provide insight into how the structures of the polymorphs influence their infrared absorption spectra. Two out of the three structures are only partial structures. It became clear that 2N0A[2] and 8ADV[3] are most likely distinguishable. The differences in absorption between the spectra are minor, likely because differing secondary structures did not have large differences in absorption. However, some things could be done to improve these results. Structures with complete chains would create more uses for this data as that would allow for even better comparison with physical structures. Larger structures could be used to limit edge effects. 2DIRRaman spectroscopy could be used to get a more detailed view of the absorption. The last two options are however very computationally expensive. Using isotope labeling during an experiment would make the computational results of partial structures presented in this thesis more comparable to experimental results.

Contents

1 Introduction

The overall goal of this computational experiment is to enable the distinguishing of α-synuclein polymorphs by measuring their infrared absorption spectra. Measuring infrared absorption spectra is relatively inexpensive and fast compared to more sophisticated methods so it still has a purpose. Experiments performed to uncover infrared absorption spectra also do not require complex sample preparation.

α-synuclein is a highly abundant protein in the brain 140 amino acids long. Its function is largely unknown. [4] It is associated with synaptic vesicles, even though it doesn't contain transmembrane domain or lipid anchors, which are structures used by proteins to attach themselves to membranes. Because it cannot attach itself to the membrane α-synuclein is a peripheral membrane protein, it merely floats next to the membrane. The protein remains unfolded and proximity to highly curved membranes and acidic phospholipid headgroups causes folding into the proper formation, which is α-helical. In non-neural cells, the concentration of α-synuclein is diffuse but it localizes to the axon terminal in neurons. The axon terminal is the specialized structure at the end of an axon which is responsible for signal transmission in the brain. Axon terminals provide a highly curved surrounding, which induces folding of α-synuclein. α-synuclein weakly interacts with the cell membrane, which is proven by the protein localizing relatively late to developing synapses in cultured neurons. Its location in axon terminals indicates a role in neurotransmitter release. [4] Aggregation of α-synuclein is also related to Parkinson's disease. The structure of folded α-synuclein is shown below in figure 1.

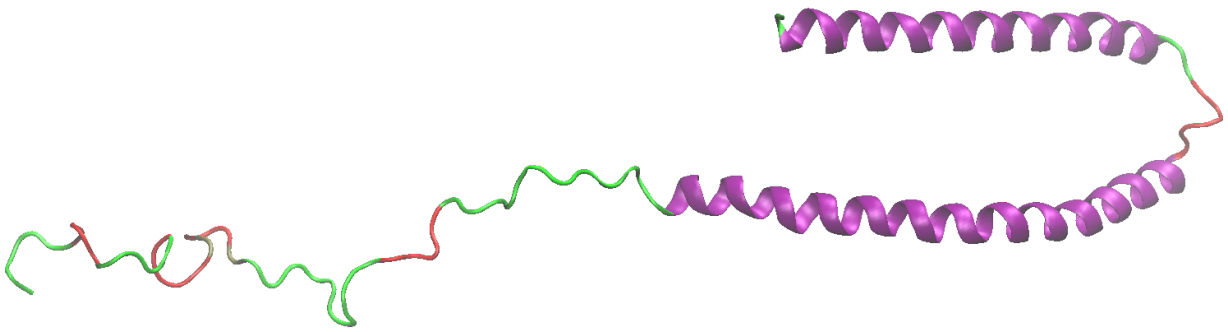


Figure 1: This image shows all 140 amino acids of 1XQ8 [5] which is micelle bound α-synuclein, this structure was discovered with solution NMR. The colors depict secondary structure, with purple being the α-helix characteristic of α-synuclein. The turns are colored red and the coils are green.

Parkinson's disease is a neurodegenerative disease disorder. It has many motor and non-motor symptoms, like resting tremors, drooling, swallowing problems, depression, memory impairment and confusion. A pathological sign of Parkinson's disease is the presence of Lewy Bodies, which are intraneuronal inclusions containing immunoreactive α-synuclein aggregates, neurofilament proteins and proteins involved with proteolysis [1].

These α-synuclein aggregates are associated with Parkinson's disease and knowledge of their formation may lead to insights into Parkinson's pathology. There are different ways of aggregating α-synuclein into fibrils, where they are stacked differently. These different aggregates are also known as polymorphs. The specific polymorph that will develop is heavily correlated with the ionic strength of the solution that the α-synuclein inhabits. This is most likely related to different charges present on the molecule at different pH. However, these fibrils are more stable and don't change in response to changes in pH levels once formed. Because the polymorphs have identical primary structures, they are very difficult to identify [6].

With the help of specialized techniques, however, it is possible to uncover the structures of these polymorphs. In this thesis, the structures will be used to calculate their infrared absorption spectra. Those absorption spectra can subsequently be used to differentiate between the structures more easily in future experiments.

These different aggregates may indicate different pathways by which Parkinson's disease is established. The aim of this thesis is therefore to provide insight into the 1D infrared absorption spectrum of three α-synuclein amyloid fibril polymorphs: 2N0A[2], 6H6B[7] and 8ADV[3] and distinguish them. These structures were taken from the Protein Data Bank[8] and are shown below. 2N0A was chosen because of its completeness since it contains all 140 amino acids. 6H6B and 8ADV were both chosen because they contain a relatively high number of protein chains, namely 10, which will limit edge effects compared to smaller aggregates. All these proteins were selected out of a list of suggestions by Steven J. Roeters [1, 6].

The aggregates that will examined have other similarities as well. They all have a core section consisting of large parts of parallel β-sheets. All structures also contain at least one more disorganized tail section in a terminal direction. The reason why the tails are only visible for 2N0A is likely because of the disorganized nature of those sections mentioned earlier.

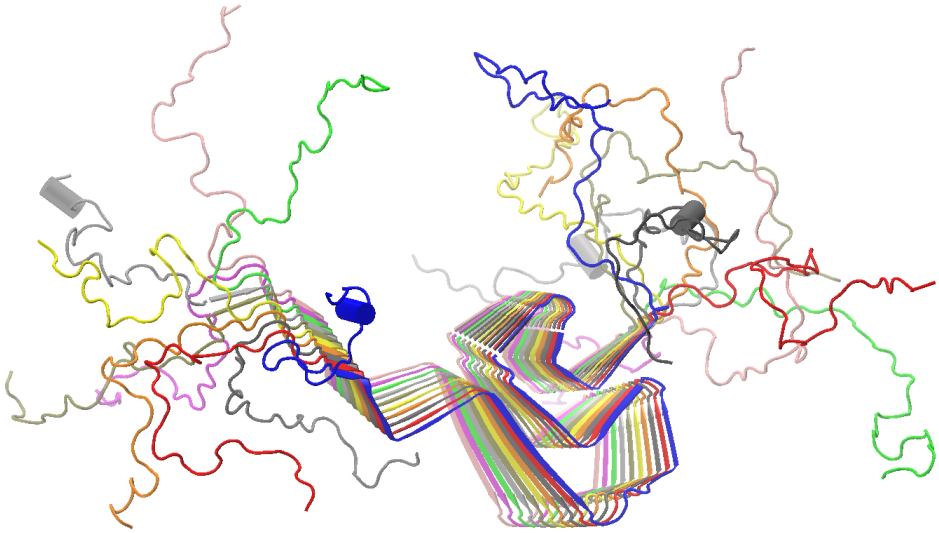


Figure 2: An image of 2N0A[2] made with VMD[9], showing the full 140 residue chains. The fibril contains 10 chains. The chains are colored in order from A-J: blue, red, dark gray, orange, yellow, olive, light gray, green, magenta and pink. It is also very noticeable that the individual chains can be divided into separate sections. The chains contain a solid section in the middle and two looser "tails" at the beginning and end of the chains. The first section, a tail, contains the first amides 1-38. The second section, which is more solid, contains the next amides 39-110. Lastly, the third section which is also a tail, contains the amides 111-148. This structure was uncovered with the help of solid-state NMR.

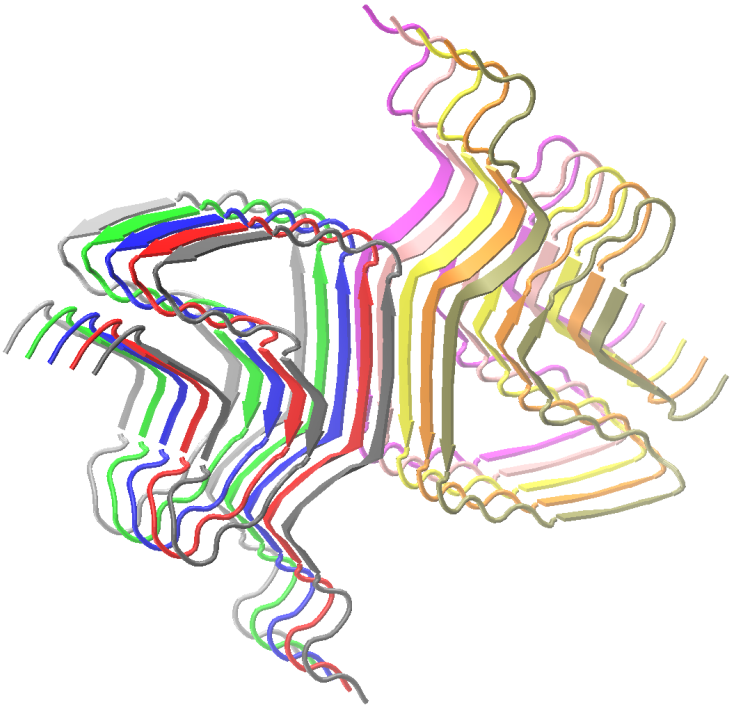


Figure 3: An image of 6H6B[7] made with VMD[9], showing only residues 38-95 per protein. The fibril contains 10 chains. The chains are colored in the same way as 2N0A in figure 2. The structure was uncovered with the help of electron microscopy.

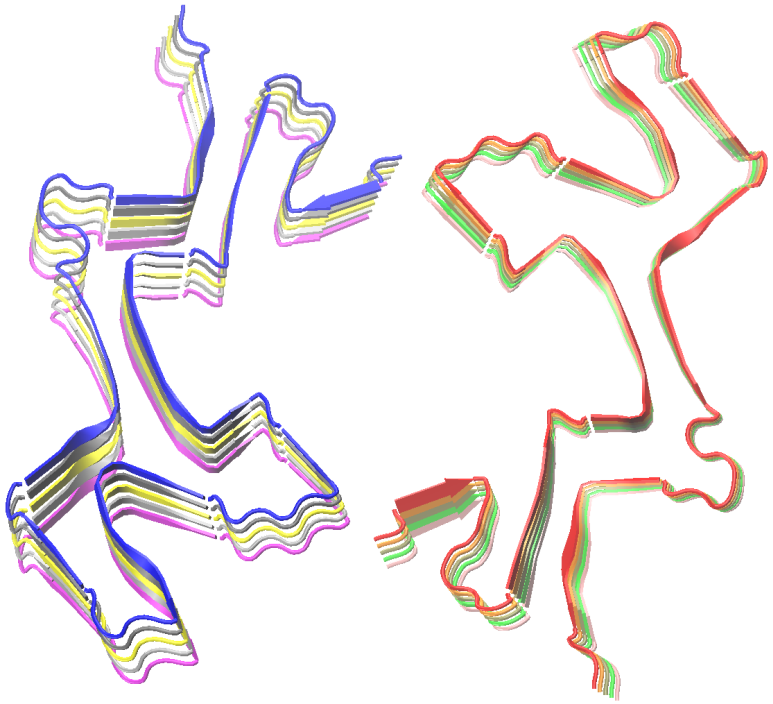


Figure 4: An image of 8ADV[3] made with VMD[9], showing only residues 1-99 per protein. The fibril contains 10 chains. The chains are colored in the same way as 2N0A in figure 2. The structure was also uncovered with the help of electron microscopy.

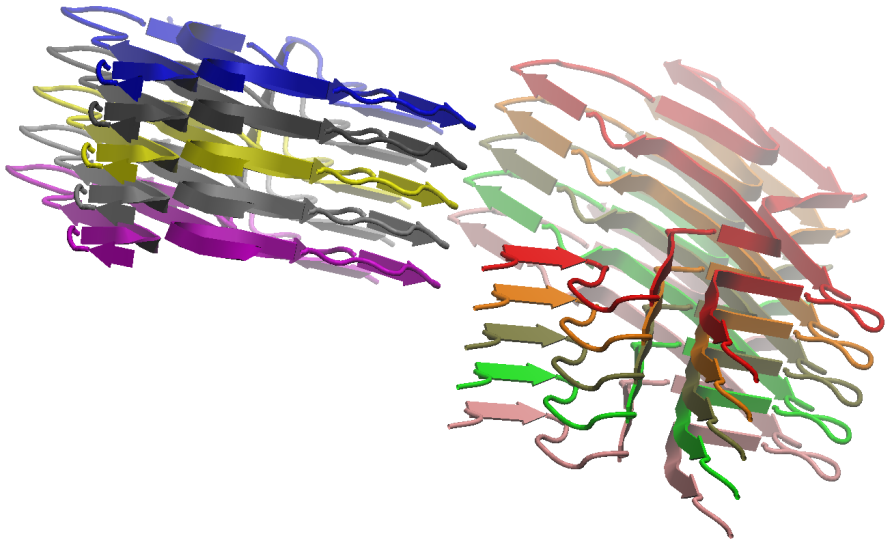


Figure 5: An image of 8ADV[3] made with VMD[9], showing only residues 1-99 per protein. The fibril contains 10 chains. This image is taken from a different angle to highlight the twisted shape of the polymorph.

The tail sections of 6H6B and 8ADV could not be resolved experimentally. This is likely because those sections are very disordered and flexible.

The remainder of this thesis will be divided as follows: firstly the theory behind infrared absorption will be explored, then the software and physical models will be discussed in the methods. Afterwards, the results will be studied. Finally, possible issues with the results are highlighted, an outlook will be presented a conclusion will be drawn.

2 Theory

Infrared spectroscopy is a tool that is used for the study of molecular structures and dynamics. The amide I vibrational mode corresponds to the stretching of carbonyl(C=O) bonds in amides and is utilized in spectroscopy for structural analysis. Amide I vibrational modes couple into the amide I band, the frequencies, bandwidths and intensities of which depend on the electric field. Patterns in inter- and intramolecular couplings also influence the properties of these vibrational modes. To illustrate, α-helices absorb light well at a wavenumber of 1650 cm^{-1} and all amyloid β-sheets absorb light well at a wavenumber of 1620 cm^{-1}. Anti-parallel β-sheets also absorb light well at a wavelength of 1675 cm^{-1}. All β-sheets in this thesis are parallel unless specified otherwise [10–12].

Calculations of amide I spectra consider amide I stretches quantum mechanically and all other degrees of freedom classically. IR absorption spectra are heavily determined by the Hamiltonian. The Hamiltonian takes the form of a matrix, the diagonal terms representing the frequencies and the other off-diagonal terms the couplings between amide groups.

The Hamiltonian is given by [13]

$$H(t) = \sum_n \epsilon_n(t) B_n^\dagger B_n - \frac{1}{2}\sum_n \Delta_n(t) B_n^\dagger B_n^\dagger B_n B_n + \sum_{nm} J_{nm}(t) B_n^\dagger B_m - \sum_n \vec{E}(t)\cdot\vec{\mu}_n(t)[B_n^\dagger + B_n], \quad (1)$$

where $B_n^\dagger$ and B_n are creation and annihilation operators. $\epsilon_n(t)$ represents the energy gap between the ground state and the first excited states on amide group n. The coupling strength of amide group n with the external electric field $\vec{E}(t)$ is controlled by the transition dipole μ_n. The anharmonicity $\Delta_n(t)$ defines the difference in energy energy gap between the ground and first excited state and the first excited and second excited state, where a positive anharmonicity indicates that the latter is smaller than the former. J_{nm} represents the transition dipole coupling. The reason why $\vec{E}(t)$, $J_{nm}(t)$ and $\vec{\mu}_n$ are dependent on time is because the structure of the polymorphs is not completely static and constantly shifts ever so slightly, and will thus be examined at discrete frames.

The transition dipole coupling used in equation 1 is given by [13]

$$J_{nm} = \frac{1}{4\pi\epsilon_0}\left(\frac{\vec{\mu}_n\cdot\vec{\mu}_m}{r_{nm}^3} - 3\frac{\vec{\mu}_n\cdot\vec{r}_{nm}\vec{\mu}_m\cdot\vec{r}_{nm}}{r_{nm}^5}\right), \quad (2)$$

where μ_n is the transition dipole, for a specific transition, r_{nm} the distance between the centers of coupled amide groups in this case.

For each frame of the system, the Hamiltonian will be determined with electronic structure maps [14]. These maps work by taking the frequencies of the amides as if the amides were gas and applying the effects of electric fields to determine the frequencies of the Hamiltonian. This then leads to coupling maps which are used to obtain all of the coupling pairs that the amides form and fill out the matrix.

The amide groups that are important to this research usually don't live "on" amino acid residues but between them. The peptide bond that connects them forms an amide group. Some amino acids also contain an amide group on side chains, namely asparagine and glutamine [15]. The structures in figures 6 and 7 were taken from the book *Molecular Biology of the Cell, 6th ed.* [15].

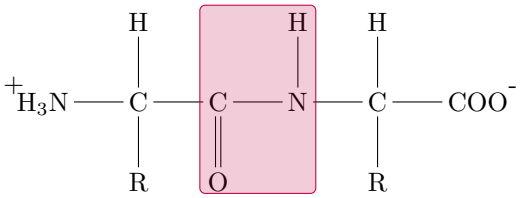
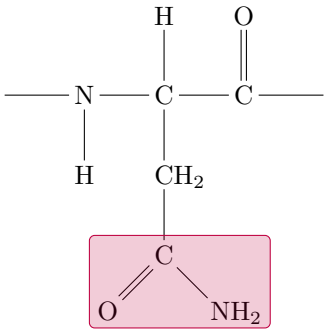
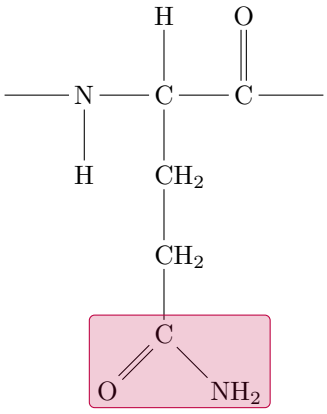


Figure 6: This image shows a protein consisting of two arbitrary amino acid residues. The bond between the amino acid residues forms an amide group, which is highlighted purple. The R in the structure represents the side chain of an amino acid which is dependent on the type of amino acid.



(a) aparagine



(b) glutamine

Figure 7: This image shows asparagine (a, left) and glutamine (b, right) which are the amino acids that contain amide groups in their side chains. The amide groups are highlighted purple.

3 Methods

The structures[2, 3, 7] from the Protein Data Bank[8] were modified using GROMACS version 2023.1[16]. To start, a solution of SPC/E water[17] was added around each polymorph, subsequently, the systems were made to have a neutral charge by adding Na^+ or Cl^{-1} ions. 90 NA^+ ions were added to the 2N0A system, 10 Cl^{-1} ions were added to the 6H6B system and 40 CL^- ions were added to the 8ADV system. The water covered the polymorphs by at least 1 nm for 6H6B and 8ADV by using 36796 and 95629 water molecules respectively. For 2N0A, it was for at least 3 nm by using 526001 water molecules. The OPLS-AA force field for proteins[18] was used. After energy minimizations, 100 ps NVT runs and 100 ps NPT runs were applied to the systems. The production runs for the 1D spectra lasted for 1 ns with 1 frame saved every 20 fs for a total of 50001 frames. The production runs used for the graphical visualizations of the Hamiltonians only had 1 frame saved every 20 ps for a total of 51 frames. The other parameters for the production runs were identical in all other respects. The simulation time step used was 2 fs and the simulation was run at a temperature of 300 K and 1 bar of pressure. The particle-mesh-Ewald scheme[19] was used for long-range electrostatics. The cut-off distance for short-range Van der Waals and Electrostatic effects was 1 nm.

The vibrational Hamiltonian of the amide-I vibrations was calculated with AIM[20]. The program used the Skinner frequency map[10] for backbone and sidechain amide I. The TDCTasumi[21] map was used for transition-dipole coupling. The simulation used Torii dipoles [22]. This resulted in the creation of a matrix with the expected frequencies of the vibrations on the diagonal. These frequencies include electrostatic effects from the environment but not coupling with nearby amide groups. The strengths of the couplings are on the off-diagonal. Influencers had been set to all, which means that the influences of the water, ions and the proteins themselves on the electric field, potential and gradient are taken into consideration. Oscillators had been set to AmideBB AmideSC. This will cause AIM to consider all amides between amino acids on the backbone and those on sidechains respectively. Apply_dd_coupling was set to Same. This setting simply causes the amides in the structures to couple, if it were to be set to None, the coupling between all of the amides would be zero for instance. This setting would be more interesting if there were multiple types of oscillators in the system and not just amides. NN_coupling_choice was set to GLDP. This will cause the coupling between amides that are nearest neighbors to be calculated with the GLDP method. GLDP uses a map based on Ramachandran angles surrounding the amide bond. SphereSize had been set to 20. This meant that only the electrostatic effects of residues within 20 angstroms (Å) were considered for the calculation of the electric potential, field and gradient. TreatNN was set to True. This caused some atoms, depending on primary structure, frequency mapping and force field interaction, to be excluded from electrostatic interaction and instead be accounted for using Ramachandran angles.

The simulation of infrared absorption spectra was performed with NISE[23–28] using the first-order response function. Which takes the form [29]

$$I(t) = \sum_{\alpha}^{x,y,z} = \langle \mu_\alpha(t) U(t,0) \mu_\alpha(0) \rangle \exp(-t/2T_1). \tag{3}$$

Here, $I(t)$ represents the linear response function, α the cartesian coordinate, μ the transition dipole moment, $U(t,0)$ is the time evolution operator, T_1 the lifetime and t time. The time evolution operator is dependent on the Hamiltonian $H(t)$ from equation 1. The Fourier transform of $I(t)$ gives the absorption with respect to wavenumber.

NISE in all cases used a samplerate of 100, meaning that every 100 frames are used for averaging. The frequencies considered by NISE ranged between 1525 and 1725 cm^{-1}. Couplingcut was set to 0, meaning no couplings were neglected because they were under the given frequency. Runtimes were set to 1023 0 256, meaning that t_1, t_2 and t_3 respectively took on those values, this is important for equation 3. The technique was set to absorption, causing equation 3 to be used instead of some other method. Timestep was 20 to indicate that there exists 20 fs of time between each frame. For 2N0A, 6H6B and 8ADV, Sites were set to 1480, 600 and 1030 respectively to reflect the total number of amides in those polymorphs. Length was 50000 to process all 50,000 frames in the Hamiltonian. Propagation was set to Coupling, which is a faster computational method than Sparse.

All of the calculations were performed on the Hábrók computer cluster of the RUG. The standard storage partitions assigned to users are too small to store the large files generated by GROMACS and AIM. This challenge was overcome by designing scripts that ran GROMACS, AIM and NISE in one job and only returning the final results to the personal partition. This sidesteps the issues caused by the large files by leaving them in the internal storage of the Hábrók nodes. The total charge in the system also caused problems. AIM only accepts close to integer values for total charge, and rounding errors made by GROMACS will cause the system to no longer have a close to integer total charge for systems as large as 2N0A. A slight alteration was therefore made to the installation of AIM to still allow it to work with 2N0A.

4 Results

This section will first cover the Hamiltonians and then the infrared absorption spectra of the 2N0A, 6H6B and 8ADV, afterwards the absorption spectra are considered.

4.1 Hamiltonians

A-A	A-B	A-C	A-D	A-E	A-F	A-G	A-H	A-I	A-J
B-A	B-B	B-C	B-D	B-E	B-F	B-G	B-H	B-I	B-J
C-A	C-B	C-C	C-D	C-E	C-F	C-G	C-H	C-I	C-J
D-A	D-B	D-C	D-D	D-E	D-F	D-G	D-H	D-I	D-J
E-A	E-B	E-C	E-D	E-E	E-F	E-G	E-H	E-I	E-J
F-A	F-B	F-C	F-D	F-E	F-F	F-G	F-H	F-I	F-J
G-A	G-B	G-C	G-D	G-E	G-F	G-G	G-H	G-I	G-J
H-A	H-B	H-C	H-D	H-E	H-F	H-G	H-H	H-I	H-J
I-A	I-B	I-C	I-D	I-E	I-F	I-G	I-H	I-I	I-J
J-A	J-B	J-C	J-D	J-E	J-F	J-G	J-H	J-I	J-J

Figure 8: This image shows a map that can be used to read the Hamiltonians. The diagonal squares correspond to a chain's coupling between internal amide groups, the off-diagonal squares contain the coupling between residues of two different chains. The chains are color-coded and labeled with letters.

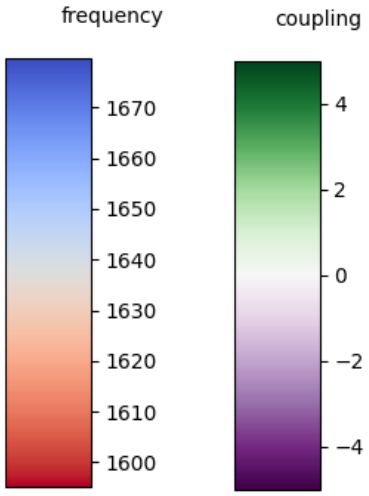


Figure 9: This figure shows all of the color bars relevant to the Hamiltonians on the following pages. The frequency is in cm^{-1} and the coupling strength is given in cm^{-1}.

The figures of the Hamiltonians in this section all start counting from zero on both axes. The count continues with each next chain. The backbones of 2N0A, 6H6B and 8ADV are 139, 56 and 98 amide groups long respectively. If one wanted to look up the coupling strength of the first amide group on chain B with the coupling strength of amide group 50 on chain I in polymorph 6H6B, the correct place on the graph of the Hamiltonian would be either (57,505) or (505,57).

Each (n,n) coordinate represents the vibrational frequency of amide group n.

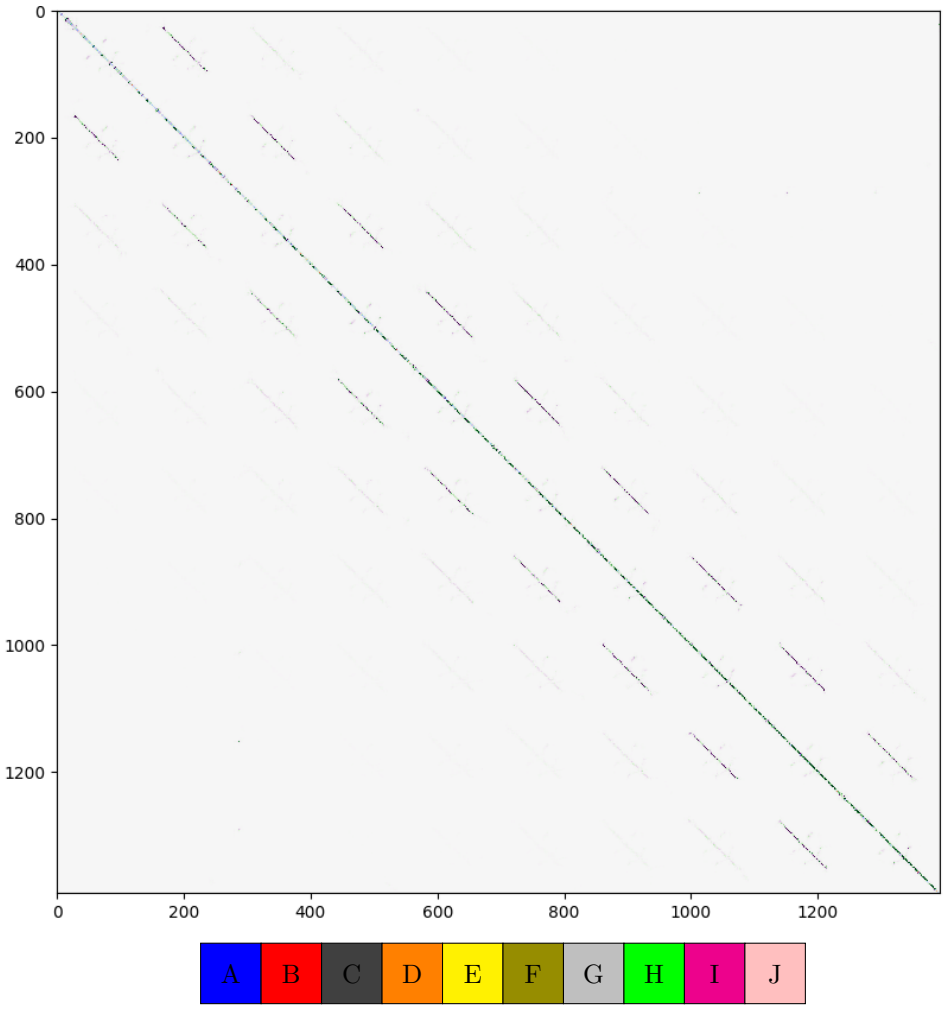


Figure 10: This image shows all the couplings between amide groups between amino acids in 2N0A. Positive couplings are green and negative couplings are purple. The more vivid the color, the stronger the coupling. The polymorph is displayed schematically below the graph. This image was made by taking the average Hamiltonian of all of the frames.

Figure 10 shows strong couplings with neighbors on the same chain. Couplings with the same or nearby amide group on adjacent chains are also strong if they happen to be on the solid part of the chain. The solid part contains amino acids 38-97. All other couplings are barely visible if at all. To more clearly see couplings, a figure zoomed into A-A and with a small range of coupling strength is in the appendix.

The strongest couplings not described already exist between amide groups that live on β-sheets that are adjacent in space. This concerns coupling between amide groups 45-51 and 72-78 and amide groups 67-70 and 90-93. The coupling between these groups is also strong if they live on adjacent chains, although only inconsistently so. There are also a few strong couplings inside of the tail sections, but they happen infrequently.

Looking at equation 1 and equation 2 provides more insight into the couplings inside different sections. It can be seen in those equations that the coupling depends on $\frac{1}{r_{nm}^3}$. Therefore adjacent amide groups in the same chain which will always have a small r_{nm} will couple strongly and so will neighboring amide groups in adjacent chains in the middle section where the chains are very close together. In the loose tail sections, the distances will be larger because the chains are more spread out so the weaker coupling is expected, although sometimes the chains passing near each other does lead to infrequently strong couplings as is observed.

To more clearly see the couplings there are additional graphs of the Hamiltonian of 2N0A in the appendix.

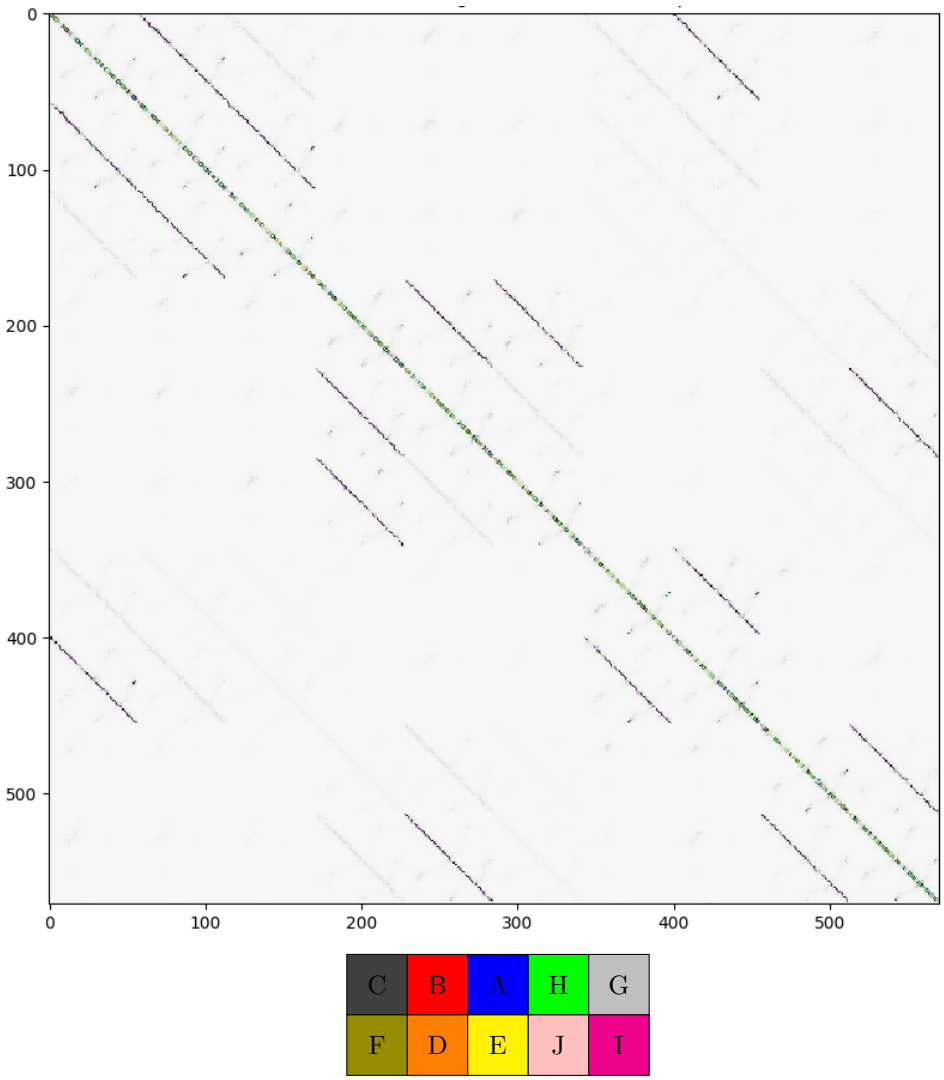


Figure 11: This is the same figure as figure 10, but showcases the Hamiltonian of 6H6B instead.

Figure 11 shows that the couplings between amide groups that are adjacent on the same chain are very strong, as is the coupling between the counterparts of the amide groups on adjacent chains. The coupling of groups across rows is extremely weak, barely even registering.

The coupling between amide groups 9-12 and 39-41 and amide groups 30-34 and 52-55. These groups are also relatively strongly coupled across chains as long as they are in the same row. Across rows, the amide groups 14-20 and 13-18 couple most strongly. Which corresponds to the β-sheet forming a bridge between the two rows on opposite chains.

To more clearly see the couplings there are additional figures of the Hamiltonian of 6H6B in the appendix.

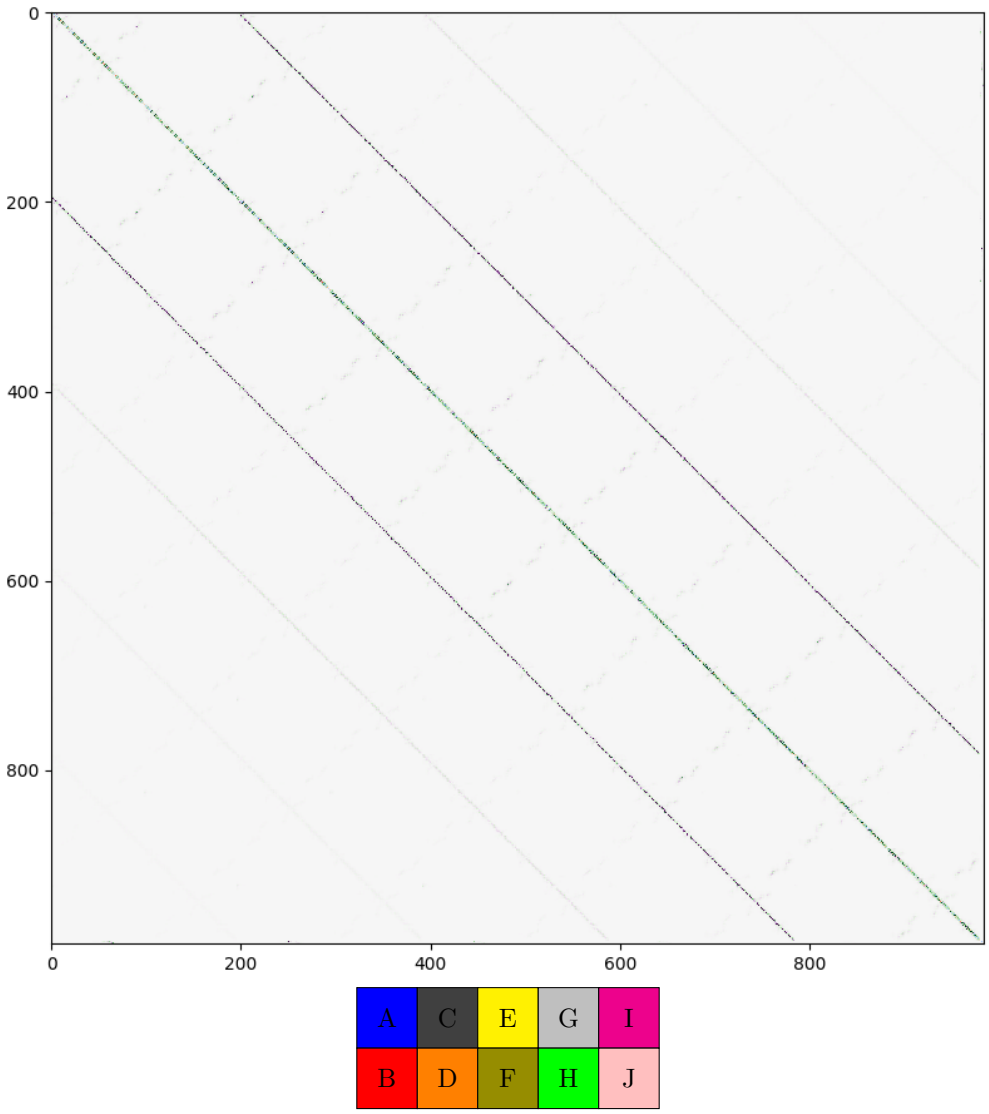


Figure 12: This is the same figure as figure 10, but showcases the Hamiltonian of 8ADV instead.

Figure 12 tells us that the coupling between adjacent amide groups in the same chain is very strong. The coupling between amide groups on the adjacent chain on the same row is also easily visible, although the coupling is not as potent as the coupling in the same chain.

The figure also highlights that the very strongest couplings between amide groups are between groups 25-28 and 71-75 and groups 13-17 and 88-92 inside of the same chains. It also becomes apparent that the coupling between rows is weaker than in 6H6B, no coupling is even visible. Observation of the structure reveals that this makes a lot of sense as the bridges between the rows are very narrow in 8ADV whilst the connection is very broad for 6H6B.

To more clearly see the couplings there are additional figures of the Hamiltonian of 8ADV in the appendix.

4.2 Absorption Spectra

The purpose of comparing the shapes of the absorption spectra of the polymorphs is to be able to uncover similarities. Those similarities can subsequently be utilized in 1D infrared absorption experiments to tell apart 2N0A, 6H6B and 8ADV The absorption spectra of the three polymorphs follow in the next graph.

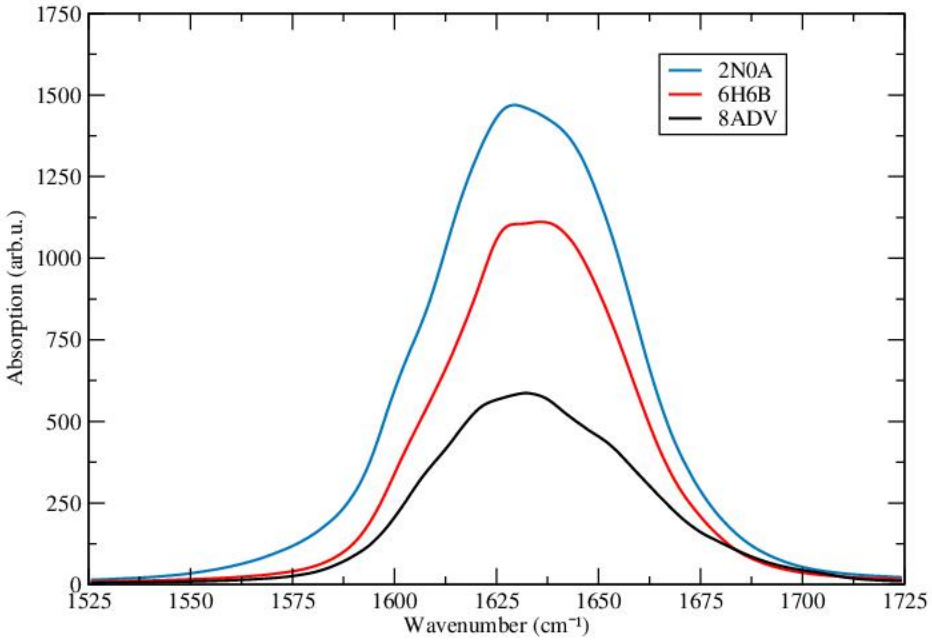


Figure 13: This graph shows the full absorption spectrum of 2N0A, 6H6B and 8ADV. The wavenumbers are plotted on the x-axis in cm^{-1} (per centimeter) and the absorption on the y-axis in arb.u. (arbitrary units).

Figure 13 shows the absorption spectra of the polymorphs quite well. However, it should be noted that 6H6B and 8ADV are not full structures, but instead part of structures. Containing only residues 38-95 and 1-99 out of 1-140 residues respectively. All differences could simply be because of a difference in residues considered and might not be influenced by the different tertiary and quaternary structures This means that to determine whether or not the real structures differ substantially in their infrared absorption spectra, the absorption of 2N0A was calculated taking into consideration only residues 38-95 and 1-99.

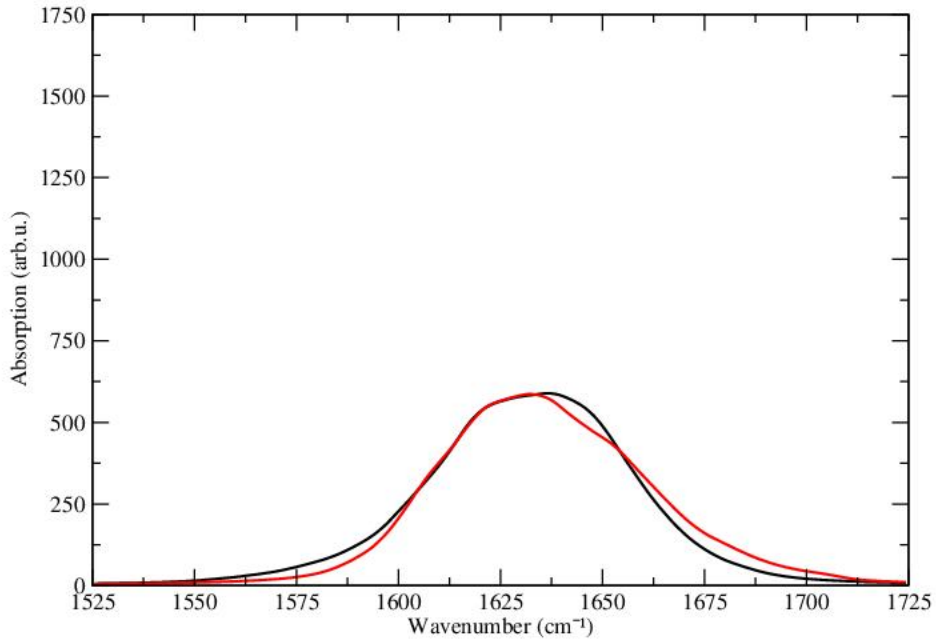


Figure 14: This graph shows the absorption spectrum of 6H6B in red and the absorption spectrum of 2N0A with only the residues 38-95 considered.

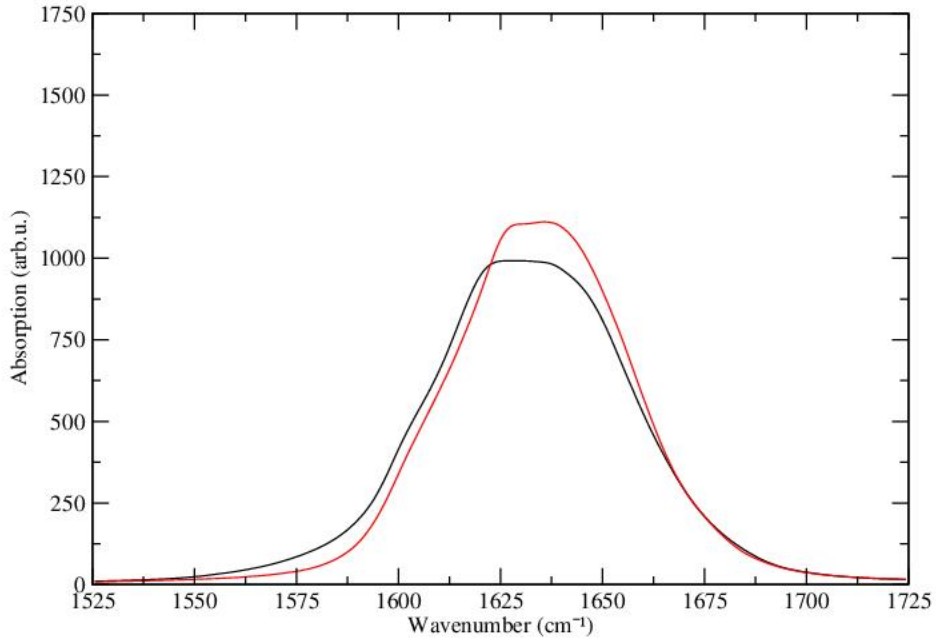


Figure 15: This graph shows the absorption spectrum of 8ADV in red and the absorption spectrum of 2N0A with only the residues 1-99 considered.

Figure 14 shows that the absorption spectra are very similar. The most notable difference is that the absorption spectrum of 2N0A appears far more Gaussian than that of 6H6B. That means that one could potentially distinguish between 2N0A and 6H6B based on an experimentally obtained infrared absorption spectrum by looking at the shape of the absorption spectrum.

Figure 15 shows that the absorption spectrum of 8ADV has a less sharp peak than the absorption spectrum of 2N0A. 2N0A is also shifted slightly to the left. One could distinguish between 2N0A and 8ADV by taking an experimentally obtained infrared absorption spectrum and investigating the width of the peak and wavenumber around which the peak is centered.

Inspection of the structure of 2N0A shows that the chains in the structure contain 3 sections. The first is a loose tail containing amide groups 1-38 that can bend and change shape containing residues, a second section that is more solid containing amide groups 49-110 and a third section that is also loose containing the amide groups 111-148. The first and third sections are all in different shapes, which might cause very different absorption spectra for each. This raises the question of whether or not 2N0A is representative of all fibrils of its shape since 10 chains might not be enough to converge on a proper average, which should happen in a physical system. The absorption spectra are graphed below in figures 16 and 17 for the first and third sections containing each chain separately. By examining these graphs one could get a sense if the first and third sections are indeed representative.

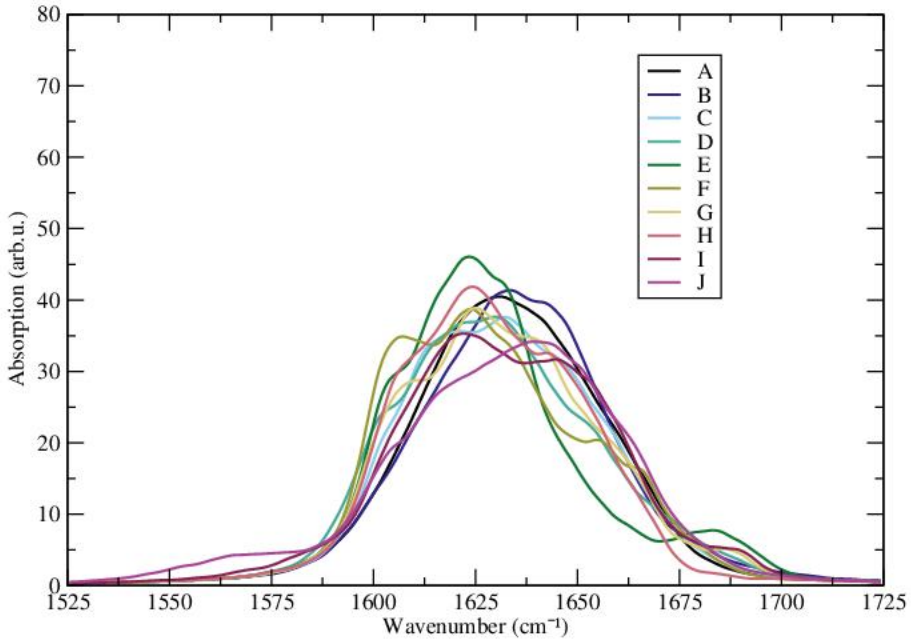


Figure 16: This figure shows the absorption spectra of the first section of all chains in 2N0A, consisting of amides 1-38 in each chain.

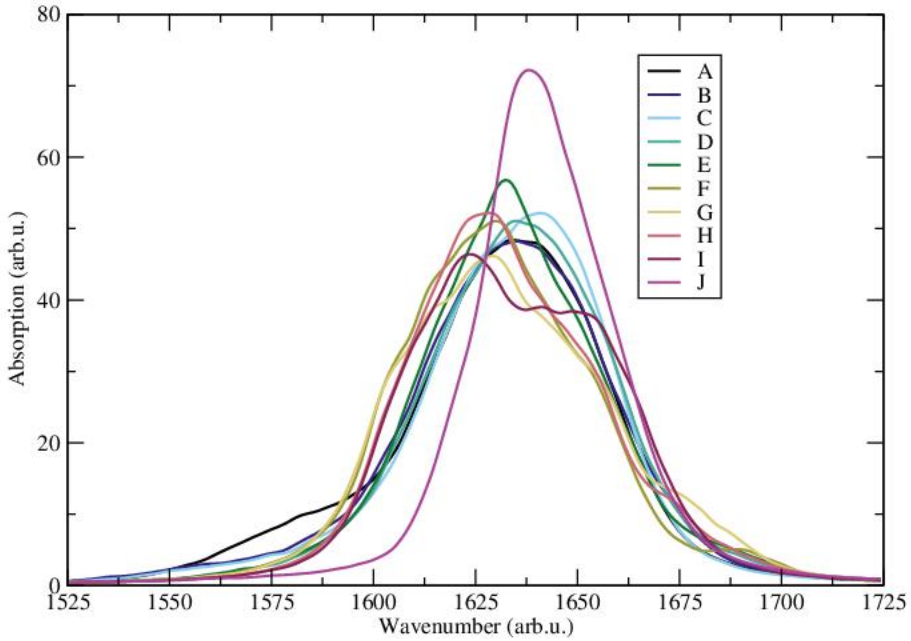


Figure 17: This figure shows the absorption spectra of the third section of all chains in 2N0A, consisting of amides 111-148 in each chain.

In figures 16 and 17, it can be seen that there is variation in the absorption spectra of all of the sections in the different chains. Whilst there are some outliers, most of the spectra in each group are very similar. There is noticeably less variation between the spectra shown in figure 16 than in figure 17. Chain J looks very peculiar, it seems to have a much sharper peak than the other chains. Visual inspection of section 3 of chain J doesn't show any obvious differences that might be the cause of this, however.

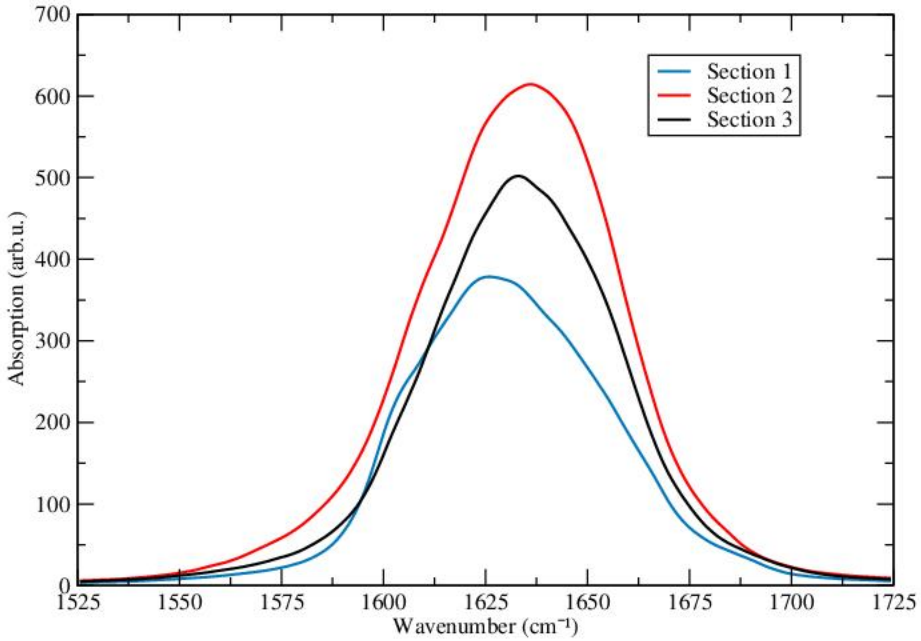


Figure 18: This figure shows the absorption spectra of the three sections of 2N0A, showing section 1 containing amides 1-38, section 2 containing amides 39 - 110, and section 3 containing amides 111-148.

From figure 18 it is evident that the difference between sections is very small despite the presence of β-sheets in section 2 and the random coils in sections 1 and 3. This highlights the limitations of infrared absorption spectra and could indicate the need for alternative methods.

5 Discussion

The Hamiltonians of all three polymorphs show that the couplings between amide groups on adjacent chains inside of the same row are very strong, so these couplings are important for the absorption spectra. This is shown in figures 10, 11 and 12. The polymorphs that were investigated only contain 10 chains. This is very short. 2N0A, which is the longest polymorph of α-synuclein investigated has a length of less than 50 nm [2], compared to physical α-synuclein fibrils that can easily be longer than 10 μm [6]. These two facts combined mean that these edge effects have a strong influence on the infrared absorption spectrum which could have caused a deviation from physical systems. This effect should be especially strong on 6H6B and 8ADV, as these polymorphs only contain rows that are 5 chains long, which is only half the length of 2N0A, which is 10.

The lack of contact between the two rows in 8ADV as shown in figure 5 could be causing deviation in the absorption spectrum of the polymorph. This would probably not be because of a lack of coupling between rows but more so because the polymorph can attain slightly different shapes because of the extra room created by a lack of opposite chains. The missing coupling across rows is probably not very relevant because the coupling across rows in 8ADV is very weak regardless.

Comparing figure 13 with results from other research shows some differences. A paper by Saxena et al. in *The Journal of Chemical Physics* [30] computationally examined amylin fibrils using amylin's infrared absorption spectrum. The amylin fibrils contain many stacked β-sheets and thus should have similar absorption spectra as the examined α-synuclein fibrils. Figure 2 in their paper shows that the amylin structures absorb light most strongly at a wavenumber around 1628 cm^{-1}. This corresponds to the results in figure 13, although the peaks in figure 13 are broader. The paper by Saxena also highlights a high-frequency side band that peaks at 1700 cm^{-1} in one of the amylin fibrils. This does not show up in any of the spectra in figure 13. There is no obvious reason for this.

In another paper by Roeters et al. in *Scientific Reports*[6], the infrared absorption spectrum of α-synuclein fibrils was measured among other things. For all measured fibrils, Roeters found a major absorption peak attributed to β-sheet structure at 1620 cm^{-1}, and a minor absorption peak attributed to coil secondary structure at 1657 cm^{-1} as is shown in figure 2A of the paper. The major peak has a slightly lower wavenumber than the absorption peak in figure 13. The minor peak is not represented in figure 13 though. This could mean that either the software used in this thesis does not model coil secondary structure well or that coils are relatively underrepresented in 2N0A, 6H6B and 8ADV. If the coils are merely relatively underrepresented the absorption spectra in this thesis are still accurate. If there is an issue with the software that causes coils to not be modeled properly the absorption spectra in this thesis could still be compared to experimental data if the contributions from coils were to be removed from the experimental data.

The variance between the absorption spectra of section 1 in all the different chains in 2N0A as shown in figure 16 is visible, however, the error this induces is into the absorption spectra is likely not very big. The wavenumbers of the peaks vary by less than one percent from 1630 cm^{-1}. The heights of the peaks are spread out evenly between 35 and 47 arb.u., which means all peaks are within 15 percent of a central value of 41 arb.u..

Absorption spectra of section 3 of 2N0A are displayed in figure 17. The spectra have their highest peaks between 1625 and 1645 cm^{-1}, which means that they vary by less than one percent from 1635 cm^{-1}. Their peaks vary between 50 and 55 and arb.u. with one outlier having its peak at 75 arb.u. from chain J. Excluding the outlier, all peaks are within a 5 percent of a central value of 52.5 arb.u.. If the outlier is included, however, this jumps to within 20 percent of a central value of 62.5 arb.u.. The errors caused by variation in sections 1 and 3 are likely not substantial enough to warrant discarding any conclusions based on any of the absorption spectra of 2N0A in this thesis.

The absorption spectrum of 2N0A as displayed in figure 13 is likely fairly accurate, however, it may not be good enough to distinguish between very similar polymorphs because the difference between polymorphs could be comparable to the potential errors.

The absorption spectra of 6H6B and 8ADV are displayed in figure 13. These structures are not complete and thus should never be expected to accurately resemble the absorption spectra of the complete polymorphs.

However figures 14 and 15 do highlight that subtle differences between even the infrared absorption spectra exist and those subtle differences could potentially persist even when the full structure is examined, however utilizing isotope labeling would allow for better distinction of various α-synuclein fibrils.

Labeling areas of proteins with ^{13}C allows for distinction between the labeled and unlabeled parts in infrared absorption spectra [31]. If one were to label only the parts of α-synuclein that are not contained in either 6H6B or 8AD, it becomes possible to filter out the computed absorption spectrum obtained in this thesis out of experimentally obtained absorption spectra of α-synuclein fibrils. By matching the filtered-out absorption spectrum to a spectrum discovered in this thesis, one could pinpoint the exact α-synuclein fibril being experimented upon.

Labeling of the side chains of amino acid residues in α-synuclein could allow for the isolation of the side chain absorption spectrum. α-synuclein only contains 9 amino acid residues that have an amide group in their side chain that are spread out throughout the protein. By isolating the absorption of those 9 sidechains a new unique spectrum is attained. If one has computational absorption spectra one would have a secondary way of comparing experimental and computational results.

Figures 14 and 15 can tell us if the absorption spectra of 2N0A are distinguishable from those of 6H6B and 8ADV. As discussed earlier, there are several sources of error influencing the absorption spectra, which are not precisely quantified. This means that 2N0A and 6H6B are probably not easily differentiable based solely on their infrared absorption spectrum. 2N0A and 8ADV however are likely much more easily distinguishable based on their infrared absorption spectra as there is a substantial difference in the height of the absorption peak.

Figure 18 shows that differences in secondary structure do not guarantee a large difference in infrared absorption spectra. To obtain more detailed graphs that would allow for better distinction, 2DIR or 2DIRRaman absorption could be used instead of infrared absorption. 2DIR results in graphs that reveal more structural information which would help differentiate secondary structures, with 2DIRRaman spectra being broader [28]. Whilst 2DIR and 2DIRRaman NISE calculations take much more computing time than 1D NISE calculations, it would significantly help distinguish the polymorphs.

This thesis is a necessary step for the comparison between α-synuclein polymorphs. The thesis shows that the infrared absorption spectra of α-synuclein polymorphs are potentially distinct enough to distinguish them. It also highlights that the use of isotope labeling, 2DIR or 2DIRRaman would make distinguishing the polymorphs much easier. It would be unwise to use 2DIR or 2DIRRaman without first knowing that infrared absorption spectra are of limited use due to the large amount of computing hours these techniques require.

6 Conclusion

This thesis was performed to provide insight into the infrared absorption spectrum of the 2N0A, 6H6B and 8ADV α-synuclein polymorphs. This is a worthwhile goal because these different aggregates may play a different role in the pathological pathways of Parkinson's disease, which is a neurodegenerative disorder characterized by Lewy bodies.

The structures of these aggregates were taken from the Protein Data Bank. Notably, only 2N0A is a complete structure with 6H6B and 8ADV being only partial structures. These structures all contained 10 chains with 2N0A having all chains in one row and 6H6B and 8ADV having 2 rows with 5 chains each.

To be able to better understand why the infrared absorption spectra are different and the coupling of which amide groups contribute significantly to the absorption spectra, the Hamiltonians of all the examined structures were investigated. For all structures, the most important couplings are those between adjacent amide groups and with nearby amide groups on adjacent chains. The couplings across rows are relatively minor. For 6H6B the coupling across rows has an impact on the absorption spectrum, however for 8ADV, this is not the case. 8ADV however might be able to attain slightly different shapes than is possible due to its twisted shape.

The absorption spectra are of varying usefulness. The graph of 2N0A is likely fairly close to reality as there are no significant sources of error and 2N0A contains a complete polymorph. Since 6H6B and 8ADV are incomplete structures, their absorption spectra cannot be directly compared to the absorption spectra of physical structures, however, isotope labeling could enable this direct comparison to still take place.

Comparison between 2N0A and the other two polymorphs was made possible by only considering the parts of 2N0A that it has in common with 6H6Band 8ADV when calculating the absorption spectrum. When 2N0A is compared to 6H6B the difference between the two spectra is very small. This means that it is likely not very easy to distinguish them with their infrared absorption spectrum. 2N0A and 8ADV are more easily differentiable, however. One could likely tell them apart with only their infrared absorption spectrum, it might be necessary to remove the absorption peak caused by coils from the experimental data however, as the comparison between the results in this thesis and the experimental results of a paper by Reuters et al. shows that the absorption peak of the coils is absent from the figures shown in this thesis. The results from this thesis do correspond well to the computational results of a paper by Saxena et al.

Whilst it is possible to differentiate the polymorphs based on their infrared absorption spectra, using 2DIR or 2DIRRaman would be a better way to distinguish the different structures, as well as provide more structural information. However, these protein aggregates are very large structures. The entire computation to create the infrared absorption spectra from a .pdb file to a graph took approximately 3 days of computation with 64 cores at 2.45 GHz. These more detailed computations would take a very long time to complete, even with powerful hardware.

Furthermore, if knowledge of the precise absorption spectra is desired, more advanced methods of computing these spectra would do nothing if the complete polymorphs are not known.

Another improvement on this thesis would be to make the structures closer in size to their physical counterparts. The scaling in necessary computing power for larger structures also makes this difficult, however, the improved results might make it worth it. The combination of larger structures with 2DIRRaman would lead to an even larger increase in necessary computing hours.

7 Acknowledgments

I would like to thank Thomas L.C. Jansen and Kim E. van Adrichem for their feedback and guidance for this research. I would also like to thank Steven J. Roeters for his useful suggestions.

8 References

[1] E. Srinivasan et al. "Alpha-Synuclein Aggregation in Parkinson's Disease". In: *Frontiers in Medicine* 8 (2021), p. 736978.

[2] Marcus D. Tuttle et al. "Solid-state NMR structure of a pathogenic fibril of full-length human alpha-synuclein". In: *Nature Structural and Molecular Biology* 23 (2016), pp. 409–415.

[3] B. Frieg et al. "The 3D Structure of Lipidic Fibrils of Alpha-Synuclein". In: *Nature Communications* 13 (2022), p. 6810.

[4] David Sulzer and Robert H. Edwards. "The Physiological Role of Alpha-Synuclein and its Relationship to Parkinson's Disease". In: *Journal of Neurochemistry* 150 (5 2019), pp. 475–486.

[5] Tobias S. Ulmer et al. "Structure and Dynamics of Micelle-Bound Human Alpha-Synuclein". In: *Journal of Biological Chemistry* 280 (2005), pp. 9595–9603.

[6] Steven J. Roeters et al. "Evidence for Intramolecular Antiparallel Beta-Sheet Structure in Alpha-Synuclein Fibrils from a Combination of Two-Dimensional Infrared Spectroscopy and Atomic Force Microscopy". In: *Scientific Reports* 7 (2017), p. 41051.

[7] Ricardo Guerrero-Ferreira et al. "Cryo-EM Structure of Alpha-Synuclein Fibrils". In: *eLIFE* 7:e36402 (2018).

[8] Helen M. Berman et al. "The Protein Data Bank". In: *Nucleic Acids Research* 28 (1 2000), pp. 235–242.

[9] William Humphrey, Andrew Dalke, and Klaus Schulten. "VMD: Visual in Molecular Dynamics". In: *Journal of Molecular Graphics* 14 (1 1996), pp. 33–38.

[10] L. Wang, C.T. Middleton, and J.L. Skinner. "Development and Validation of Transferable Amide I Vibrational Frequency Maps for Peptides". In: *The Journal of Chemical Physics* 115 (2011), pp. 3713–3724.

[11] Mike Reppert and Andrei Tokmakoff. "Computational Amide I 2DIR Spectroscopy as a Probe of Protein Structure and Dynamics". In: *Annual Review of Physical Chemistry* 67 (1 2016), pp. 359–386.

[12] Justin P. Lomont et al. "Not All Beta-Sheets Are the Same: Amyloid Infrared Spectra, Transition Dipole Strengths, and Couplings Investigated by 2D IR Spectroscopy". In: *Journal of Physical Chemistry* 121 (38 2017), pp. 8935–8945.

[13] Thomas L.C. Jansen. "Computational Spectroscopy of Complex Systems". In: *The Journal of Chemical Physics* 155 (170901 2021).

[14] Carloz R. Baiz et al. "Vibrational Spectroscopic Map, Vibrational Spectroscopy, and Intermolecular Interaction". In: *Chemical Reviews* 120 (15 2020), pp. 7152–7218.

[15] Bruce Alberts et al. *Molecular Biology of the Cell*. 6th. Garland Science, 2015. Chap. 3.

[16] Mark A. James et al. "GROMACS: High Performance Molecular Simulations through Multi-level Parallelism from Laptops to Supercomputers". In: *SoftwareX* 1 (2 2015), pp. 19–25.

[17] H.J.C. Berendsen, J.R. Grigera, and T.P. Straatsma. "The Missing Term in Effective Pair Potentials". In: *The Journal of Chemical Physics* 91 (1987), pp. 6269–6271.

[18] William L. Jorgensen and Julian Tirado-Rives. "The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin". In: *Journal of the American Chemical Society* 110 (6 1988), pp. 1657–1666.

[19] Ulrich Essmann et al. "A Smooth Particle Mesh Ewald Method". In: *The Journal of Chemical Physics* 103 (19 1995), pp. 8577–8593.

[20] Kim E. van Adrichem and Thomas L. C. Jansen. "AIM: A Mapping Program for Infrared Spectroscopy of Proteins". In: *Journal of Chemical Theory and Computation* 18 (5 2022), pp. 3089–3098.

[21] Thomas L. C. Jansen et al. "Modeling the Amide I Bands of Small Peptides". In: *The Journal of Chemical Physics* 125 (4 2012).

[22] Mitsuo Tasumi and Hajime Torii. "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". In: *Journal of Raman Spectroscopy* 29 (1 1998), pp. 81–86.

[23] Thomas L.C. Jansen and Jasper Knoester. "Nonadiabatic Effects in the Two-Dimensional Infrared Spectra of Peptides: Application to Alanine Dipeptide". In: *The Journal of Physical Chemistry* 110 (45 2006), pp. 22910–22916.

[24] Thomas L.C. Jansen and Jasper Knoester. "Waiting Time Dynamics in Two-Dimensional Infrared Spectroscopy". In: *Accounts of Chemical Research* 42 (9 2009), pp. 1405–1411.

[25] Thomas L.C. Jansen et al. "Two-Dimensional Infrared Spectroscopy and Ultrafast Anisotropy of Water". In: *The Journal of Chemical Physics* 132 (22 2010), p. 224503.

[26] Chungwen Liang and Thomas L.C. Jansen. "An Efficient N3-Scaling Propagation Scheme for Simulating Two-Dimensional Infrared and Visible Spectra". In: *Journal of Chemical Theory and Computing* 8 (5 2012), pp. 1706–1713.

[27] Chungwen Liang et al. "Vibrational Spectra of a Mechanosensitive Channel". In: *The Journal of Physical Chemistry Letters* 4 (3 2013), pp. 448–452.

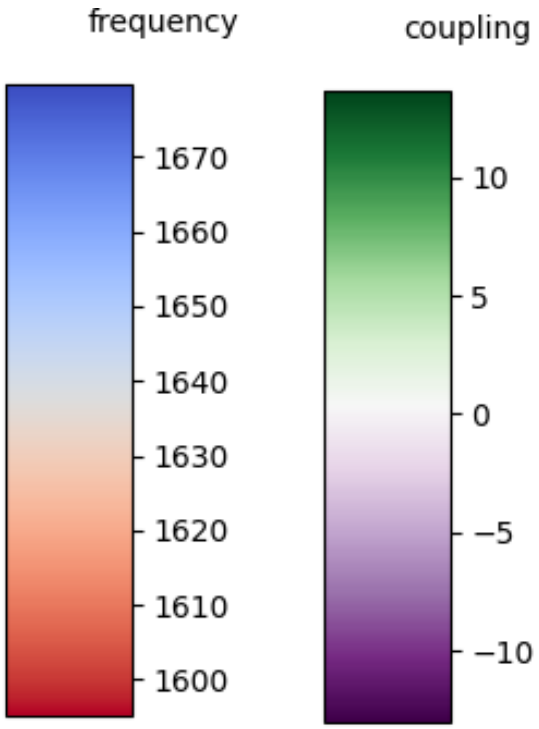
[28] Carleen D.N. van Hengel, Kim E. van Adrichem, and Thomas L. C. Jansen. "Simulation of Two-Dimensional Infrared Raman Spectroscopy with Application to Proteins". In: *The Journal of Chemical Physics* 158 (6 2023).

[29] Hong-Guang Duan et al. "Two-Dimensional Electronic Spectroscopy of Light-Harvesting Complex II at Ambient Temperature: A Joint Experimental and Theoretical Study". In: *The Journal of Physical Chemistry* 119 (36 2015), pp. 11969–12154.

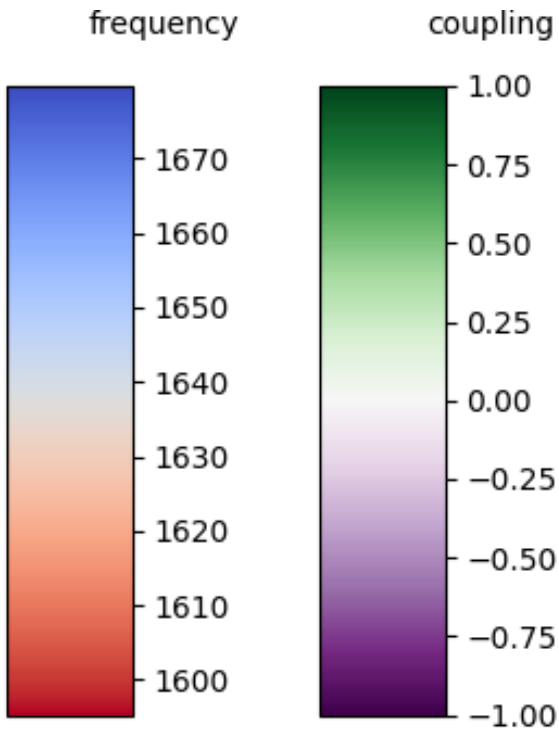
[30] Vishesh Saxena, Ruben Steendam, and Thomas L.C. Jansen. "Distinguishing Islet Amyloid Polypeptide Fibril Structures with Infrared Isotope-Label Spectroscopy". In: *The Journal of Chemical Physics* 156 (5 2021), p. 055101.

[31] Suren A. Tatulian. "Structural Analysis of Proteins by Isotope edited FTIR Spectroscopy". In: *Spectroscopy* 24 (1-2 2010), pp. 37–43.

9 Appendix



(a) Color bars for logarithmic Hamiltonian graph



(b) Color bars for a linear Hamiltonian graph with a reduced range

Figure 19: This figure shows all of the color bars relevant to the Hamiltonians on the following pages. The left part (a) is relevant for logarithmically graphed coupling intensities. The scale works in powers of 2, thus increasing the displayed value in such a graph by 1 indicates a coupling twice as strong. The right part (b) shows color bars relevant for linearly plotted coupling intensities with a smaller range, -1 to 1. Everything outside of the range will be displayed as the closest color inside of this range. All frequencies are in cm^{-1}. The linear coupling strengths are given in cm^{-1}.

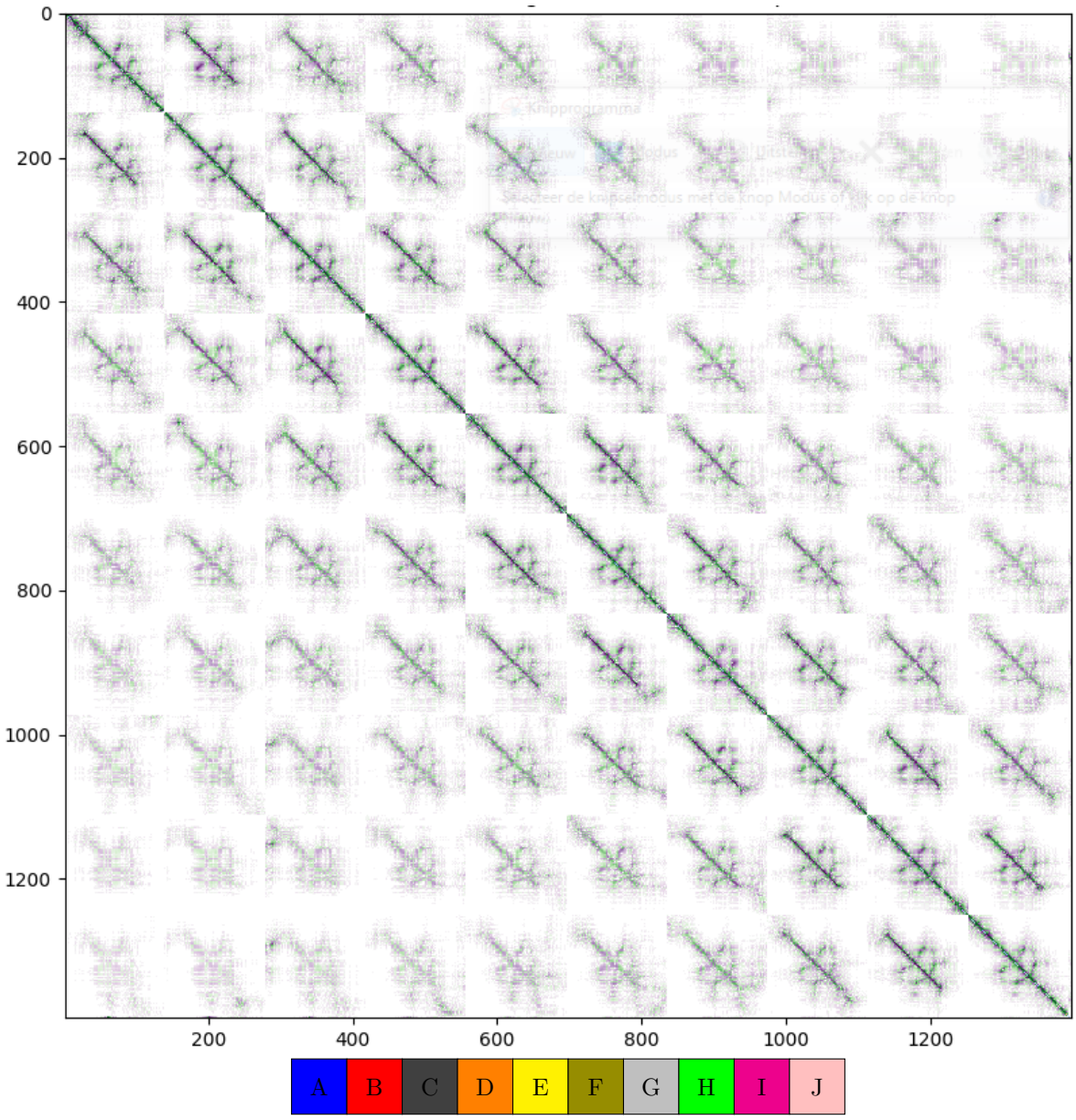


Figure 20: This image shows all the couplings between amide groups between amino acids in 2N0A[2]. Positive couplings are green and negative couplings are purple. The more vivid the color, the stronger the coupling. The coupling strength is plotted logarithmically. The polymorph is displayed schematically below the graph. This image was made by taking the average Hamiltonian of all of the frames. Color bars are shown in figure 19

It can be seen that all chains have some coupling between them, even on the opposite side of the polymorph, even if very weakly. The coupling across chains seems to be more consistently strong on the solid part of the chain. The coupling strength across adjacent tails varies dramatically. This can be explained by the varying shapes of the tail sections of this polymorph.

If we look at chain A and look at the coupling between amides 46 and 74, we discover that it is -9.8. The coupling between amide 46 on chain A and amide 74 on chain B is -5.45. The coupling of 46 and 74 inside of A is 20.39 times stronger than the coupling between amide 46 in chain A and amide 74 in chain B. If we look at the coupling between amide 46 on chain A and amide 74 on chain J this strength is only -2.68, which is 6.8 times weaker. If we look at the coupling between amide 32 on chain A and on chain B, it is -12.51, this is an extremely strong coupling. If we replace chain B with chain J we find a coupling that is only -2.58, which is 976 times weaker. Couplings similar to those mentioned here follow the same trends.

In figure 20 it can be seen when parts of the chain are near each other. The tails of the chain and the solid part all form their sections of relatively strong coupling, with crossover at the edges.

Figure 20 provides an idea of the coupling strength between all considered amide groups, however, it is difficult to visually distinguish the very strongest couplings because of the logarithmic plotting of coupling strength causing a halving of coupling strength to only cause a drop in displayed coupling strength of 1.

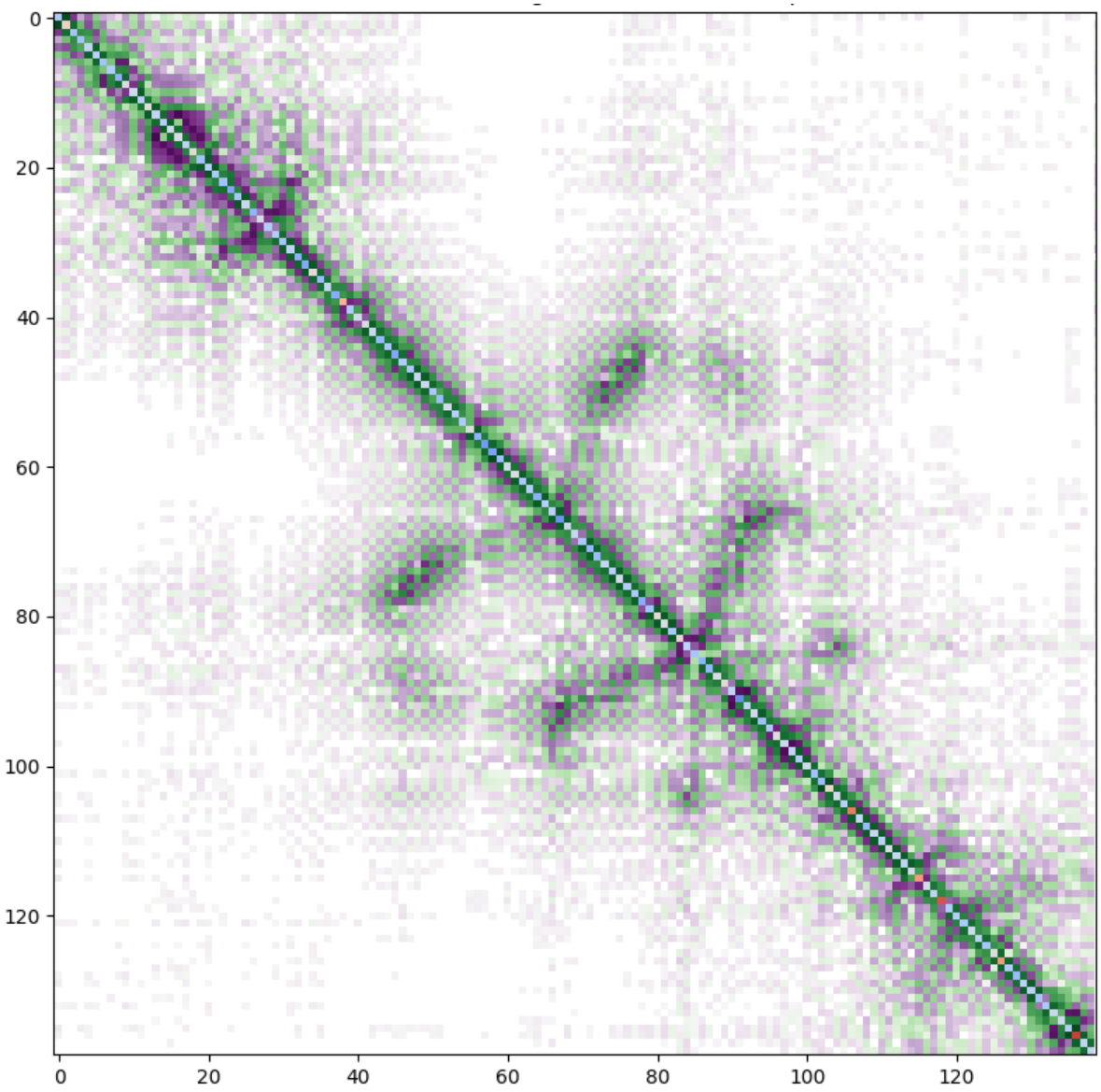


Figure 21: The same as graph as figure 20 but zoomed on A-A. The frequencies on the diagonal can now also be seen. They range between 1550-1700 cm^{-1}

In figure 21 it can be seen when parts of the chain are close to each other. The tails of the chain and the solid part all form their sections of relatively strong coupling, with crossover at the edges. The amides on the border of the sections are part of the tails.

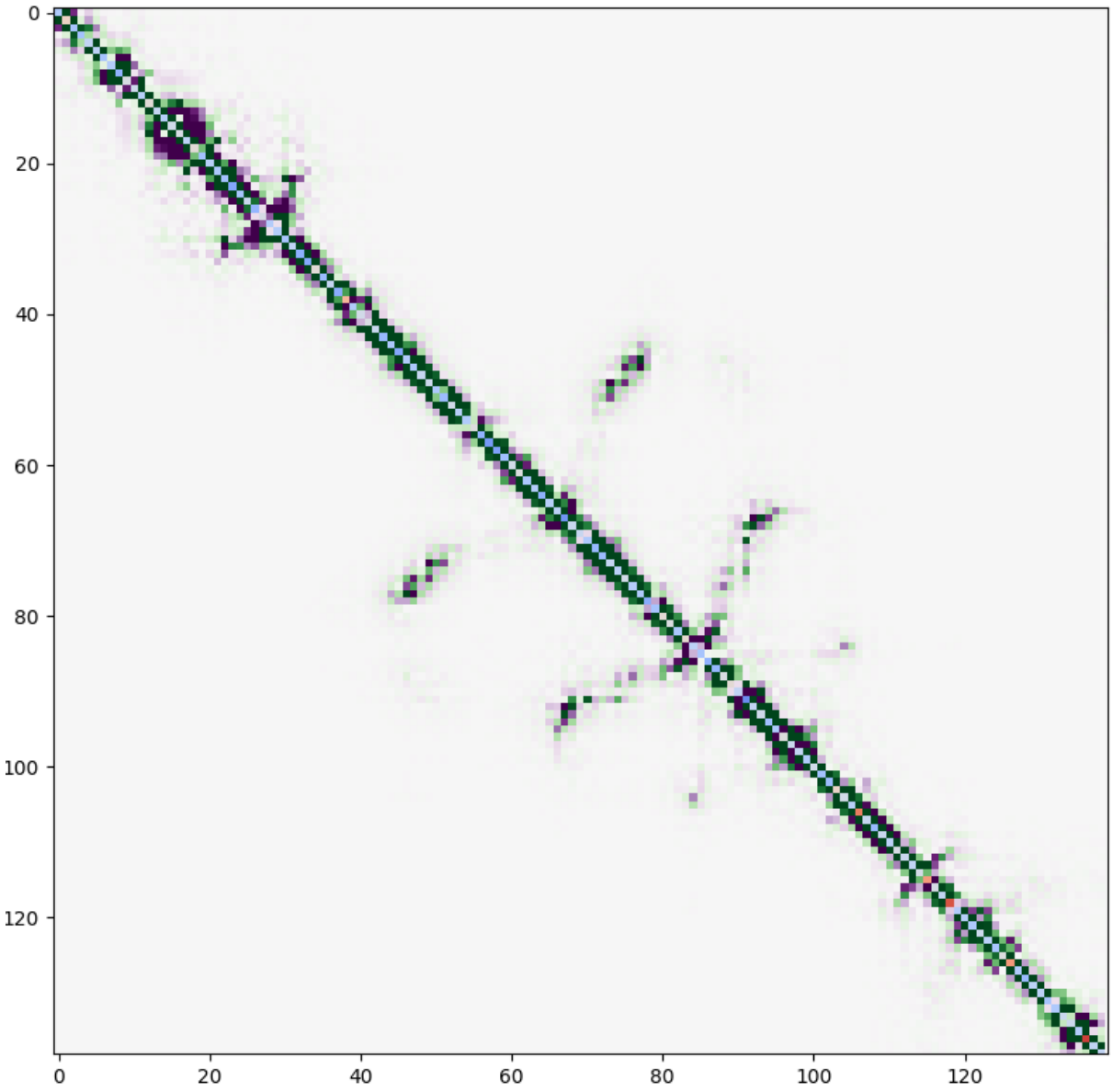


Figure 22: This figure is the same as figure 21, but the intensity of the coupling is plotted linearly and with a low range instead.

It can be seen that there exists a strong coupling between amino acids 45-51 and 72-78, amino acids 67-70 and 90-93 and some amino acids on the first tail section. These couplings are also inconsistently but relatively strong between adjacent chains. When inspecting the structure of 2N0A, these groups of amides with strong coupling are very near each other and on β-sheets.

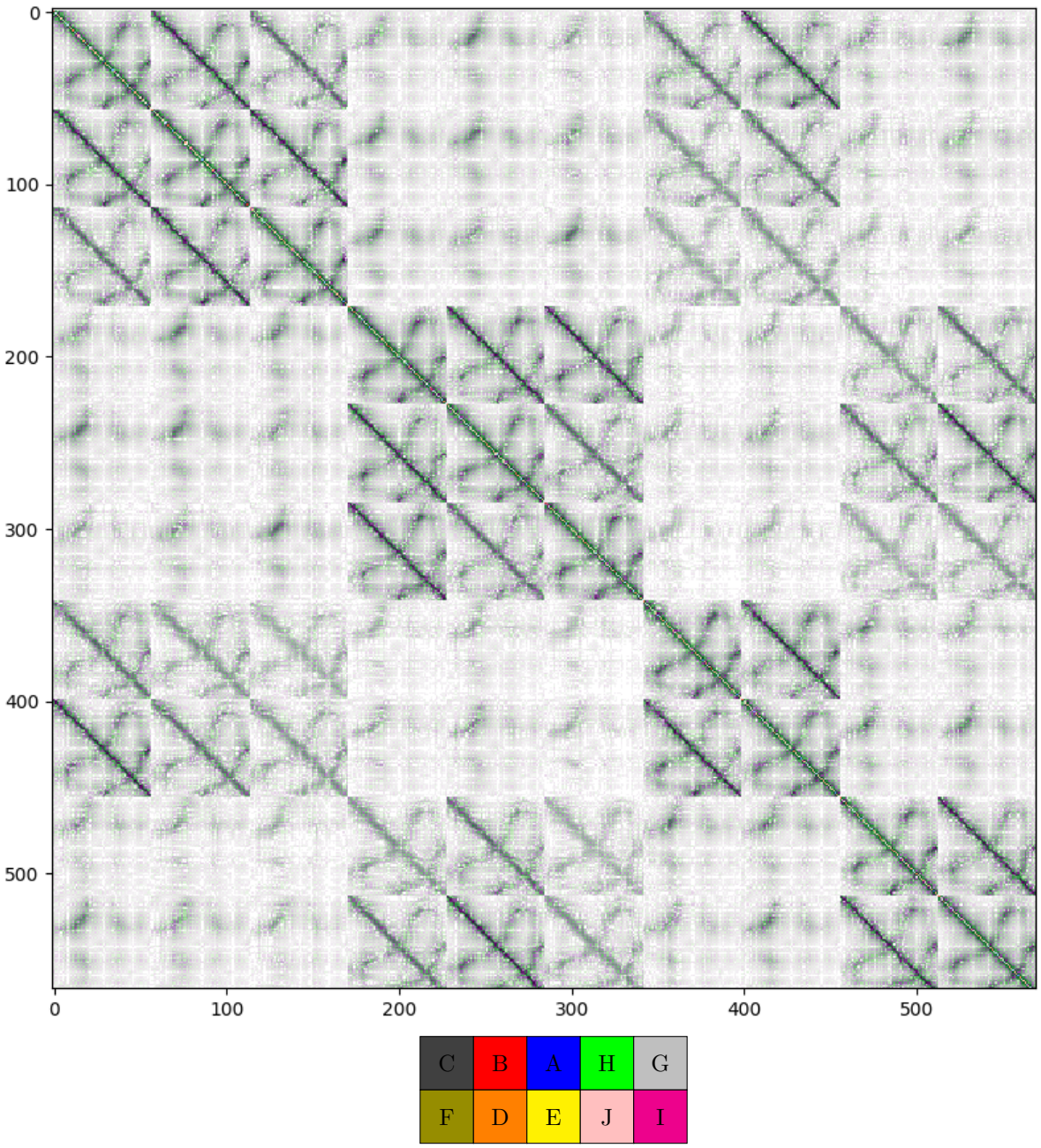


Figure 23: This image shows all the couplings between amide groups between amino acids in 6H6B[7]. Positive couplings are green and negative couplings are purple. The more vivid the color, the stronger the coupling. The coupling strength is plotted logarithmically. The polymorph is displayed schematically below the graph.

Figure 23 shows that the amide groups between chains on 6H6B couple more strongly inside of the same row than outside of it. The amide groups 11-20 couple relatively strongly with each other across the rows, these groups coincide with the location of stacked β-sheets. Whilst in figure 20 all off-diagonal squares resemble more weakly coupled images of on-diagonal squares, in figure 23 the squares that depict coupling across rows are much more unique.

The figure shows that there exists a strong coupling between amide groups 6-18 and 27-43 and between amide groups 24-45 and 47-57, this holds across adjacent chains and within but not across rows. Across adjacent rows, amide groups 11-20 couple most strongly amongst themselves. These correspond to stacked β-sheets.

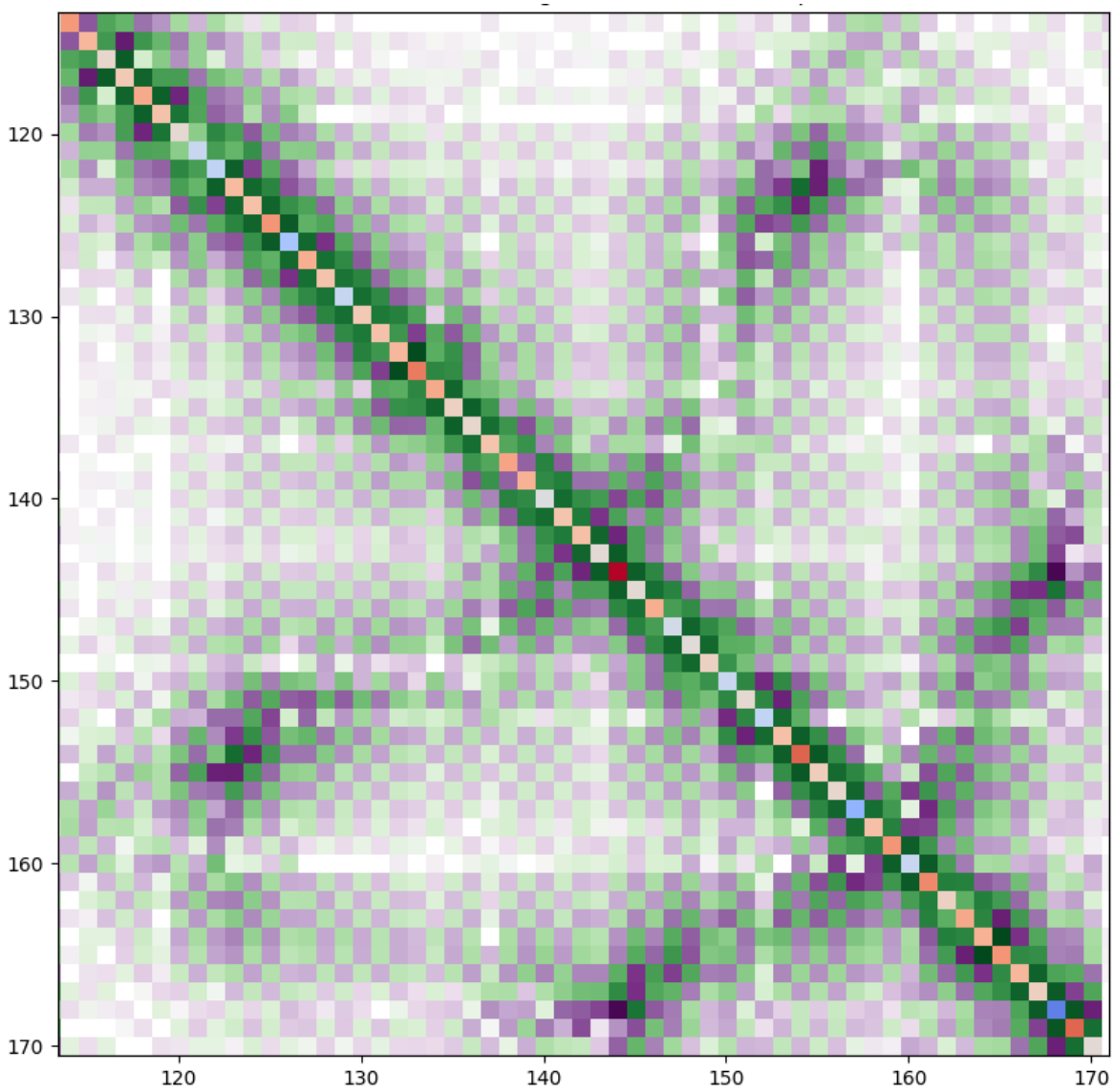


Figure 24: The same as graph as figure 23 zoomed on C-C. The frequencies on the diagonal can now also be seen. They range between 1550-1700 cm^{-1}.

Figure 24 shows strong coupling between amide groups 6-18 and 27-43 and between amide groups 24-45 and 47-57. These correspond to stacked pieces of chain C containing β-sheets.

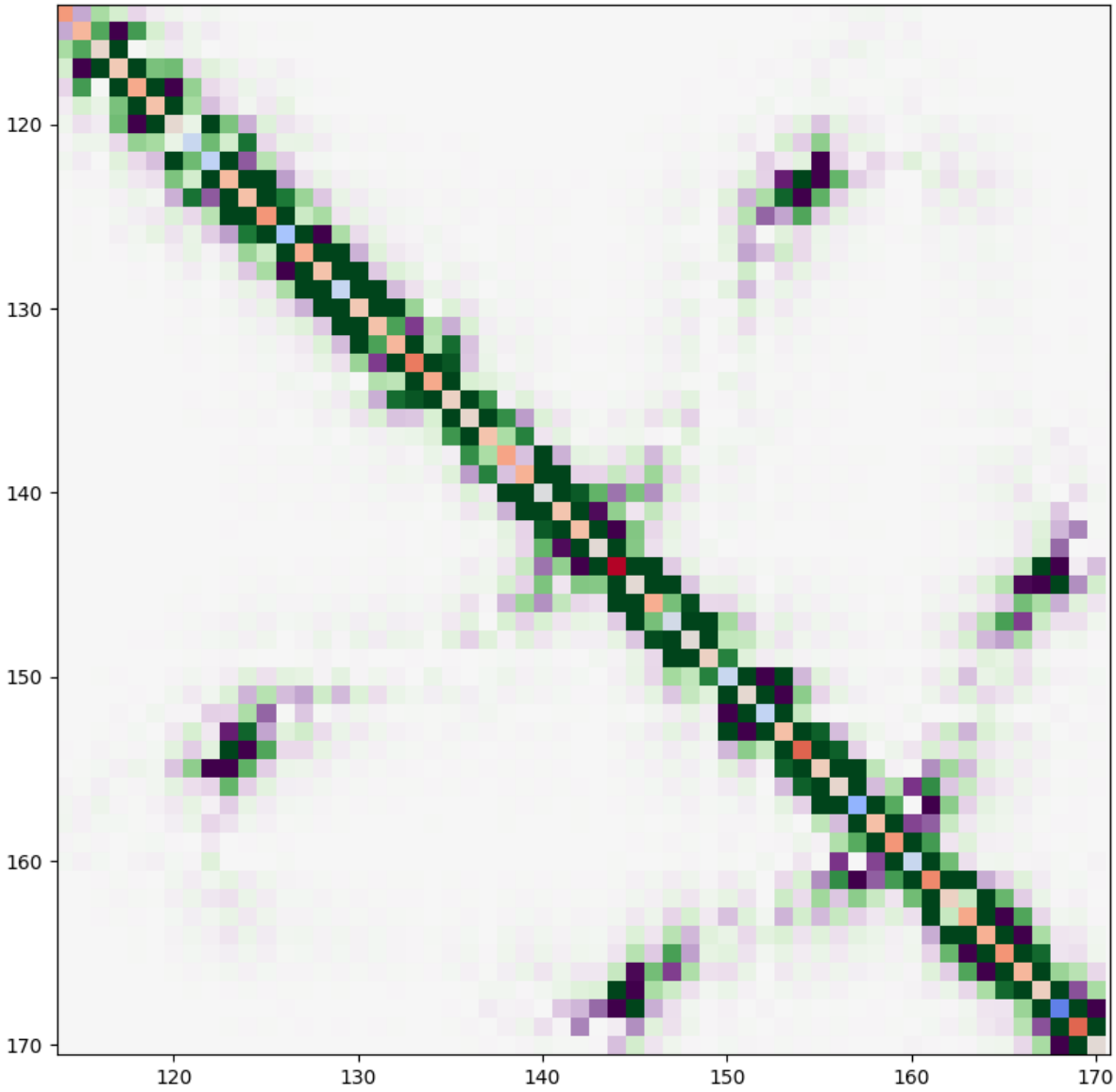


Figure 25: This figure is the same as figure 24, but the intensity of the coupling is plotted linearly and with a low range instead.

Figure 25 shows that the strongest coupling is between amide groups 9-12 and 39-41 and amide groups 30-34 and 52-55. These groups are also relatively strongly coupled across chains as long as they are in the same row.

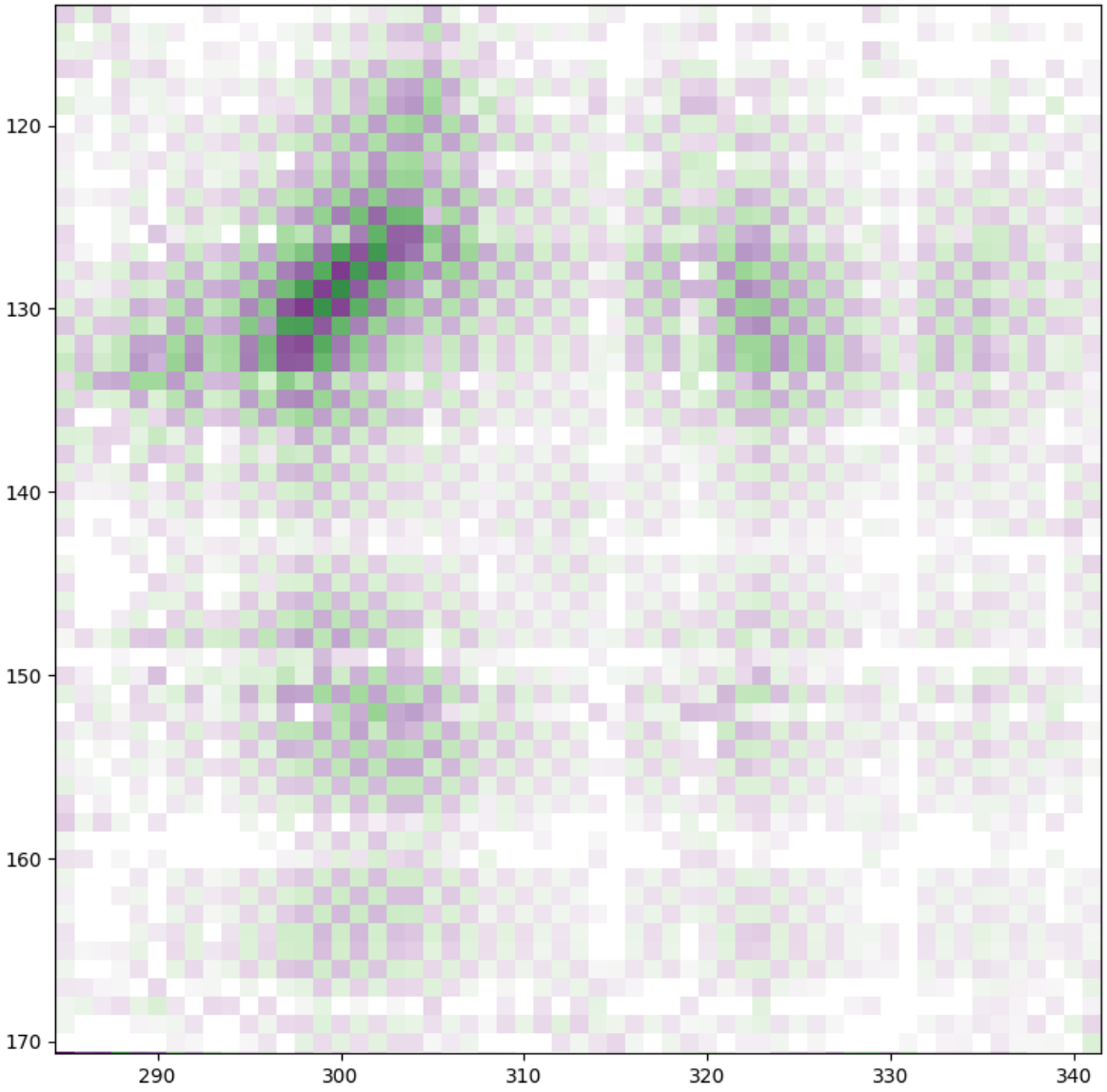


Figure 26: The same as graph as figure 23 zoomed on C-F. Because this square is not on the diagonal the frequencies cannot be seen.

Amide groups 11-20 interact most strongly among themselves in figure 26. To get a better sense of the relative strength of the couplings there is a figure below showing the same graph as figure 23 but with the strength of the dipole couplings plotted linearly.

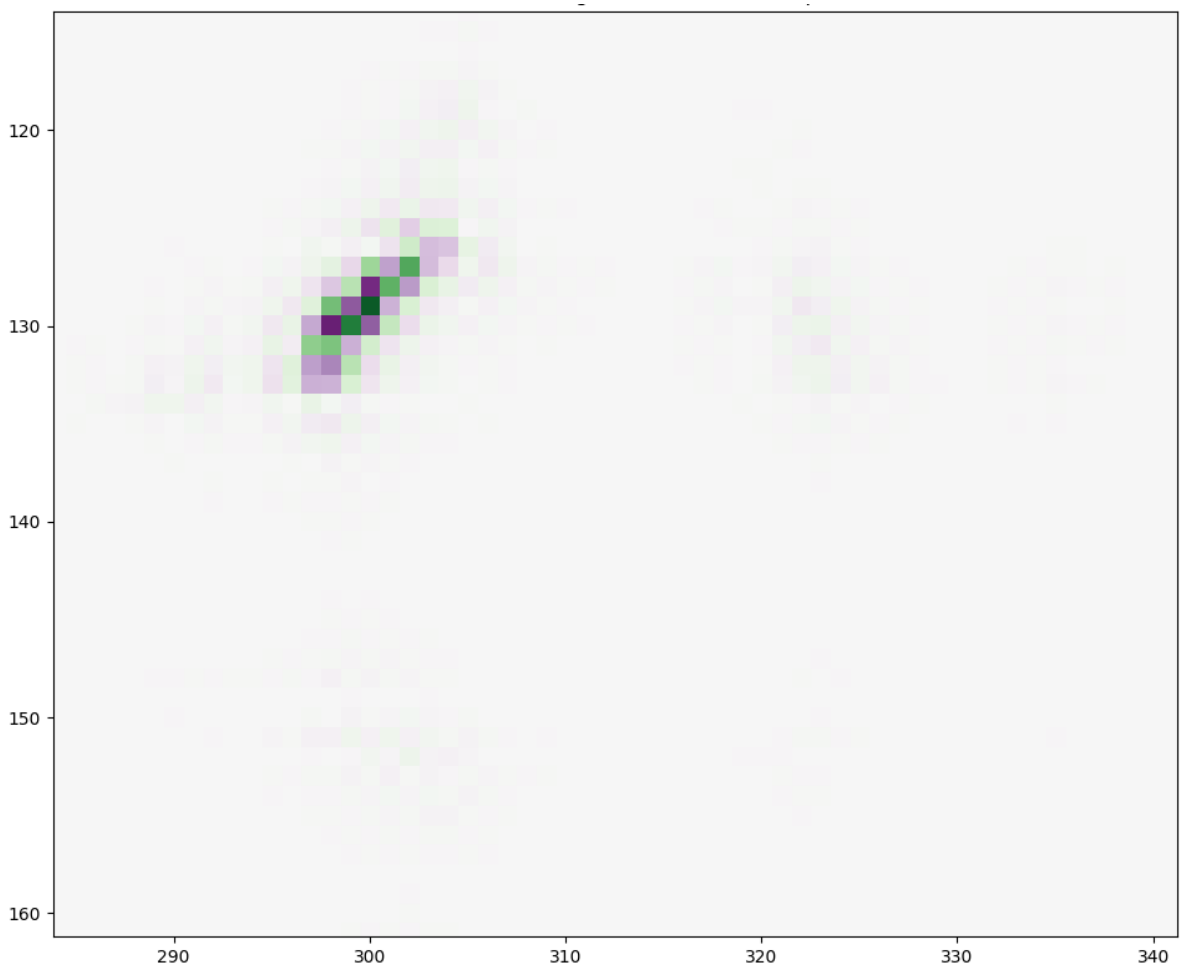


Figure 27: This figure is the same as figure 26 but with the intensities of the coupling displayed linearly.

Figure 27 shows more precisely that the only couplings that are relevant are between the amide groups 14-20 and 13-18. These correspond to the β-sheets that form a bridge between the two rows on opposite chains.

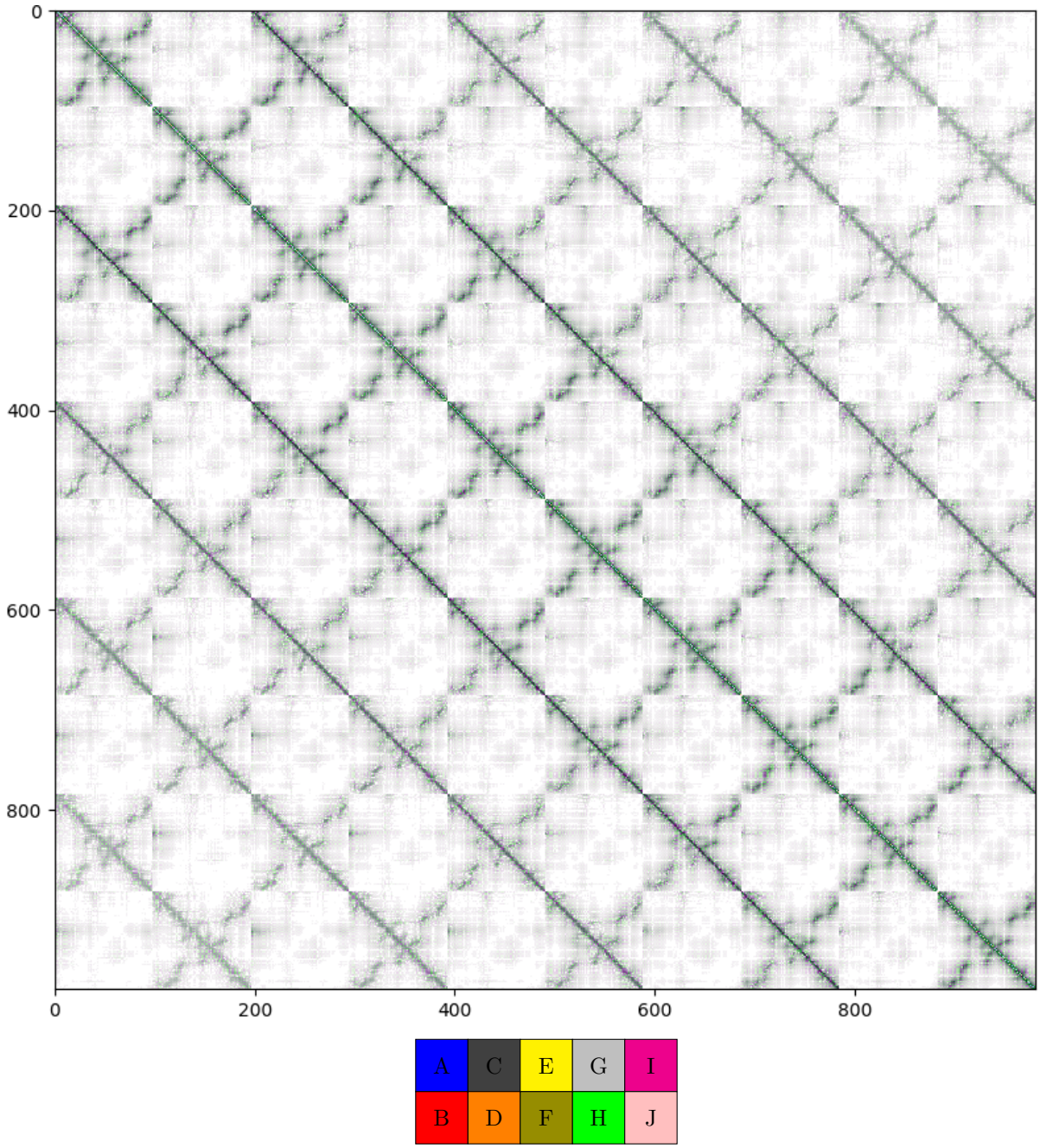


Figure 28: This image shows all the couplings between amide groups between amino acids in 8ADV[3]. Positive couplings are green and negative couplings are purple. The more vivid the color, the stronger the coupling. The coupling strength is plotted logarithmically. The polymorph is displayed schematically below the graph. This image was made by taking the average Hamiltonian of all of the frames.

It is evident from figure 28 that the coupling between rows is a lot weaker in 8ADV[3] than it is in 6H6B[7]. Observation of the structure reveals that this makes a lot of sense as the bridges between the rows are very narrow in 8ADV[3] whilst the connection is very broad for 6H6B[7].

Figure 28 looks very different from the solid section of figure 20 or 23. It is noticeable that amide groups in the middle area of the chain are very weakly coupled to the end and beginning of the chain. There is particularly strong coupling between amide groups 22-33 and 73-81, amide groups 11-20 and 87-97 and amide groups 48-54 and 68-75.

The figure also shows that the coupling across rows is very weak, even with the exact opposite chain across a row. There are two regions of relatively strong coupling across chains, namely around the coupling of 1-40 and 40-1. Amide groups 1 and 40 form the narrow connection between the two rows.

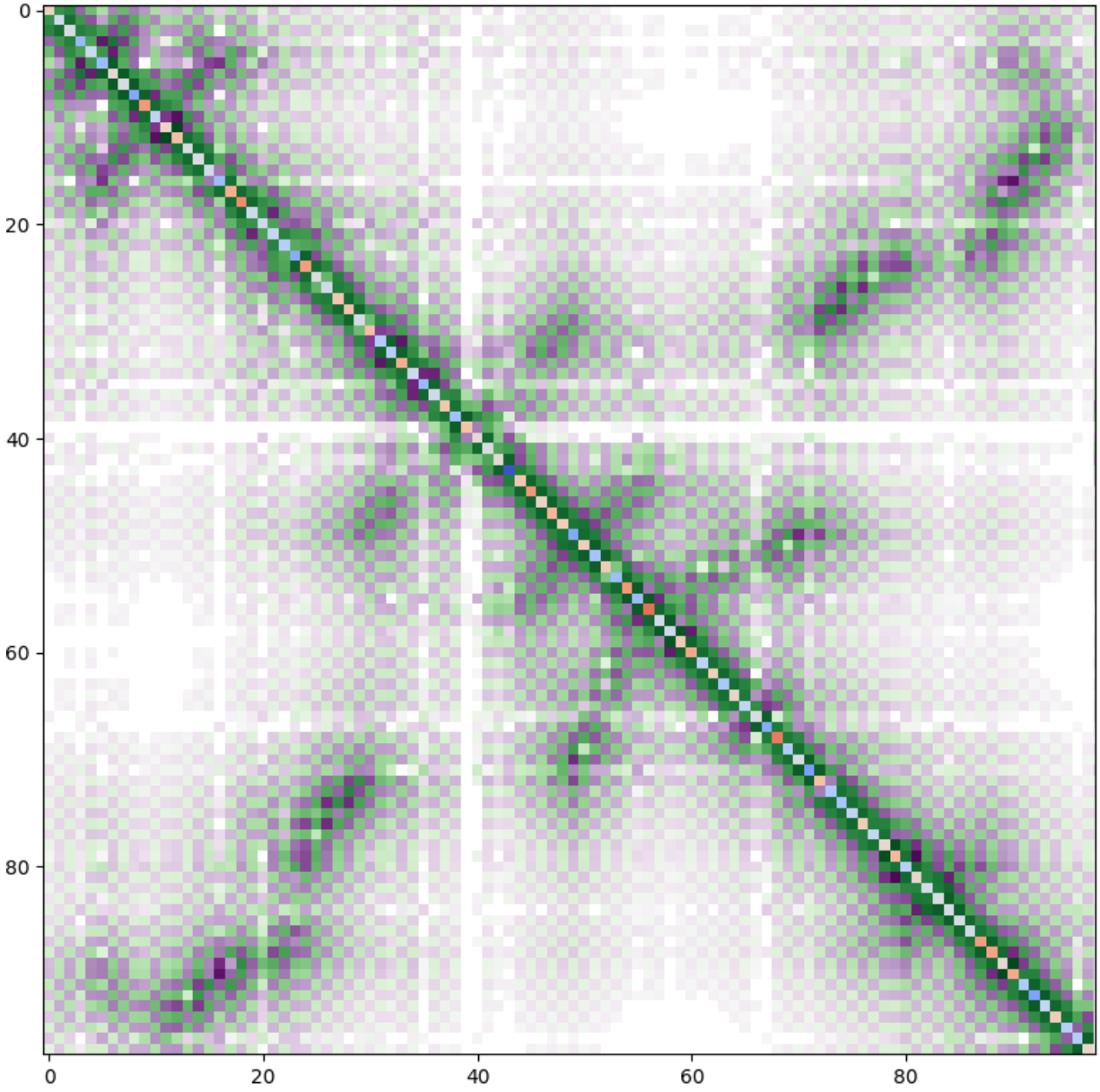


Figure 29: The contains the same as graph as figure 28 but zoomed on A-A. The frequencies on the diagonal can now also be seen. They range between 1550-1700 cm^{-1}.

Figure 29 looks very different from the solid section of figure 21 or 24. It is noticeable that amide groups in the middle section of the chain are very weakly coupled to those closer to the end and beginning of the chain. There is particularly strong coupling between amide groups 22-33 and 73-81, amide groups 11-20 and 87-97 and amide groups 48-54 and 68-75.

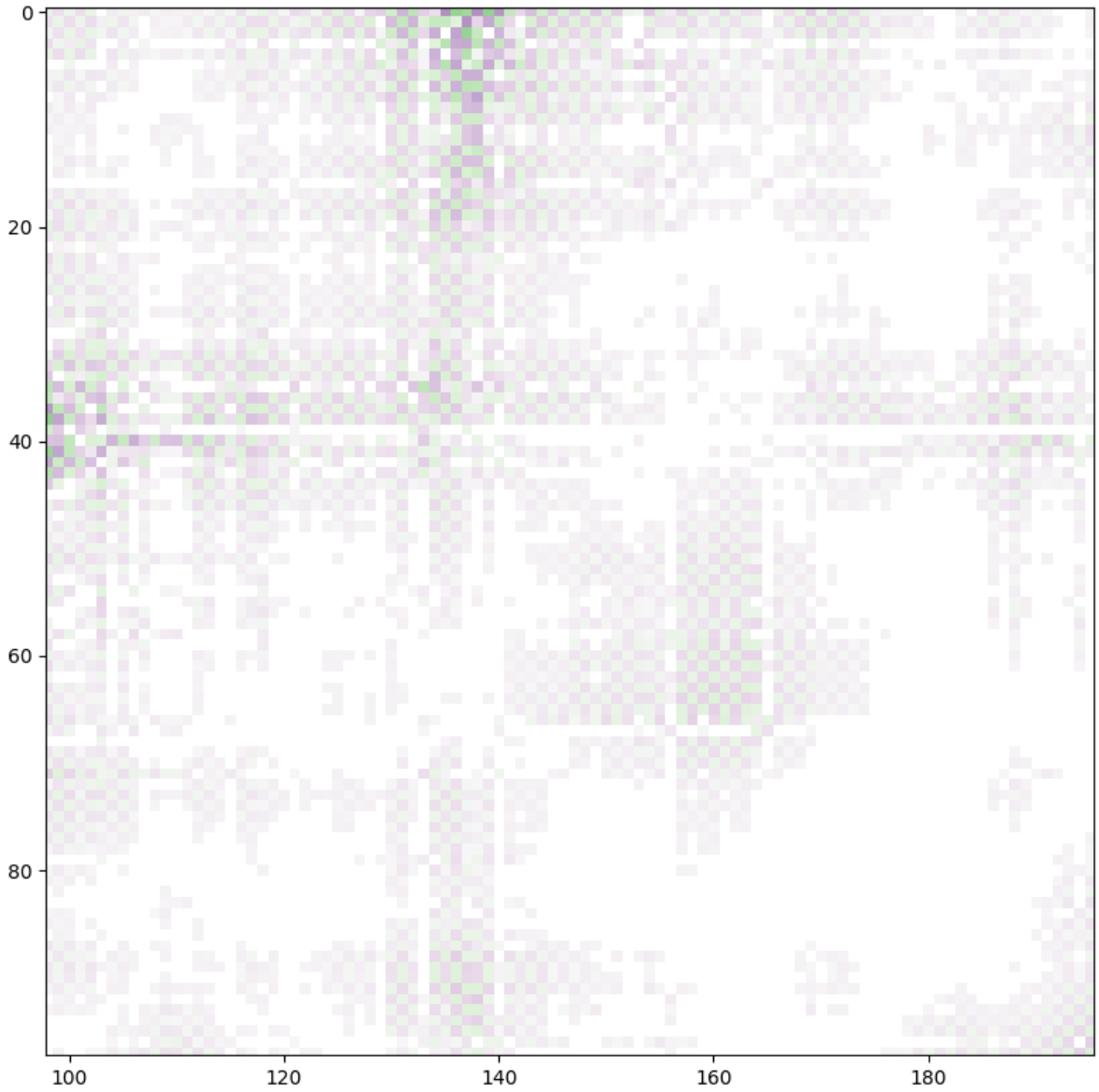


Figure 30: This figure is the same as figure 29 but zoomed in on C-F. Because this square is not on the diagonal the frequencies cannot be seen.

Figure 30 shows that the coupling across rows is very weak, even with the exact opposite chain of a particular chain. There are two regions of relatively strong coupling, namely around the coupling of 1-40 and 40-1. Amide groups 1 and 40 form the narrow connection between the two rows. To more precisely look at the couplings inside of this polymorph, those couplings are graphed with linearly scaling intensity below.

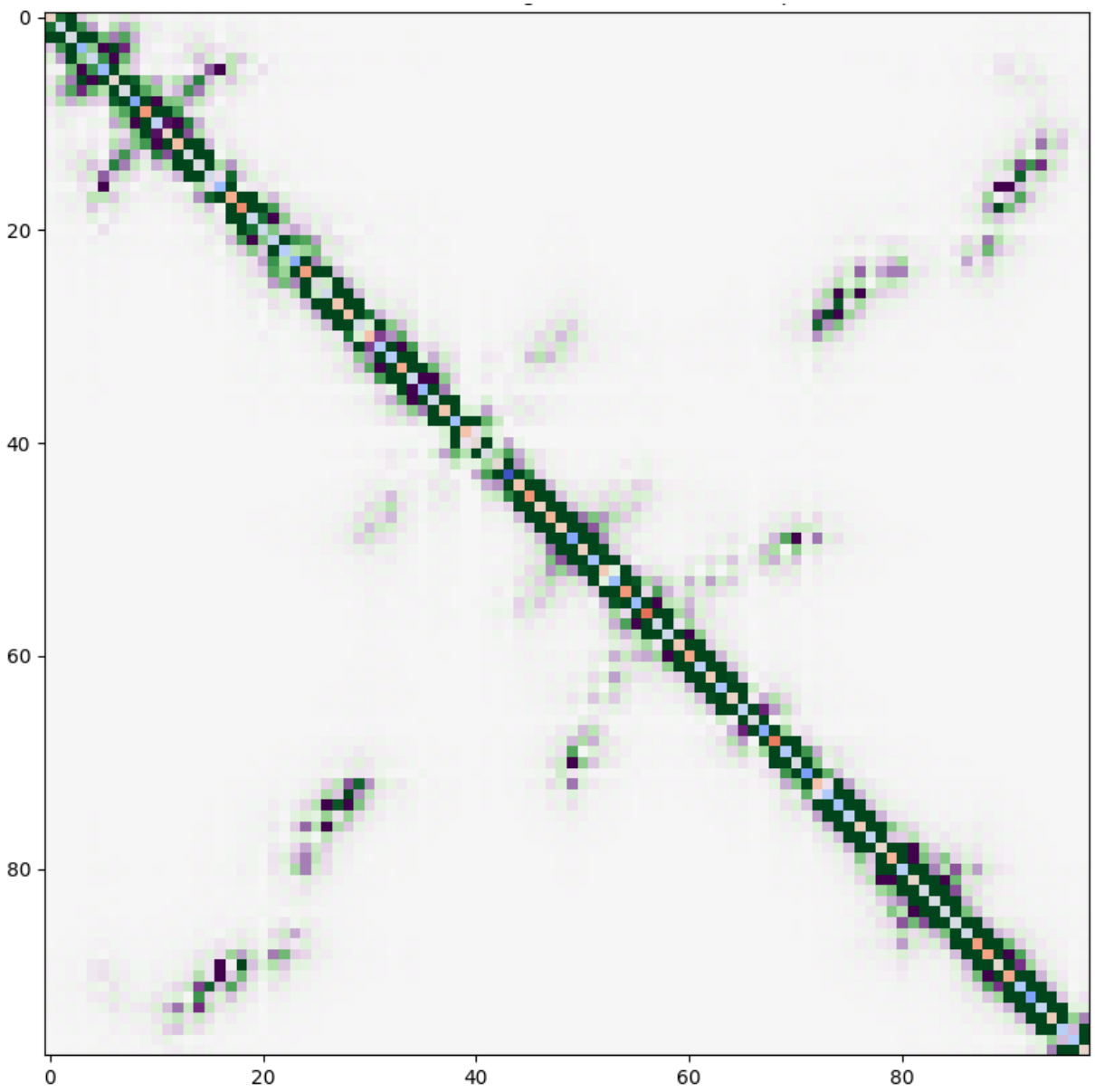


Figure 31: This figure is the same as figure 29, but the intensity of the coupling is graphed linearly and with a low range instead of logarithmically.

Figure 31 shows that the very strongest couplings between relatively far away amide groups are between groups 25-28 and 71-75 and groups 13-17 and 88-92.

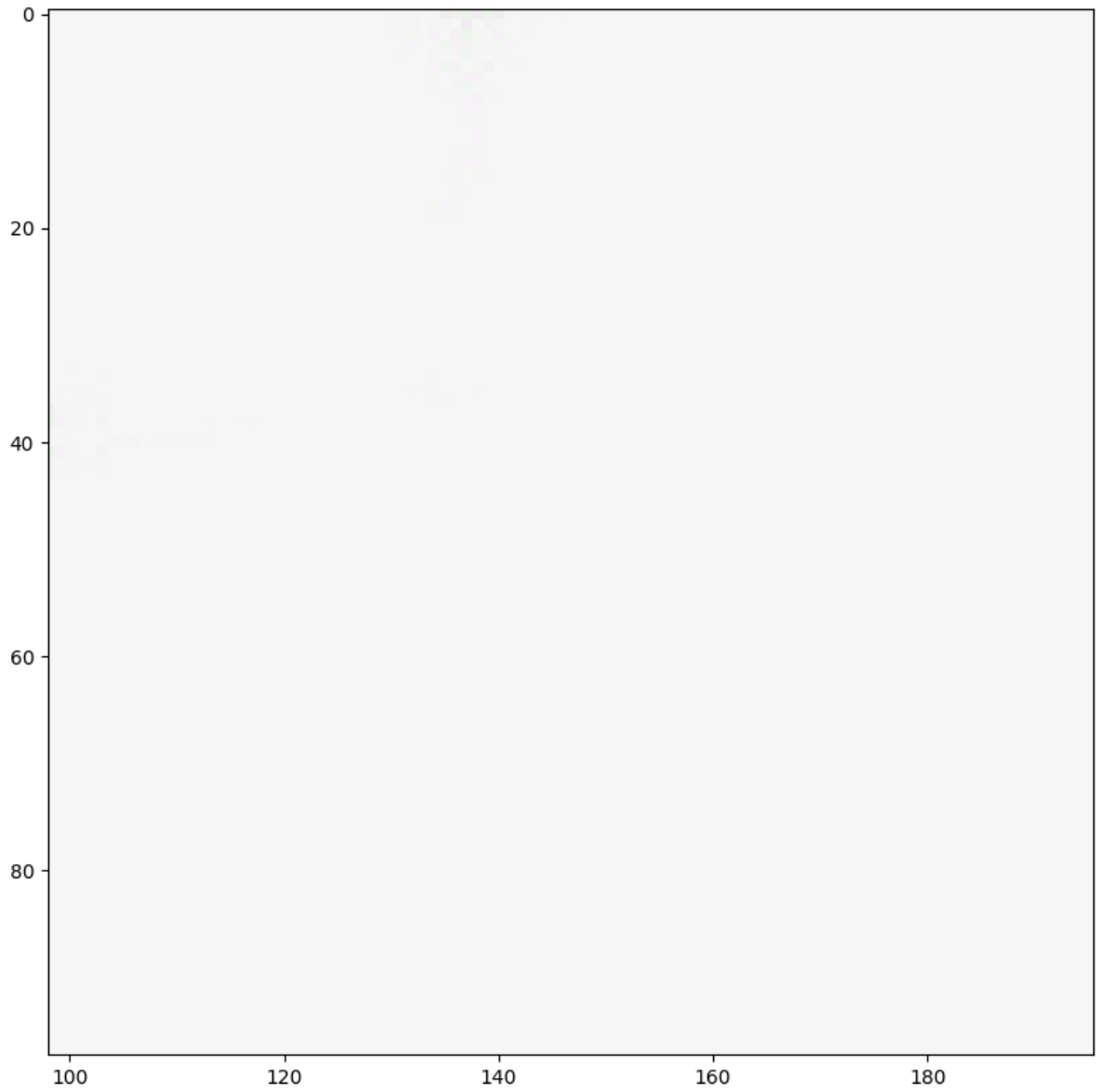


Figure 32: This figure is the same as figure 30, but the intensity of the coupling is graphed linearly and with a low range instead of logarithmically.

It is evident from figure 32 that the coupling between rows is minimal in this polymorph as no coupling is even visible.