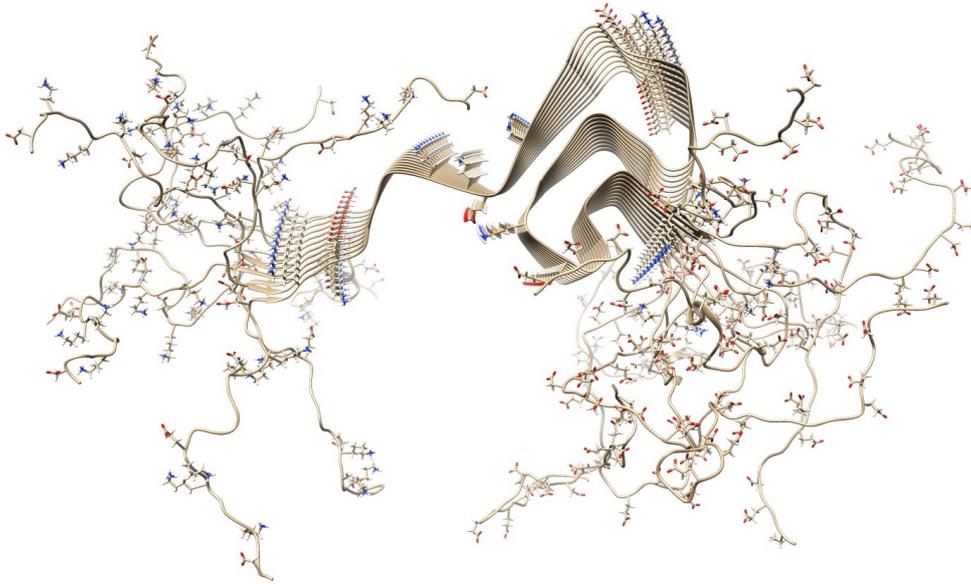




CHALMERS
UNIVERSITY OF TECHNOLOGY



Structure of amyloid fibrils

A computational analysis

Bachelor of Science Thesis

Josefin Erngren
Kerstin Karlsson
Jan Qvick
Anders Stigsson
Johan Tobin

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2016

The Authors grant to Chalmers University of Technology and University of Gothenburg the non-exclusive right to publish the Work electronically and in a non-commercial purpose make it accessible on the Internet. The Authors warrant that they are the authors to the Work, and warrant that the Work does not contain text, pictures or other material that violates copyright law.

The Authors shall, when transferring the rights of the Work to a third party (for example a publisher or a company), acknowledge the third party about this agreement. If the Authors have signed a copyright agreement with a third party regarding the Work, the Author warrant hereby that they have obtained any necessary permission from this third party to let Chalmers University of Technology and University of Gothenburg store the Work electronically and make it accessible on the Internet.

Structure of amyloid fibrils

A computational analysis

Josefin Erngren
Kerstin Karlsson
Jan Qvick
Anders Stigsson
Johan Tobin

- © Josefin Erngren, June 2016
- © Kerstin Karlsson, June 2016
- © Jan Qvick, June 2016
- © Anders Stigsson, June 2016
- © Johan Tobin, June 2016

Supervisor: Graham Kemp, Department of Computer Science and Engineering
Examiner: Niklas Broberg, Department of Computer Science and Engineering

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
SE-412 96 Göteborg
Telephone +46 31 772 1000

Cover:

An α -synuclein fibril, PDB-entry: 2N0A. The structure is generated in Chimera, a molecular graphics program.

Gothenburg, Sweden June 2016

Abstract

Amyloid fibrils are misfolded proteins related to several different diseases, such as Alzheimer's and Parkinson's disease. The aims of the report are to computationally analyze significant patterns and distributions of certain amino acids in the structure of amyloid fibrils. The direction relative to the fibril core is investigated for the charged amino acids and those present in bends. Amino acids in the amyloid fibrils tend to stack upon each other, these stacks are identified for the amino acids containing aromatic rings and the tilt of the aromatic rings are calculated relative to the axis of the fibril. The result from the analysis of the bends does not show significant patterns exclusive to amyloid fibrils, while the analysis of the charged amino acids indicates that they tend to face outwards relative to the fibril core. The tilt of the aromatic rings tends to be distributed around 55° . The results are consistent with previous findings and provide some information which could be of use when predicting the structure of amyloid fibrils.

Sammandrag

Amyloida fibriller är felveckade protein med koppling till flertalet sjukdomar så som Alzheimers och Parkinsons sjukdom. Målet med denna rapport är att utföra datoriserade analyser av betydande mönster samt fördelningar av specifika aminosyror i den amyloida fibrillens struktur. Riktning hos laddade aminosyror och de som finns närvarande i svängarna undersöks. Aminosyror närvarande i den amyloida fibrillens struktur tenderar att stapla sig på varandra, dessa staplar identifieras för aminosyror innehållande aromatiska ringarna och deras lutning relativt den amyloida fibrillens axel beräknas. Resultatet visar inte på några specifika mönster gällande svängarna i amyloida fibriller, medan analysen för de laddade aminosyrorerna visar att de föredrar att ordna sig utåt relativt fibrillens kärna. Lutningen hos de aromatiska ringarna tenderar att ha en fördelning runt 55° . Resultatet stämmer överens med tidigare iakttagelser och presenterar information som skulle kunna komma till användning vid modellering av strukturen hos amyloida fibriller.

Acknowledgements

We would like to thank our supervisor Graham Kemp for all the help with this project. Without your mini-lectures, ideas and encouragement this project would not have been the same. You have managed to make us appreciate the beauty of proteins as well as the field of bioinformatics. We have appreciated that you have cared about our personal development just as much as the development of the project.

Glossary

α-carbon	The central atom in the amino acid which carries the side chain.
Dry interface	Two surfaces facing each other, generating an environment which does not allow water to be located between them.
Expression system Extracellular	Organism used to produce proteins. Outside the cell.
FASTA format	The primary sequence of a protein represented, each amino acid is represented by their one letter abbreviation.
Heterocyclic	A ring structure formed by at least two different elements.
Homologous sequences	100% identical amino acid sequences.
Hydrophilic	Water attracting.
Hydrophobic	Water repelling.
In vitro	Studies performed outside the living cell, such as in a test tube.
In vivo	Studies performed in the living cell or tissue.
Intracellular	Within the cell.
Macromolecular	Large structures built of small subunits, such as nucleic acids and proteins.
Native form	The structure a protein is intended to form when produced.
Non-polar	A compound lacking the ability to form a dipole moment.

PDB-entry	Identification code for proteins available from the Protein Data Bank.
Polar	A compound having the ability to form partial charges which gives rise to a dipole moment.
Primary structure	The amino acid sequence of a protein.
Protein Data Bank	Database containing structural information about proteins.
Proteolytic cleavage	Degradation of proteins and polypeptide chains.
Quaternary structure	A protein structure including several subunits of proteins.
Secondary structure	The level of protein folding which includes α -helices and β -sheets.
Steric hindrance	Preventing conformations due to size and nearby structures.
Tertiary structure	The assembly of α -helices, β -strands and β -sheets.
Torsion angle	Angle between bonds to two specified groups within an amino acid.

Contents

1	Introduction	1
1.1	Importance of amyloid fibrils	1
1.2	Purpose and objectives	3
1.3	Delimitations	4
1.4	Overview of thesis	4
2	Background	5
2.1	Basics of protein structure	5
2.1.1	Amino acids and their chemical properties	5
2.1.2	Protein folding and levels of protein structure	7
2.2	Structure of amyloid fibrils	8
2.2.1	Steric zippers	12
2.3	β -solenoid proteins	13
3	Data sets	15
3.1	Data set of amyloid fibrils and steric zippers	15
3.2	Data set of β -solenoids	17
3.3	Evaluation of the data sets	18
4	Analysis methods	19
4.1	Identifying direction and distribution of amino acids	19
4.2	Identifying the axis of amyloid fibrils	20
4.2.1	Finding stacks in amyloid fibrils	21
4.2.2	Plane of best fit and line of best fit	24
4.2.3	Advantages and disadvantages	25
4.3	Tilt of aromatic rings	26
5	Results	29
5.1	Direction and distribution of amino acids	29
5.1.1	Bends of amyloid fibrils	29
5.1.2	Bends of β -solenoid proteins	31
5.1.3	Charged amino acids in amyloid fibrils	33
5.1.4	Charged amino acids in β -solenoids	38
5.2	Tilt and stacking of aromatic rings	39
6	Discussion	46
6.1	Direction and distribution of amino acids	46

6.1.1	Bends of amyloid fibrils and β -solenoids	46
6.1.2	Charged amino acids in amyloid fibrils and β -solenoids	48
6.2	Angles and tilt of aromatic rings	49
6.3	Reflections and future development	51
7	Conclusion	52
A	List of the 20 amino acids and their abbreviations	I
B	Distances between charged amino acids in amyloid fibrils	II
C	Angles of aromatic rings in amyloid fibrils and steric zippers	IV
D	Adjusted angles of aromatic rings in amyloid fibrils and steric zippers	VI
E	Data from the direction and distribution analysis	VIII
F	Number of stacks present in a subset of the amyloid fibrils	XI

1

Introduction

The introduction covers a short historical background of amyloid fibrils, the diseases they are related to and social issues that come with them. This is followed by the purpose and the more specific objectives of the report. Last, an overview of the thesis is presented.

1.1 Importance of amyloid fibrils

In 1854, the German physician and scientist Rudolph Virchow introduced the term ‘amyloid’. He was examining brain tissue with abnormal appearance and mistakenly concluded, after a positive iodine staining reaction, that the unusual macroscopic defects consisted of starch. He named the starch-like structures amyloid, originating from the Latin word *amylum*, meaning starch. In 1859, it was however determined that the abnormalities in the tissues did not consist of carbohydrates but rather contained aggregated proteins. Since then, amyloid fibrils have been recognized and studied as a group of proteins capable of forming fibrillar structures [1].

A major reason for the research interest in amyloid fibrils is their association with a large group of diseases referred to as amyloidosis¹. There are at least 27 diseases in this group, including Alzheimer’s disease, Parkinson’s disease and type II diabetes. Proteins related to amyloidosis have lost their native form due to misfolding [3]. The correct three-dimensional structure of proteins is essential in order for them to function properly. When misfolding, they have the ability to cause harm to their cellular surroundings [4].

The diseases of amyloidosis can differ from each other in several aspects depending on which protein and what organ is affected. In Alzheimer’s disease for example, the protein $A\beta$ aggregates in the brain and in type II diabetes the protein IAPP aggregates in the pancreas [3]. The feature common to the diseases of amyloidosis is the presence of misfolded fibrous proteins forming the characteristic structure of amyloid fibrils [5]. An example of an amyloid fibril is shown in Figure 1.1. The proteins involved in amyloidosis can, in their native form, differ vastly from each other regarding primary structure, function and size. When existing in their amyloid state however, their structures become very similar despite differences in amino acid sequence. This suggests that the characteristic structure of amyloid fibrils is accessible to many different proteins [6].

¹The word *-osis* means an increase in or abnormal amount of. The word amyloidosis meaning an abnormal amount of amyloid [2].

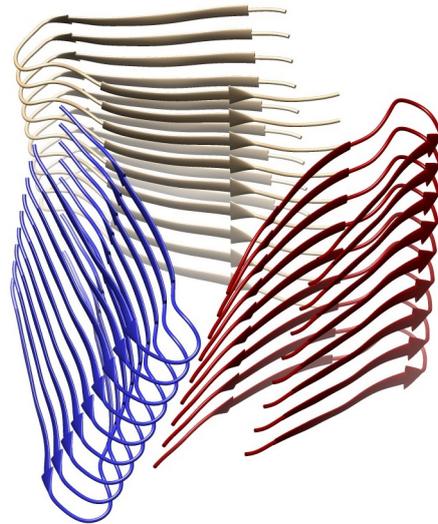


Figure 1.1: A three-dimensional model of a fibrillar structure, classified as an amyloid fibril. PDB-entry: 2MPZ.

Regardless of the number of proteins theoretically capable of folding into amyloid fibrils, the impact of the ones already identified and related to amyloidosis is indisputable, affecting a large number of people all over the world. The conditions of neurodegenerative diseases, such as previously mentioned Alzheimer's, does not only affect the afflicted but also family members or friends providing with every day care. Because of the behavioural and mental symptoms of the disease, the care can be both physically and psychologically intense [7]. Due to the great number of people affected, the social aspects are undeniable.

Alzheimer's disease, which is one of several diseases referred to as dementia, is more frequently occurring among elderly. With a rapid increase in both global population and life expectancy, the number of people diagnosed with Alzheimer's is expected to increase during the coming decades [7]. In 2010, the number of people in the world living with dementia reached 35.6 million which represented approximately 0.5% of the world's population [8]. Of those diagnosed with dementia, 50-75% is estimated to have Alzheimer's disease [7]. The same year, the costs of dementia were approximated to US\$604 billion. In 20 years, the number of people affected will double and the costs are likely to follow [8].

There is no doubt a treatment is of importance, for those affected directly as well as people in their close surroundings. To fully understand the impact of amyloid fibrils, researchers study the molecular structure of the proteins involved. Both experimental as well as computational methods have been used in order to calculate, model and predict how and why amyloid fibrils are formed in the human body. To build reliable models of protein structure further investigation is still required in this complex area of research.

1.2 Purpose and objectives

This report presents an investigation into the molecular structure of amyloid fibrils. A motivation for studying these structures is the future development of drugs or treatments for the conditions of amyloidosis. Knowledge about protein structure is important in order to understand protein property and function, and thereby determine the possible effects on human health. Despite the large amount of research being conducted on amyloid fibrils and their association to disease, more has to be done in order to determine this relation. This vast area of research is complex and therefore research objectives needs to be specified.

This report aims to investigate amyloid fibrils by computationally analyzing significant patterns and distribution of amino acids. More specific, there will be an investigation of the direction and distribution of certain amino acids. The amino acids chosen for this investigation are the ones present in the bends of the fibril as well as the charged amino acids throughout the fibril. There will also be an analysis of the arrangement of amino acids containing aromatic rings. There are several other patterns and amino acids that could be investigated apart from the ones just mentioned. Motivations for the chosen approaches follows below, as well as hypotheses regarding the outcome of the analyses.

The direction and distribution of the amino acids in the bends are analyzed because they are a recurring segment of the amyloid. Despite the amyloid fibril being a misfolded structure, it is still highly ordered and stable. This limits the structural alternatives for the amino acids within the bends, suggesting that amino acids with smaller side chains are more likely to face inwards towards the fibril core. Another characteristic of the amino acids pointing inward might be that they are hydrophobic since most proteins have a hydrophobic core. The hydrophobic core of proteins would also reduce the likelihood of finding charged residues facing inwards.

The aromatic rings have a relatively bulky and stiff structure compared to other amino acid side chains. This is an interesting feature when considering that amyloid fibrils consists of layers, causing identical amino acids to stack. Due to its shape, a single stack of aromatic rings is expected to be organized in a uniform manner. In order to analyze the aromatic rings, the fibril axis will be generated. This will provide a reference point to which angles of the aromatic rings can be decided. The angles generated will be used to compare the tilt of the aromatic rings.

The foundation of these analyses is a data set composed of proteins known to exist in a fibril state. In order to distinguish which structural attributes that are specific for the amyloid fibril, a data set of β -solenoids will be used for a corresponding, but less extensive analysis. The β -solenoid proteins are a relevant comparison since they have a similar structure to the amyloid fibril.

1.3 Delimitations

When collecting the data set, some information about the protein and its structure will not be taken into consideration. Firstly, there will not be any distinction regarding what experimental method is used when determining the structure. Secondly, the environmental conditions used during experimentation such as pH, pressure, temperature and solvent will be ignored. Lastly, no attention will be given to the type of organism the protein originates from or what expression system is being used to produce it.

1.4 Overview of thesis

In Chapter 2, protein structures and amyloid fibrils are presented. The method and results of gathering the data sets are provided in Chapter 3. The analysis methods are given in Chapter 4. Chapter 5 highlights the results of the report and the results are discussed in Chapter 6. Conclusions are drawn in Chapter 7. Finally, additional information and data from the analyses are available in the appendices.

2

Background

In the previous chapter, amyloid fibrils were introduced as misfolded proteins related to several diseases. In this chapter, their characteristic structure will be defined and further explained. The chapter begins with a theoretical background on protein structure, which can be disregarded by those familiar with the basic concepts of proteins. This is followed by a theoretical background of amyloid fibrils as well as steric zippers and β -solenoid proteins. Steric zippers being a model of amyloid fibril and β -solenoids shares some structural properties with the amyloid fibrils.

2.1 Basics of protein structure

Proteins perform a wide range of tasks in the human body, an example is as catalysts in chemical reactions. The function of a protein depends on its three-dimensional structure. In this section, the basics of protein structure are presented along with important chemical properties which stabilizes them.

2.1.1 Amino acids and their chemical properties

A protein molecule is a long chain formed by smaller building blocks called amino acids. There are 20 different amino acids that, in different combinations, make up all existing proteins. Each amino acid is linked to its neighbour through a peptide bond creating an array of amino acids called a polypeptide chain, more commonly referred to as a protein. Most proteins contain between 50-2000 amino acids, but larger proteins do exist. Any polypeptide containing fewer than 50 amino acids is simply called a peptide [9].

An example of a peptide chain is shown in Figure 2.1 where the location of the peptide bond is pointed out. An amino acid, also referred to as an amino acid residue, consists of two parts; a backbone and a side chain. The backbone is identical among all 20 amino acids and contain a central carbon atom (α -carbon), an amino group and a carboxylic acid group. The side chain is attached to the backbone and varies depending on the amino acid. Side chains contain different functional groups which give the amino acids different properties such as size, shape, bonding capacities and chemical reactivity. The arrangement of amino acids is therefore essential to the function and shape of the complete protein [9]. The 20 amino acids are presented in Appendix A, along with their one and three letter abbreviations.

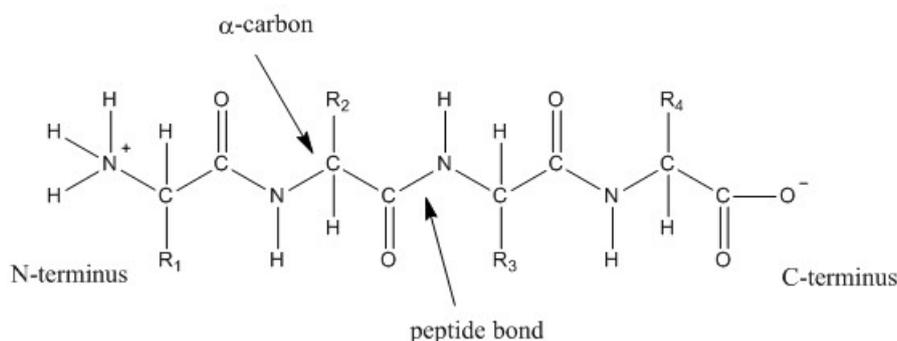


Figure 2.1: A peptide consisting of four amino acids, connected by peptide bonds. The R -groups on each amino acid represent the side chain. By convention, polypeptide chains are written from the N -terminus to the C -terminus. The N -terminus is an example of an amino group, including a nitrogen atom. The C -terminus is an example of a carboxylic acid group, including a carbon atom and two oxygen atoms.

Amino acids are classified into four groups depending on the chemical properties of the side chain; non-polar (hydrophobic), polar (hydrophilic), positively charged (basic) and negatively charged (acidic). Non-polar amino acids have hydrophobic side chains and come together to stabilize protein structures through hydrophobic effects, thus avoiding any contact with water or other charged segments. This phenomenon is involved in protein folding since it forces the non-polar amino acids to the inside of the protein. Polar amino acids have neutral side chains. They are hydrophilic and therefore more reactive than the previous group. The charged amino acids have an overall positive or negative charge which make them highly hydrophilic [9].

Aromatic rings are chemical structures present in some of the amino acid side chains. There are several ways of describing aromaticity and one approach is to use Hückel's rule. Hückel's rule defines an aromatic compound as a planar conjugated monocyclic structure with $4n + 2\pi$ -electrons, $n \in \mathbb{Z}_+$ [10]. As this is a complex definition, a visualization of the aromatic compounds present in the amino acids can be helpful. The amino acid residues containing aromatic rings are displayed in Figure 2.2. Phenylalanine and tyrosine contain the aromatic ring known as benzene. The benzene ring is a planar, symmetrical hexagon with three conjugated double bonds, thus satisfying the rule of Hückel. Many organic compounds does not contain a benzene ring, instead the aromatic systems are heterocyclic [10]. In these structures the main elements are not only carbon atoms. Such a system is present in the amino acid histidine, which contains the heterocycle imidazole. Tryptophan has a combination between benzene and a heterocyclic system, known as indole.

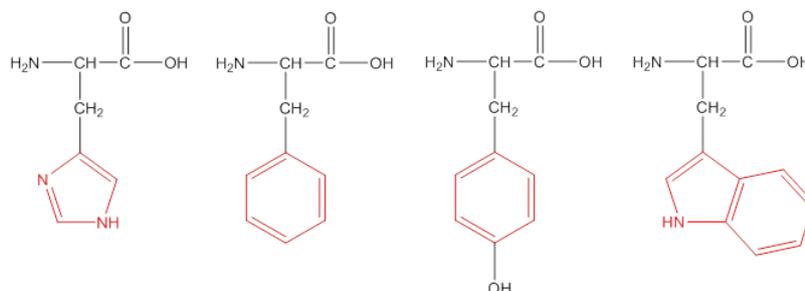


Figure 2.2: Amino acids containing aromatic rings, with the aromatic ring highlighted in red. To the left is histidine with the aromatic ring imidazole. Thereafter follows phenylalanine and tyrosine, both having a benzene ring. Lastly, tryptophan is shown. It also contains the aromatic ring benzene, but as a part of the larger aromatic complex called indole.

2.1.2 Protein folding and levels of protein structure

Proteins can be described at four structural levels; primary, secondary, tertiary and quaternary. The primary structure is the one-dimensional sequence of covalently linked amino acids, an example can be seen in Figure 2.1. The secondary structure describes the interaction of amino acids close to each other. They form hydrogen bonds and fold into three-dimensional structures known as α -helices and β -sheets. A β -sheet consists of at least two β -strands joined together by hydrogen bonds [9], [11]. The β -strand is usually visualized by an arrow pointing in the direction of the N- to the C-terminus. A β -sheet is visualized by a stack of arrows. Figure 2.3 displays two α -helices connected by a disordered segment and two β -strands assembled into a β -sheet. It is common that the hydrophobic and hydrophilic amino acids in a β -strand face the opposite side of the strand, causing the surfaces of the β -sheet to be uniformly hydrophobic and hydrophilic. This property allows a specific side of the β -sheet to associate with the environment in the most favourable way [4]. The important secondary structure in this report is the β -sheet, which are central to the structure of amyloid fibrils.

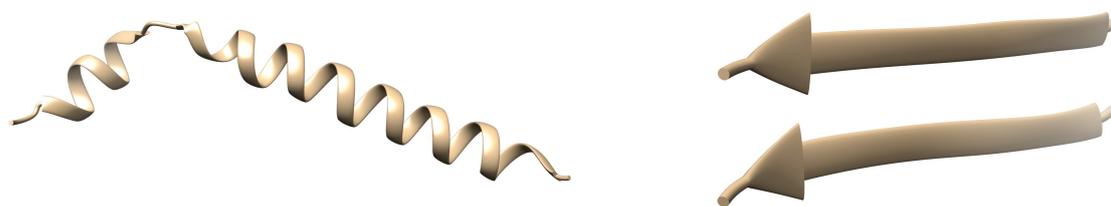


Figure 2.3: Visualization of the secondary structures. The structure to the left represents two α -helices connected by a disordered segment and the structure to the right displays two β -strands, forming a β -sheet. PDB-entry: 2LLM and 2Y3K.

The secondary structures can assemble into a higher order structure known as the tertiary structure [11]. Two common tertiary structures are the globular and the fibrous, see Figure 2.4.



Figure 2.4: Two different tertiary protein structures. The left is a fibrous protein, type I collagen, and the right is the enzyme catalase, which represents a globular protein. PDB-entries: 3HQV and 5BV2.

The tertiary structure is stabilized by a range of different electrostatic interactions. The strongest electrostatic interactions are the ionic. Oppositely charged amino acid residues can, in neutral pH, give rise to attractive forces forming an ionic bond, also known as a salt-bridge. If the pH increases or decreases beyond a certain point, the charged residues might lose their ionic character and the stabilizing property is lost. Disturbance in the protein folding is also possible if two uniformly charged residues interact, causing electrostatic repulsions [4]. Several residues are polar and have the ability to form hydrogen bonds. Despite hydrogen bonds being weaker than the ionic bonds, they are large in numbers and will therefore contribute to the stability of the tertiary structure. The last important electrostatic interaction is the van der Waals force. This type of bonding occurs between the hydrophobic residues and it is essential for the occurrence of the hydrophobic effect. The hydrophobic residues have a limited solubility in water and can therefore aggregate in the center of the protein. The hydrophobic residues arrange tightly and form a dry core. The tight packing releases water molecules, which stabilizes the protein structure significantly [4], [11].

The last level of protein folding is the quaternary structure. It is formed when several polypeptide chains assemble into larger protein complexes [11], making the polypeptide a subunit. The protein complex can either consist of several identical polypeptide chains or different ones. The quaternary structure is, like the tertiary, stabilized by electrostatic interactions [4].

2.2 Structure of amyloid fibrils

Amyloid fibrils have previously been introduced as misfolded proteins linked to several diseases known as amyloidosis. As for now, presence of amyloid fibrils can only be viewed as a pathological hallmark in the progression of disease. What role the aggregation of amyloid fibrils has in the formation of disease is yet to be discovered. Current

research investigates the process of aggregation. The process is complex, forming several intermediates, some which are believed to be responsible for the toxic effects causing disease [12], [13]. Since the main focus of this report is the structure of amyloid fibrils, rather than their correlation to disease, this section aims to present the structural attributes common to amyloid fibrils.

Amyloid fibrils refers to aggregated unbranched filamentous proteins. They originate from proteins which have lost their native structure and undergone spontaneous self-assembly. The aggregated protein contains a high extent of β -sheets, organized in a cross- β structure [14]. The cross- β structure consists of β -sheets arranged along the axis of the fibril and the assembled β -sheets contains stacked β -strands running perpendicular to the fibril axis [14], [15], [16], [17]. An example of a cross- β structure is seen in Figure 2.5.

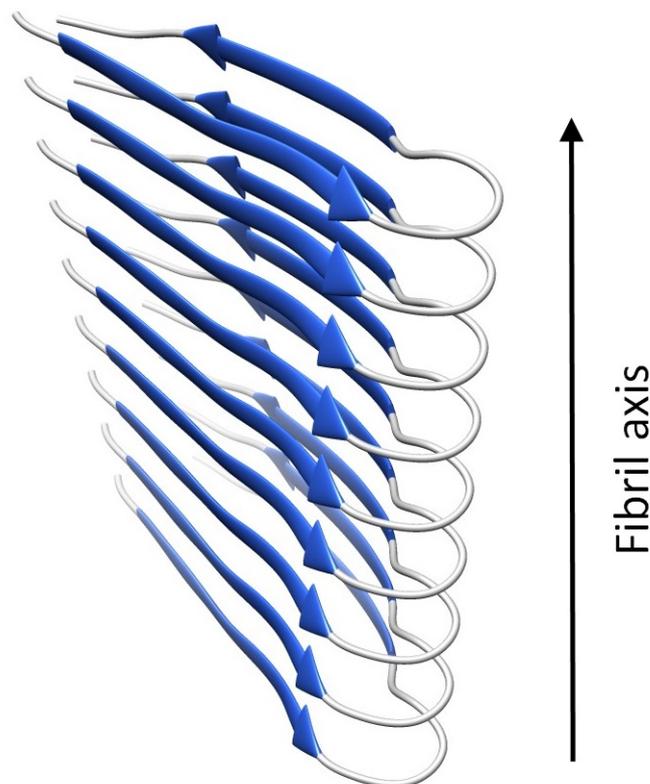


Figure 2.5: A cross- β structure. The two stacks of blue arrows represents two β -sheets which are arranged along the axis of the fibril. The β -strands in the β -sheets runs perpendicular to the fibril axis. The black arrow shows an approximation of the fibril axis. PDB-entry: 2MPZ.

Figure 2.6 shows an example of an amyloid fibril and how it stacks. Identical peptide sequences assemble into the larger cross- β structure by stacking upon each other, eventually twisting around its own axis. Figure 2.6 does contain β -strands and sheets, but these are not represented by arrows in this structure.

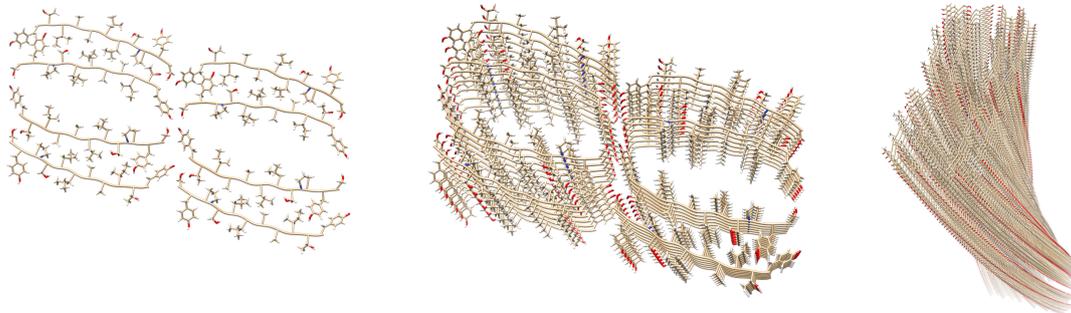


Figure 2.6: Three levels of a protein structure representing an amyloid fibril. The structure to the left is the first layer in the fibril, which is formed by an assembly of short peptide sequences. In the center, the same sequence has stacked upon each other in five layers. The structure to the right shows a fibril with 114 layers of the same peptide sequence. All three assemblies are viewed from the same angle. PDB-entry: 2M5K.

The β -sheets in the cross- β structure can vary in number and be either parallel or anti-parallel. The difference between parallel and anti-parallel sheets is displayed in Figure 2.7. It has been observed that amyloid fibrils formed by shorter peptide segments tend to assemble into anti-parallel β -sheets [14], compared to the amyloid fibrils with longer peptides, more than 20 amino acids, which tend to form parallel β -sheets [18].

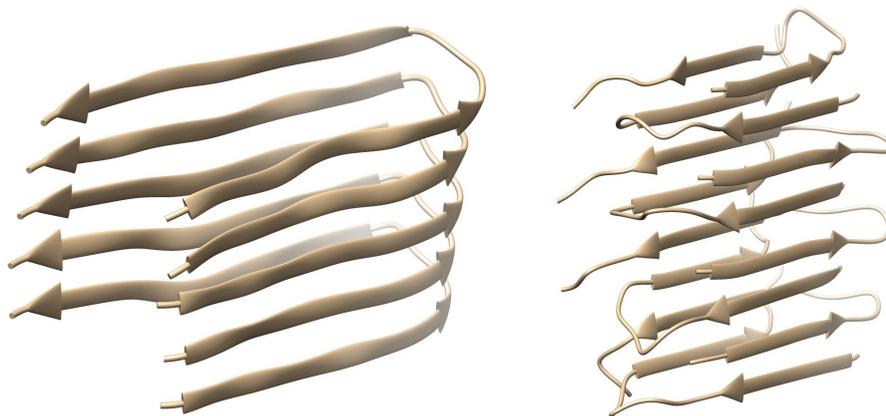


Figure 2.7: To the left is an amyloid fibril with parallel β -sheets. The structure to the right represents an amyloid fibril with anti-parallel β -sheets. The differences between the two structures are seen in the directions of the arrows. PDB-entries: 2BEG and 2LNQ.

Other variations seen in the structure of the amyloid fibrils are if the amino acids in the β -sheets tend to be in- or out-of-register. An in-register arrangement is achieved when the same kind of amino acid residues are aligned on top of each other in each level of the β -strand, throughout the β -sheet. When these alignments are missing, the amino acid residues are out-of-register. An example of in- and out-of-register formation is shown in Figure 2.8. The parallel and in-register formation is the most common arrangement seen in the amyloid fibril. This combination might be due to an increased possibility of hydrophobic interactions between the β -sheets, which in turn increases the stability of the structure [18].

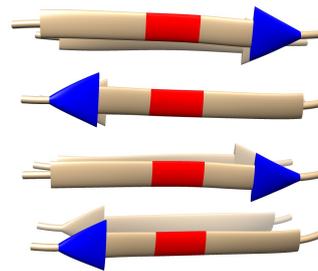


Figure 2.8: The figure shows two β -sheets, with identical amino acid sequences in all of the β -strands. The opposite direction of the arrows indicates that the sheets are anti-parallel. Highlighted in red are alanines, together they form an in-register arrangement. The amino acids coloured in blue are leucines, they display an out-of-register arrangement. PDB-entry: 2KIB.

The presence of a cross- β structure is one of the requirements for an aggregated protein to be defined as an amyloid fibril. To assign a cross- β and an amyloid fibril structure, certain distances in the assembly needs to fall within set limits [3], [14], [19]. Another characteristic property worth mentioning, because of its historical importance in assigning the structure, is the amyloid fibrils' ability to bind certain dyes, such as Congo Red [18].

The structure of amyloid fibrils is highly stable [14], it is insoluble and resistant to proteolytic cleavage. Resistance to degradation can lead to accumulation within tissues, a phenomena seen in amyloidosis [3], [18], as mentioned in Section 1.1. The insolubility of amyloid fibrils, as well as their inability to crystallize, are important aspects when trying to experimentally retrieve its structural information. These properties intervene with the methodologies used for deciding protein structures. It was not until the late 1990s that detailed information about the amyloid fibrils three-dimensional structure was obtained [14]. In 2008, there were still no complete molecular structure of amyloid fibrils available [18]. Even though there are difficulties retrieving structural information about amyloid fibrils, more can be said about their amino acid sequences. There are no common features in the sequences of the native proteins that indicates the possibility of amyloid fibril formation [14], [18]. This might be related to the high amount of β -sheets in the amyloid fibril, since these are mainly stabilized by hydrogen bonds between the backbone of the peptide rather than the side chains [20].

The structure of amyloid fibrils is stabilized by hydrophobic interfaces. Hydrophobic residues generates a low dielectric environment within the fibril. A combination of a low dielectric environment and electrostatic repulsions between charged amino acids would possibly disrupt the structure of the fibril [14]. However, electrostatic attraction might instead help to further stabilize the structure of amyloid fibrils. To approximate the distance between charged amino acids and thereby asses influences on stability, Coloumb's law

can be altered into an equation describing the energy of interactions (E).

$$E = \frac{q_1 q_2}{4\pi\epsilon_r\epsilon_0 d} \quad (2.1)$$

In Equation 2.1 q represents the charges of the amino acid residues, d is the distance between them, ϵ_0 is the permittivity in vacuum and ϵ_r is the dielectric constant of the medium. When performing calculations on biomolecules, an aquatic environment ($\epsilon_r=80$) is usually favoured. The energy then range from -13 to -17 kJ/mol [4], which corresponds to a distance of 1.02-1.34 Å. In amyloid fibrils, charged amino acids can be located in the hydrophobic interfaces where there is a low dielectric environment. The electrostatic interaction will then be stronger compared to an aqueous environment. It is therefore reasonable that a distance greater than 1.34 Å between charged residues can still form an ionic bond. This is consistent with the observations of Barlow and Thornton in 1983, which determined the most common distance between charged residues in proteins to be ≤ 4 Å [21], as well as with Tycko, who accepts the distance 3.7 Å between the carboxylate carbon and amino nitrogen atoms, as an ionic bond [14].

2.2.1 Steric zippers

In the previous section, the difficulties of retrieving information about the structure of the amyloid fibrils were introduced. One approach to manage this issue has been to create models of amyloid fibrils by using short peptide sequences, originating from the fibril sequences. Structural information is possible to generate for the short peptide sequences, compared to amyloid fibrils [22]. If the sequence aggregates into a cross- β structure, interacts in a self-complementary manner and forms a completely dry interface, it is classified as a steric zipper [23], [24]. An example of a steric zipper is seen in Figure 2.9. A steric zipper is also seen in Figure 2.6, which represents the structure of an amyloid fibril. Steric zippers are regarded as basic units of amyloid fibrils, due to its limited appearance in other protein structures [23], [25].

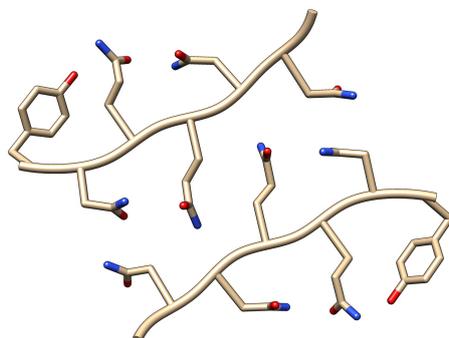


Figure 2.9: A short peptide sequence originating from the yeast prion Sup35. The segment is classified as a steric zipper, it aggregates into a cross- β structure, forms a dry interface and interacts in a self-complementary manner. PDB-entry: 1YJP.

When testing as for whether or not a sequence can be classified as a steric zipper, in vitro experiments are being used. Applying different environmental conditions to a sequence, during experimentation, does lead to variations in the arrangements of the β -sheets [16]. The variations are often referred to as polymorphism and include the formation of either parallel or anti-parallel β -sheets, as well as differences in how the assembly of β -sheets are facing each other. These variations of arrangement gives eight classes of steric zippers, all with different combinations and all being stable structures. The structural diversity occurring during in vitro experiments has also raised questions regarding the existence and distribution of polymorphism in vivo [23].

2.3 β -solenoid proteins

β -solenoid proteins, also referred to as β -solenoids, is another class of proteins relevant to include parts of, for the structural analysis due to their structural similarities to amyloid fibrils. This large group consists of proteins exhibiting a solenoid appearance, with secondary structures consisting of β -strands, hence the name β -solenoids. These include β -helices and β -rolls, but the distinction between them will not be evaluated in this report [26].

The prominent similarity between amyloid fibrils and β -solenoids, is the repetitive appearance of the structural units where amino acid side chains stack around a common axis. A difference is however that the β -strands stacked in β -solenoids are not necessarily identical, as they tend to be in amyloid fibrils [26]. A β -solenoid consist of one single polypeptide chain. The β -strands coil around the solenoid axis via turns and bends forming a helical structure similar to the one shown in Figure 2.10. Each coil consists of about 12-30 amino acids. The polypeptide chain has an axial rise of 4.8 \AA ($\pm 0.2 \text{ \AA}$) and the distance is the same on all sides of the helix. The interior of the solenoid consists of a compact and hydrophobic core, while charged and polar amino acids tend to face the solvent [27].



Figure 2.10: PDB-entry: 3P4G. An example of a β -solenoid. The arrows, representing β -strands, are stacked in a parallel manner. The bends and turns are the segments connecting the β -strands.

It has been observed that the bends of β -solenoids differ from those in globular proteins. The bends in the β -solenoids are arranged in such a manner where adjacent β -strands interact via their side chains instead of their polypeptide backbone. Figure 2.11 (a) shows a representation of a bend more common in globular proteins (left) and a bend more common among β -solenoids (right). Even though the bends are irregular segments, they have been shown to display recurring patterns [27].

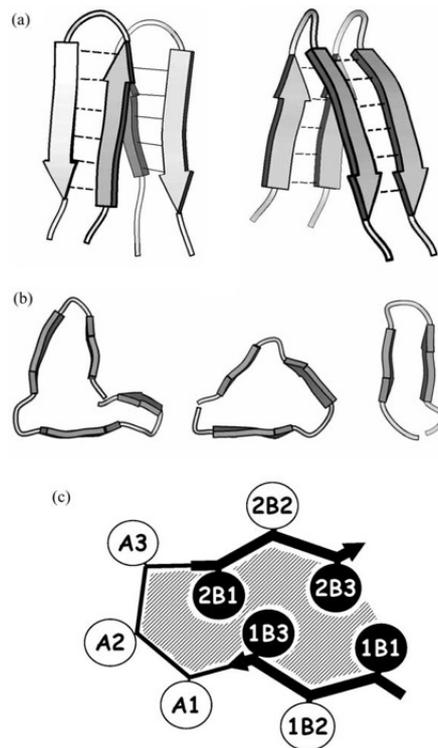


Figure 2.11: (a) To the left, a representation of a bend more common in globular proteins and to the right an example of a bend common in β -solenoids. The broken lines showing hydrogen-bonds. (b) Cross-sectional shapes occurring in β -solenoids. The more common L- and triangular shapes shown to the left and middle, respectively, and the oval shape of β -rolls to the right. (c) A representation of a bend which is the segment connecting the two β -strands (shown as arrows). The white and black circles representing positions of amino acid side chains. Reprinted from [27], with permission from Elsevier.

Cross-sectional shape is one feature considered when adding β -solenoids to the data set presented in Section 3.2. Figure 2.11 (b) shows three different cross-sectional shapes; L-shape (left), triangular (middle) and β -roll (right). β -helices show a predominantly triangular and L-formed shape, while β -rolls tend to be more oval [27].

As mentioned, β -solenoids is a large class of proteins with further classifications based on coil shape, number of strands coiled around the axis, right or left handedness etc. [28]. Investigating and discussing the different classifications is beyond the scope of this report and the term β -solenoids is used in a broad sense.

3

Data sets

This chapter presents the resulting data sets required for the upcoming analyses. The data sets consists of protein structures submitted to the Protein Data Bank (PDB) [29]. PDB contains information about more than 100 000 macromolecular structures. Each structure is assigned a unique PDB-code, also referred to as PDB-entry, and the structural information is stored in a corresponding PDB-file. The PDB-file contains information such as experimental details and coordinates describing the position of the atoms within the structure. The data sets used in the analyses are the ones containing amyloid fibrils and steric zippers, as well as a complementary data set with β -solenoids.

3.1 Data set of amyloid fibrils and steric zippers

Structures of amyloid fibrils were found in PDB using the classification of protein fibrils. The search term generated 121 entries. Five more structures were collected with other approaches. The PDB-entries were visually inspected, 17 showed the characteristic structure of an amyloid fibril and were included in the data set. A sequence alignment of the data set was performed using the FASTA format of the PDB-entries in the Clustal Omega program[30]. Any homologous sequences could then be identified.

Of the 121 entries included in the classification of protein fibrils, more than 60 were short segments of amyloid fibrils or considered to be steric zippers. Steric zippers are, as mentioned in Section 2.2, short peptides of approximately six amino acids and can be seen as the basic unit of the amyloid fibrils. These PDB-entries were excluded at first, since they only contain a limited amount of coordinates in their PDB-files. However, larger structures are generated by applying transformation matrices and are available in the file Biological Assembly, in PDB. The larger structures were collected in a separate data set, referred to as the data set of steric zippers. The structures generated in the Biological Assembly lacks bends, and has an artificial arrangement of the residues within the structure, making them less interesting in the analysis of the charges and not at all for the bends. The arrangement of aromatic rings are on the other hand more relevant and the data set was therefore narrowed down to those containing one or several residues with aromatic rings. This resulted in a data set containing 17 steric zippers.

To make a distinction between structures generated by transformation matrices and those collected as PDB-files, two separate data sets were formed. Table 3.1 displays the PDB-entries having the characteristics of an amyloid fibril and having coordinates for the full structure in a PDB-file, this data set is referred to as amyloid fibrils. In Table 3.1, some properties of the structures are also present. These include the number of β -strands in the β -sheets and the direction of the β -strands, also known as parallel or anti-parallel β -sheets, as well as homologous PDB-entries, those entries have identical numbering in the last column.

Table 3.1: Data set of amyloid fibrils. All the entries in the table have their full structure available in a PDB-file and they are all determined by NMR. The table shows if the β -sheets in the structure are arranged in a parallel or anti-parallel manner and the last column indicates homologous PDB-entries.

PDB-code	Name	Parallel β -sheets	Anti-parallel β -sheets	Number of β -strands	Homology
2RNM	HET-s(218-289) prion	x	-	10	1
2KJ3	HET-s(218-289) prion	x	-	6	1
2LMQ	A β (1-40)	x	-	6	2
2LMP	A β (1-40)	x	-	6	2
2LMN	A β (1-40)	x	-	6	2
2M4J	A β (1-40)	x	-	3	2
2LNQ	A β (1-40) Iowa mutant	-	x	8	3
2BEG	A β (1-42)	x	-	5	4
2MXU	A β (1-42)	x	-	11	4
2MPZ	A β (1-42) Iowa mutant	x	-	8	5
2MVX	A β (1-42) Osaka mutant	x	-	5	6
2M5N	Transthyretin(105-115)	x	-	8	7
2NNT	Human CA150 WW2	x	-	4	8
2N1E	MAX1 peptide fibril	-	x	8	9
2E8D	β 2-microglobulin	x	-	4	10
2KIB	hIAPP	-	x	5	11
2N0A	α -synuclein	x	-	10	12

In the data set of amyloid fibrils, Table 3.1, 12 out of 17 structures has a unique sequence. It is important to notice that five of the unique sequences originate from the same protein, the A β , not differing by more than three amino acids. On the other hand, the A β is a well established amyloid fibril and the structures do display a variation regarding the β -sheets, both in number and directions. There is also a difference in polymorphism, some displays three- or two-fold symmetry while others are a single filament. The two first entries in Table 3.1 are both the HET-s(218-289) prion, with an almost identical three-dimensional structure. It might have been of interest to exclude one, but since lacking any validation parameters for structures determined by NMR, there is no obvious way to rank the two. The same problem arises when attempting to restrict the number of A β -proteins.

The PDB-entries requiring the file Biological Assembly are collected in Table 3.2, referred to as the data set of steric zippers. The table displays, as the previous, the number and direction of the β -strands, with the addition of peptide sequence. No homology has been determined, due to their limited number of amino acids it is easy to identify any similarities. As an example, the two last entries in Table 3.2 only differ in the last amino acid. As mentioned earlier in this section, the commonality between the sequences in the data set of steric zippers, are the presence of at least one residue containing an aromatic ring. It was also mentioned that the selection was carried out because the steric zippers lack bends, and has an artificial arrangement amongst the residues, resulting in them only being used in the analysis of the aromatic rings.

Table 3.2: Data set of steric zippers containing aromatic rings. The structures are available on PDB, as the file Biological Assembly. The table includes the peptide sequence of each PDB-entry, the number and arrangement of the β -strands in the β -sheets.

PDB-code	Name	Sequence	Parallel β -sheets	Anti-parallel β -sheets	Number of β -strands
4ZNN	α -synuclein (47-56)	GVVHGVTTVA	x	-	5
3NHC	Human prion protein (127-132)	GYMLGS	-	x	6
3NHD	Human prion protein (127-132)	GYVLGS	-	x	6
4R0P	Human lysozyme C (46-61)	IFQINS	x	-	5
3NVF	Human prion (138-143)	IIHFGS	x	-	5
3OW9	A β (16-21)	KLVFFA	-	x	6
3LOZ	β 2-microglobulin	LSFSKD	-	x	4
2OMP	Human insulin chain A (13-18)	LYQLEN	-	x	4
3NVG	Mouse prion (137-142)	MIHFGN	x	-	5
3NVH	Mouse prion (137-143)	MIHFGND	x	-	5
3NVE	Syrian hamster prion (138-143)	MMHFGN	-	x	6
3FVA	Elk prion	NNQNTF	x	-	3
3FTL	Islet amyloid polypeptide	NVGSNTY	x	-	-
2OMQ	Human insulin chain B (12-17)	VEALYL	-	x	4
4NP8	Tau	VQIVYK	x	-	4
4ONK	[Leu-5]-Enkephalin mutant	YVVFLL	-	x	6
4OLR	[Leu-5]-Enkephalin mutant	YVVFV	-	x	6

Compared to the data set of amyloid fibrils, the structures in Table 3.2 has a more even distribution between parallel and anti-parallel β -sheets, but a slightly more narrow in number of β -strands. Another difference between the data sets is the methodology used for determining the structure. The amyloid fibrils are only determined by NMR, whilst the steric zippers are mainly determined by X-ray crystallography.

3.2 Data set of β -solenoids

The data set consisting of five β -solenoid proteins was, similarly to the other data sets, collected from the PDB. The data set is to be viewed as a complementary comparison to the set of amyloid fibrils, which continues to be the main focus. As a consequence,

the number of β -solenoids was intentionally kept small. Since the class of β -solenoids is large with a number of subclasses, this less extensive data set does not intend to represent the entire group, but are included in order to investigate possible similarities with the amyloid fibrils.

The data set of β -solenoids contains proteins from different origins to give the set variability. Three of the proteins were originally found in a study by Kajava [28], where a set of β -solenoids was categorized based on their cross-sectional shape. The ones chosen from the study were selected because of their common L-shaped cross section, even though they originate from different host organisms. 1Z2F, an anti-freeze protein, and 1LXA, a type of protein called transferase, were found using the search terms antifreeze protein and β -helical protein. The data set can be seen in Table 3.3.

Table 3.3: Data set of β -solenoids. Table showing name of protein, that all proteins have parallel β -sheets and the number of coils present in the helical structure.

PDB-code	Name	Parallel β -sheets	Anti-parallel β -sheets	Number of coils in helix
1Z2F	CfAFP-501	x	-	8
1LXA	UDP N-Acetylglucosamine Acetyltransferase	x	-	9
1RU4	Pectat lyase Pe19A	x	-	10
1EE6	Pel-15 pectate lyase	x	-	8
1RMG	Rahmno galacturonase A	x	-	11

3.3 Evaluation of the data sets

The β -solenoids are, as previously stated, a relevant group of proteins to use for comparison due to similar structure. The similarities overlap to such a degree that some proteins have been assigned, both an amyloid structure, as well as a β -solenoid structure. The HET-s prion protein, seen in Table 3.1, is an example of a such a case.

Since researchers believe that many different proteins can form the characteristic structure of amyloid fibrils, the analyses would benefit from a more versatile data set. Despite this, the data set of amyloid fibrils only contain 17 different structures, where 9 of them originates from the same protein. This does not represent the entire group of proteins capable of forming fibrils. Similar problems occur with the data set of β -solenoids. To improve the comparison between the two classes, the data sets would preferably be of equal size and display the same diversity. It would be possible to extend the data set of β -solenoids, but the different subclasses would have to be considered. This is harder to do with the amyloid fibrils, since there is a limited amount of experimentally determined structures available.

4

Analysis methods

In this chapter, the methods for analyzing the data sets described in the previous chapter are presented. First an analysis of the direction and distribution of charged amino acids and amino acids present in bends are described. Secondly, the methods for identifying the axis in amyloid fibrils are shown. Lastly, the method used to calculate the angle of aromatic rings is presented.

4.1 Identifying direction and distribution of amino acids

The analysis of direction and distribution was conducted on two groups of amino acids; those present in bends, and those categorized as charged. The bend is defined as the segment connecting two β -strands and the amino acids belonging to a bend are those present in that specific segment. An example of amino acids present in a bend is shown in Figure 4.1.

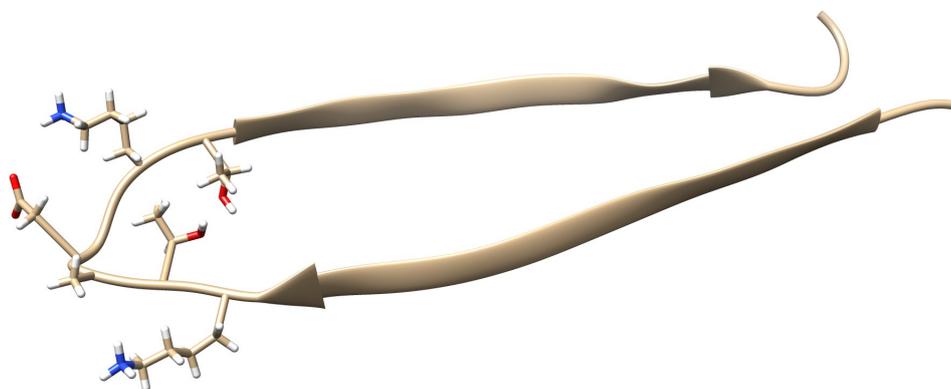


Figure 4.1: The figure displays those amino acids present in the bend, the segment connecting two β -strands. PDB-entry: 2NNT.

To perform the analysis, the molecular graphics program Chimera was used [31]. Chimera visualizes the information about the protein structure in the PDB-file as a three-dimensional model. Information about secondary structures for the three-dimensional model is available in the PDB-file. If the information is not included, Chimera uses the program DSSP (Define Secondary Structure of Proteins) [32] to assign the secondary structures. The PDB-files often contain several different models of the proteins, therefore only the first

model was extracted and saved as a new PDB-file, which was later used in Chimera. One by one, the proteins in the data sets were manually examined. The amino acids of interest were counted and the direction in which they were facing was noted.

The direction of an amino acid side chain is related to the core of the fibril and is categorized as inwards or outwards relative the core. The meaning of the term core may differ depending on the structure and it is not general for the amyloid fibril. Here, the term refers to the interface of two β -sheets, connected by at least one bend. To clarify the meaning of inwards or outwards relative the core, an example of side chain direction is presented in Figure 4.2. There are three cases where the definition of the core is insufficient but the core of the fibril still present. Firstly, when the β -sheets are not connected by a bend. Secondly, when the existing interfaces does not contain β -sheets. Lastly, when a single β -sheet can give rise to two separate interfaces, both regarded as the core.



Figure 4.2: To the left, amino acid side chains regarded having an inward direction relative to the core of the fibril. To the right, side chains facing outwards. PDB-entry: 2BEG.

No specific segment of the amyloid fibril were selected for the analysis of the charged amino acids. Instead, all of the charged amino acids in the structure were evaluated. The charged amino acids are the positively charged lysine, histidine and arginine, classified as basic, and the negatively charged aspartic acid and glutamic acid, classified as acidic. Distances between charged amino acids in close proximity were also measured in order to determine possible electrostatic interactions. The distance was measured starting from any of the oxygen atoms on the carboxylate carbon, in the residues of aspartic and glutamic acid, to one of the nitrogen atoms in lysine, asparagine or histidine. The distances were measured with a built-in tool in Chimera and saved if less than, or close to, 4 Å. The threshold of 4 Å was chosen due to it being the upper limit of ionic bond formation in proteins, see Section 2.2.

4.2 Identifying the axis of amyloid fibrils

As stated, one of the objectives is to calculate the angle between the normal of the planar aromatic rings in respect to the axis of the fibril. Multiple methods for determining the axis of amyloid fibrils were developed. All methods were developed in Matlab, with some visual identification aid from Chimera. The following subsections will present the mathematical background to these methods and will end with a comparison regarding strengths and weaknesses of each method. A flow chart showing a summary of how the methods

are used is shown in Figure 4.3.

As previously mentioned in section 3, all relevant information regarding the protein can be found in the PDB-file. This includes coordinates of all atoms in the structure which allows the use of linear algebraic operations as long as each atom is treated like a point in \mathbb{R}^3 .

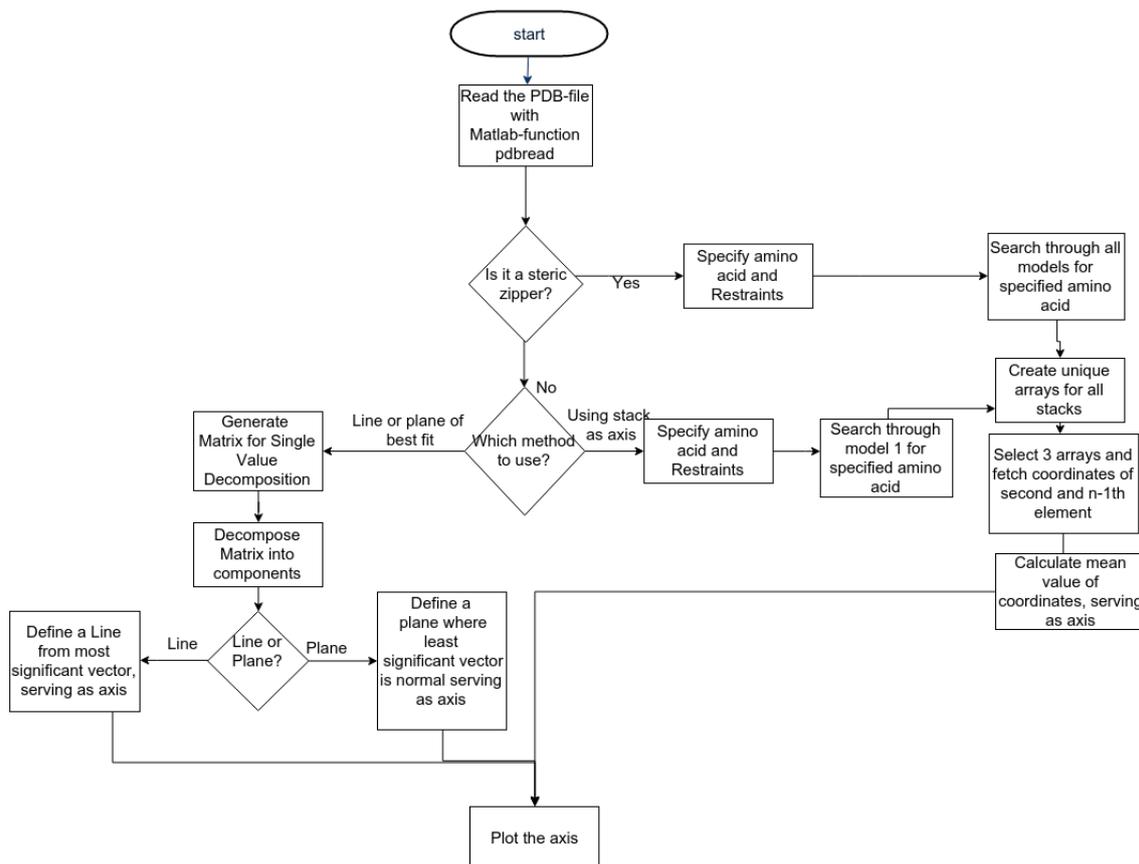


Figure 4.3: A flowchart showing how the Matlab-code is used to calculate and plot the axis of amyloid fibrils and steric zippers.

4.2.1 Finding stacks in amyloid fibrils

In amyloid fibrils, amino acids in the β -sheets tend to stack upon each other. Figure 4.4 shows an example of such stacks. The stacking residue is histidine, which contains the aromatic ring imidazole.

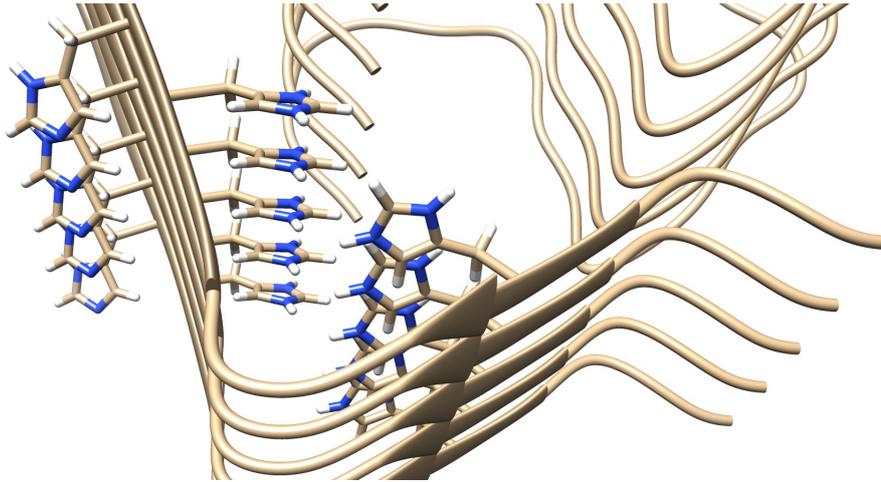


Figure 4.4: The stacking of histidines in the structure of 2MVX.

To be able to use some of the methods, the amino acids present in stacks have to be identified. The location of such amino acids is found using Chimera. A specific amino acid present in three separate stacks is chosen, since three points are needed to create a plane which one of the methods below depend upon. By retrieving the coordinates of the α -carbon in these amino acids and calculating the distances between them one by one, the following restraints were crafted to extract the coordinates of only the stacking α -carbons from the PDB-file.

$$X(C\alpha_1) - X(C\alpha_2) < 4.5 \text{ \AA} \quad (4.1)$$

$$Y(C\alpha_1) - Y(C\alpha_2) < 4 \text{ \AA} \quad (4.2)$$

$$1.5 \text{ \AA} < Z(C\alpha_1) - Z(C\alpha_2) < 15 \text{ \AA} \quad (4.3)$$

Equation 4.1 - 4.3 show the limiting distances for the difference between α -carbons, in the X , Y and Z -direction, respectively. A few of the amyloid fibrils, as well as most of the steric zippers, are expanding more in the X or Y -direction, than in the Z -direction. When this occurs, the restraints needs to be altered. If it expands more in Y , the Y - and Z -restraints are interchanged and the same is done with the X -restraints if the fibril expands in X . By using the restraints as a definition of whether different α -carbons are located in the same stack or not, it is possible to save all stacks in separate arrays. This is done by using Matlab to read through the entire PDB-file and creating a structure with the information. Thereafter, the α -carbons are located and the differences in coordinates between every pair of α -carbons are calculated. If the differences are within the restraints, the α -carbons are added to the same array. If the distances do not match any already checked α -carbon, a new array is created. These generated arrays are later used to calculate the axis of the amyloid fibril.

The method initially developed for calculating an axis uses the atoms found in the aforementioned stacks on matching levels along the fibril. This is done by generating a plane for each level of the stack, and for each plane formulate the normal related to that plane.

The normals from each plane are then combined to form the fibril axis.

Three atoms are extracted from each level in the stack, all with the designation α -carbon, and their positions are treated as the points p_1, p_2, p_3 in \mathbb{R}^3 . These points are in turn used to create two vectors, \vec{V} and \vec{U} , see Equation 4.4.

$$\begin{aligned}\vec{V} &= p_1 - p_2 \\ \vec{U} &= p_1 - p_3\end{aligned}\tag{4.4}$$

The vectors span a plane in \mathbb{R}^3 . By calculating their cross product, a normal (\vec{N}) to the plane is identified, see Equation 4.5.

$$\vec{N} = \vec{V} \times \vec{U}\tag{4.5}$$

The same procedure is repeated for all levels of the fibril, excluding only the top and bottom levels, due to the fibril's tendency to have less strictly ordered β -strands on these levels.

To summarize, a plane is created using the three points and from the plane a normal is defined, as seen in Figure 4.5. This gives a set of normals $\{\vec{N}_2, \vec{N}_3 \dots \vec{N}_{max-1}\}$ where the indices indicates which β -strand along the fibril axis the plane belongs to. A mean normal is computed from the previously calculated set and then scaled to run the entire length of the fibril. This mean normal is used as the axis of the amyloid fibril.

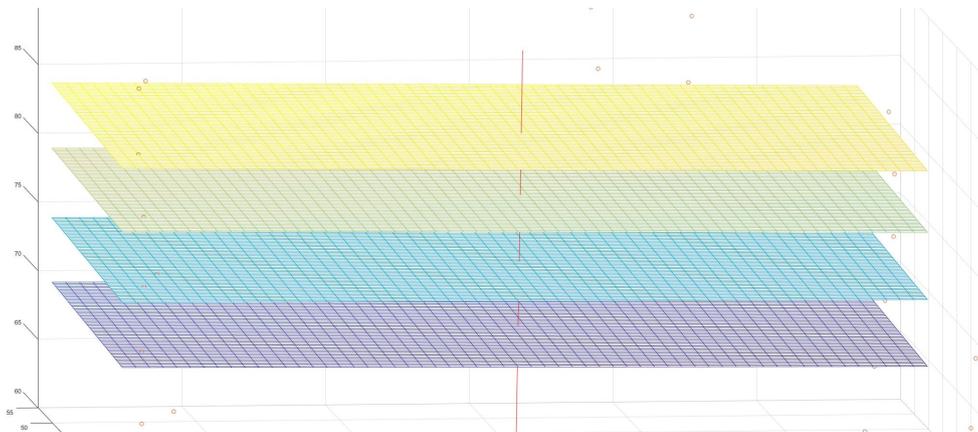


Figure 4.5: A visualization of planes that are used to calculate the axis of the amyloid fibril which is seen as the red line intersecting the planes.

While this method worked for some of the proteins, there existed quite a few for which it did not work as expected, more about this in Section 4.2.3. Because of these issues, the method was not used for the analysis and another method had to be developed which is explained below.

The second method can, due to the fact that the initial method saves all stacks in unique arrays, approximate an axis by calculating the vector between α -carbon number 2 and

$n - 1$ in the arrays, using Equation 4.4. To make the calculations more accurate, the average of three different stacks are used. This is done by saving the X-, Y- and Z-coordinates of the chosen α -carbons in three different arrays, and calculating the mean of all the X-, Y- and Z-coordinates. The average values of the coordinates are used to create a new vector, which represents the axis of the fibril. The axis is then adjusted to originate in a specified point in one of the stacks. An example of these calculations can be seen in Table 4.1.

Table 4.1: Example of how the average stack is calculated. The $StackX_y$ shows which stack (X) and which layer (Y) the values are taken from. The $Vector_{stackX}$ are the vectors calculated between the points in every stack respectively. Finally, *Averages* is the vector that is used as the axis of the fibril, calculated as the average of the three previous vectors.

	X	Y	Z	PDB-position
Stack1 ₂	-12.135	24.379	70.271	6896
Stack1 _{$n-1$}	-10.235	25.473	60.671	6421
Stack2 ₂	-5.681	42.578	62.049	6328
Stack2 _{$n-1$}	-4.127	42.647	76.121	7753
Stack3 ₂	1.737	60.38	66.546	6695
Stack3 _{$n-1$}	2.968	59.859	76.178	7645
Vector _{stack1}	-1.9	-1.094	9.6	
Vector _{stack2}	-1.554	-0.069	-14.072	
Vector _{stack3}	-1.231	0.521	-9.632	
Averages	-1.561	-0.213	-4.701	

4.2.2 Plane of best fit and line of best fit

This method is based on the idea of finding a plane or line of best fit in relation to the entire amyloid structure. It uses a matrix \mathbf{A} containing all the points in the amyloid structure and decomposes it down into useful components via the use of Singular Value Decomposition (SVD). This decomposition is available as a standard Matlab function.

The properties of the components are then used to define a line, or plane, of best fit in \mathbb{R}^3 . The SVD decompose a matrix \mathbf{A} as seen in Equation 4.6.

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (4.6)$$

Matrix \mathbf{A} is compiled from the coordinates of the atoms in the fibril structure and it is decomposed down to the three components \mathbf{U} , $\mathbf{\Sigma}$ and \mathbf{V} . The two components of interest being $\mathbf{\Sigma}$ and \mathbf{V} . Matrix $\mathbf{\Sigma}$ contains singular values that are used in order to extract correct values and matrix \mathbf{V} contains three vectors acting as a basis in \mathbb{R}^3 . The column vectors in this matrix describes a set of unit vectors which are rotated in response to the values of matrix \mathbf{A} . This is shown in figure 4.6, where it is plotted with the centroid of the structure acting as origin.

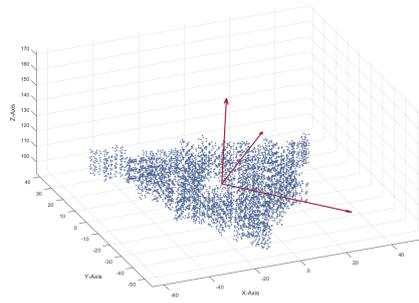


Figure 4.6: Shows the unit vectors as described by the matrix \mathbf{V} originating from the centroid of the fibril structure. Note that the vectors have been scaled by a factor 60. PDB-entry: 2M4J.

In order to generate a plane, the vector in \mathbf{V} with the lowest corresponding value in matrix Σ is extracted. Due to the orthogonal nature of the vectors found in \mathbf{V} , the two remaining vectors span a plane that best fits the structure, and the extracted vector acts as a normal to this plane. In figure 4.6 the normal can be observed as the vector expanding along the vertical axis and the two vectors with highest impact can be seen expanding in the two horizontal directions.

In some cases when the structure mainly expands in one direction, a plane cannot be fitted in a good way. Instead the highest value from matrix Σ is extracted, which means it corresponds to the vector with the best fit to the structure. This vector is defined as the direction of a line, originating from the centroid. In the first case, the normal to the plane acts as the axis and in the later case the line acts as the axis.

4.2.3 Advantages and disadvantages

Each previously mentioned method has its own disadvantages, starting with the deprecated method first mentioned in Section 4.2.1. It used a basic principle of generating planes for each level of the β -sheets, essentially treating each level as a plane and combining the direction of these planes to the axis. This proved inadequate when PDB-entry 2BEG was analysed since the resulting axis was skewed. This leads to the situation seen in Figure 4.7, where the angle is tilted and therefore not representative of the fibril axis. In response, two new methods were developed, the second method presented in Section 4.2.1 and the method presented in Section 4.2.2. Each method has its own advantages and disadvantages.

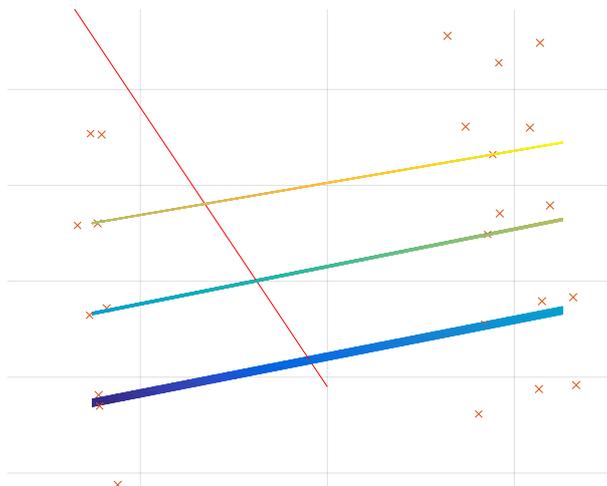


Figure 4.7: A case when the stacks are tilted leading to tilted planes and thus a faulty axis is calculated.

The method plane of best fit has an advantage when analysing large structures expanding primarily in two dimensions, see Figure 4.6. When analyzing smaller structures expanding primarily in one dimension, the plane becomes tilted. This problem can be seen in Figure 4.8, where the plane of best fit intersects the correct axis of the fibril, creating a normal which is rotated from its intended position.

The second method used stacks in the structure and compose an axis from the mean of these stacks. This method seems preferable since it can easily be applied directly on the β -sheets with the right configuration. The problem with this method is that it requires more configuration, finding the right restrictions and choosing specific amino acids which are present in the β -sheets in order to correctly find the stacked that are sought.

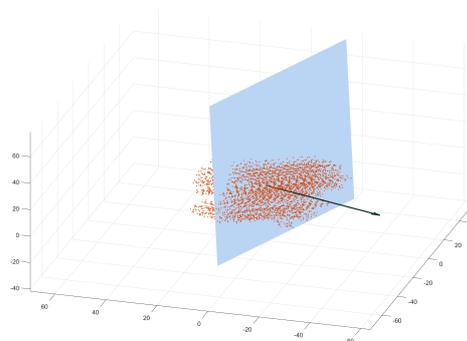


Figure 4.8: Shows how the plane of best fit is intersecting the growth of the fibril due to a structure expanding in primarily one direction. PDB-entry: 2MXU

For the method using the direction of the stacks, issues may arise if the chosen stacks are tilted, however the true axis of the fibril might not be tilted. To solve this issue it is possible to first use Chimera to visually decide which stacks to use.

4.3 Tilt of aromatic rings

As mentioned in Section 2.1.1 there are chemical structures known as aromatic rings. These structures are found as side chains in some of the amino acids; histidine, phenylalanine, tyrosine and tryptophan. This method is taking advantage of the flat ring structure

of the aromatic rings. Each ring is treated as a geometrical plane in \mathbb{R}^3 . Using the normal to this plane and the axis defined in Section 4.2, the angle between them is calculated.

Initially, the coordinates of three atoms in the ring are extracted, these are b_1 , b_3 and b_5 for benzene, see Figure 4.9, and i_1 , i_3 and i_4 for imidazole, see Figure 4.10.

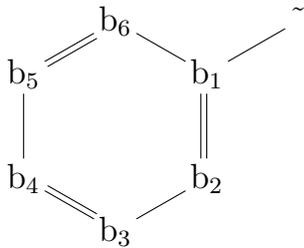


Figure 4.9: Benzene ring, present in phenylalanine, tryptophan and tyrosine.

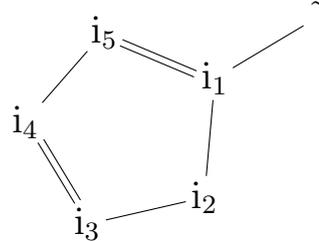


Figure 4.10: Imidazole present in histidine.

These points are in turn used to create two vectors \vec{V} and \vec{U} . For benzene rings see Equation 4.7 and for imidazole see Equation 4.8.

$$\begin{aligned}\vec{V}_{\text{benzene}} &= b_1 - b_3 \\ \vec{U}_{\text{benzene}} &= b_1 - b_5\end{aligned}\tag{4.7}$$

$$\begin{aligned}\vec{V}_{\text{imidazole}} &= i_1 - i_3 \\ \vec{U}_{\text{imidazole}} &= i_1 - i_4\end{aligned}\tag{4.8}$$

In each case, the calculated vectors span a plane in \mathbb{R}^3 and by calculating the cross product of the vectors, a new vector \vec{N} is obtained, which acts as normal to the aromatic ring, see Equation 4.5.

To identify the angle θ between the normal to the aromatic ring and the axis of the amyloid fibril, the dot product operation is used on the normal \vec{N} and the axis \vec{A} , see Equation 4.9

$$\vec{N} \cdot \vec{A} = \|\vec{N}\| \|\vec{A}\| \cos(\theta) \rightarrow \theta = \arccos\left(\frac{\vec{N} \cdot \vec{A}}{\|\vec{N}\| \|\vec{A}\|}\right)\tag{4.9}$$

Figure 4.11 shows an example of the normals to the aromatic rings plotted along with the axis of the fibril.

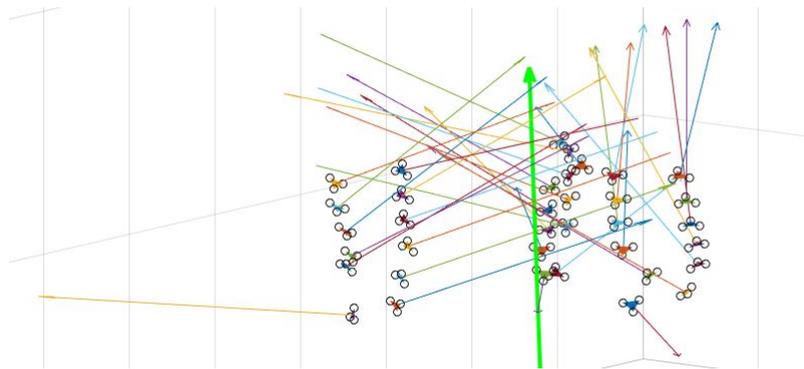


Figure 4.11: Normals of the aromatic rings of the amino acid phenylalanine, calculated and plotted alongside the axis of the fibril. PDB-entry: 2LMQ.

This information is in turn compiled into a table, together with additional information about the scanned structure, namely which protein structure, which amino acid, which data set and where in the structure the ring was located.

5

Results

This chapter begins with displaying the results from the analysis of direction and distribution of amino acid side chains. Firstly, results from the analysis of the bends of amyloid fibrils and β -solenoids are presented. Secondly, results from the analysis of the charged amino acids within amyloid fibrils and β -solenoids are presented. Followed by this are the results from the analysis of the aromatic rings. The raw data which calculations are based on are available as Supplementary material¹ and the data which are used to produce the charts seen in this section are available in Appendices B-F.

5.1 Direction and distribution of amino acids

The results are presented in regards to both individual amino acids as well as their organization according to their groups based on chemical properties. It has to be pointed out that the different groups are not equally represented in regards to the number of different kinds in each group. There are nine non-polar, six polar, three basic and two acidic amino acids. The distribution of the groups might differ because of their different sizes. The classifications of amino acid residues are available in Appendix A.

5.1.1 Bends of amyloid fibrils

The analysis of the bend segment was performed on 15 amyloid fibrils. The original data set, presented in Table 3.1, contains 17 fibrils but two of the structures (2KIB and 2M5N) did not contain any bends and were therefore excluded. In total, the set of amyloid fibrils contained 214 bends consisting of 1251 amino acids.

The total distribution of the amino acids involved in the bends is presented in Figure 5.1. The pie chart shows the distribution of the chemical groups and the bar chart shows the distribution of individual amino acid residues. The pie chart shows that 54% of the amino acids present within the bends are non-polar, 22% are polar, 14% are acidic and 10% are basic. When observing the bar chart to the right, it becomes clear that glycine (G) is most common, occurring 306 times. Glycine is followed by alanine (A), serine (S), lysine (K) and asparagine (N), all of them occurring between 114-164 times. It is interesting to notice that tryptophan (W), glutamin (Q) and cysteine (C) do not appear at all in the bend segment of the amyloids.

¹ https://github.com/DATX021607/DATX02_16_07

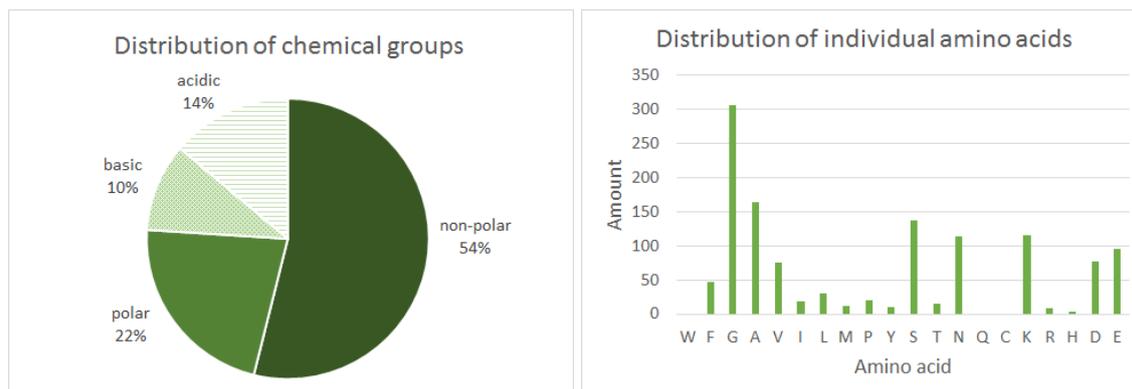


Figure 5.1: Distribution of amino acids involved in the bends. In the pie chart (left), showing the proportion of the chemical groups, the non-polar group is the largest, followed by the polar, the acidic and finally the basic. In the bar chart (right), showing the proportion of individual amino acids, it is obvious that glycine (G) is very common in the bends. Tryptophan (W), glutamin (Q) and cysteine (C) are the only ones not present at all in the bend segment of amyloids.

Moving on to the distribution between the two directions, Figure 5.2 shows the total distribution. The pie chart shows that the majority, 65%, of the amino acid side chains point outwards. The fact that more amino acids prefer facing the outside of the protein structure could have two explanations. Firstly, the definition of the direction favours the outward direction. If an amino acid side chain is located between the two directions it is classified as outwards. A small amount of the side chains were positioned between the two directions which means that this has affected the result in some degree. Secondly, it could be a consequence of environmental factors. The chemical surroundings could differ between the inside and outside of the fibril and therefore favour this arrangement. The inside of the fibril could also be more crowded and will therefore prevent too many residues from arranging themselves with their side chains inwards.

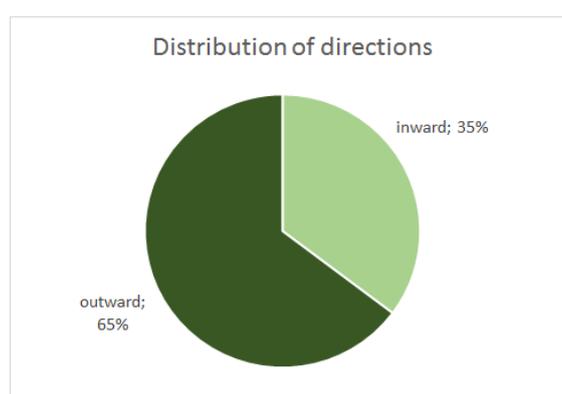


Figure 5.2: The distribution of the amino acids occurring within the bends of the amyloid fibrils. The majority, 65%, point outwards relative the fibril core.

When observing the distribution within each direction, as seen in Figure 5.3, it is possible to see that the distribution of amino acid side chains, in regards to their chemical groups, differs a bit. The non-polar amino acids make up almost the same proportions; 53% of

the side chains facing outwards and 55% of the ones facing inwards. The polar amino acids are relatively more common amongst the ones facing inwards, representing 30%, than the ones facing outwards, representing only 18%. The charged amino acids show the opposite result. The acidic and basic amino acids representing 17% and 12% of the outward residues, respectively, and only 9% and 6% of the inward residues, respectively.

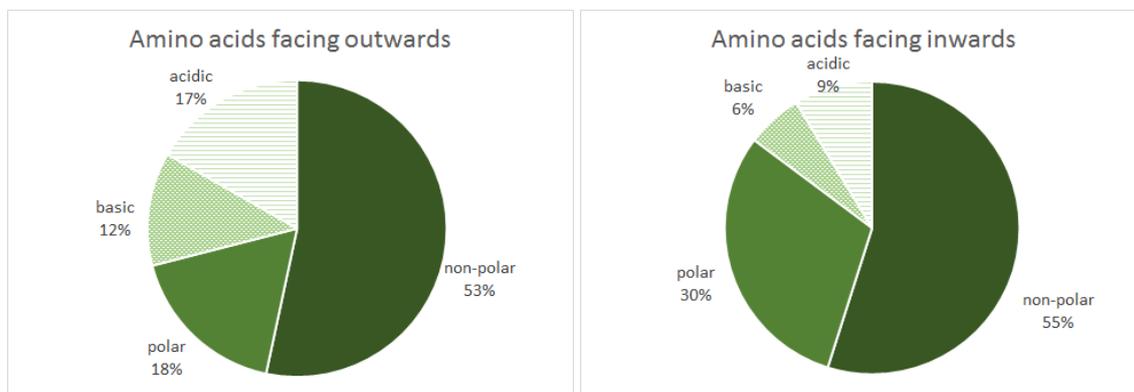


Figure 5.3: The diagram to the left shows the distribution of the amino acids facing outwards and the diagram to the right shows the distribution of amino acids facing inwards. The non-polar groups are most represented in both cases while the polar amino acids make up a larger portion of the amino acids facing inwards than outwards.

5.1.2 Bends of β -solenoid proteins

The data set of five β -solenoid proteins contained 135 bends with a total number of 494 amino acids present within the bends. The distribution of individual amino acids is presented in Figure 5.4, where it is evident that glycine (G) is the most abundant amino acid appearing 80 times. Glycine is followed by asparagine (N) and serine (S), appearing 69 and 58 times, respectively. When dividing the amino acids into their chemical groups it is clear that the frequency of non-polar and polar amino acids is almost the same, 39% and 40% respectively. The charged groups are also about equally large with the acidic amino acids representing 11% and the basic 10% of the total distribution.

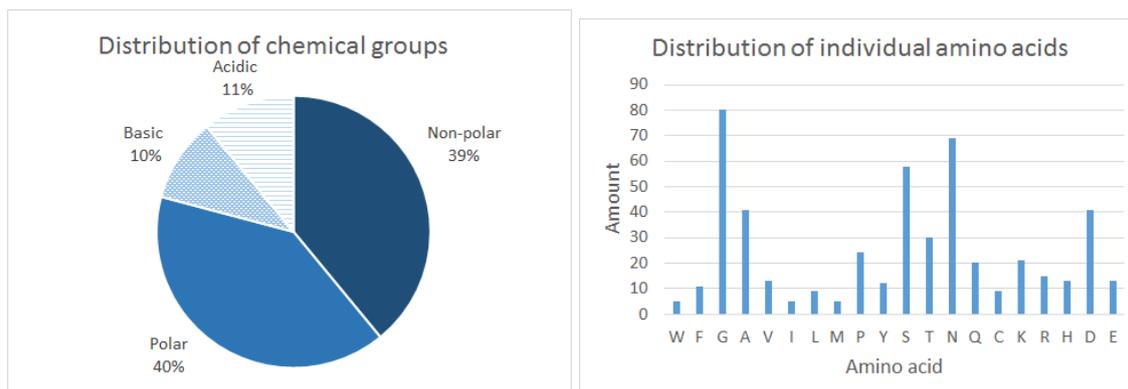


Figure 5.4: The bar diagram shows the distribution of individual amino acids. Glycine (G) is overrepresented occurring 80 times, followed by asparagine (N) occurring 69 times and serine (S) occurring 58 times. The pie diagram shows the distribution of the amino acids present in the bends when divided into their chemical groups. The occurrence of polar and non-polar amino acids is equally common. The occurrence of basic and acidic amino acids is equally common.

When analyzing the bends of the β -solenoids, it was not possible to register the direction of glycine since the PDB-files did not contain information about the hydrogen atoms in some of the β -solenoids. This only affected glycine since its side chain consists of a single hydrogen atom. Glycine is therefore separated from the results concerning the directions of the amino acids in β -solenoids. Figure 5.5 displays the distribution of the directions among all amino acids except glycine. The proportion of glycine is instead shown separately to show the large amount of glycine in the analysis. Since glycine is very common within the bends, as seen in the Figure 5.4, the exclusion of glycine will have a substantial effect on the result concerning the directions of amino acids.

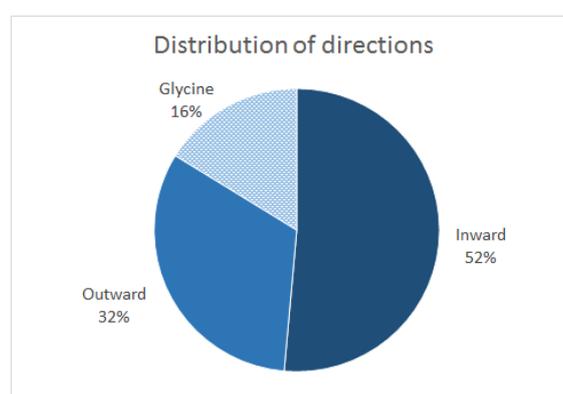


Figure 5.5: Distribution of the directions among the amino acids in bends of β -solenoids. Glycine is excluded from the analysis but included in the diagram to show that glycine represents a substantial amount of the amino acids.

The distribution of the amino acids facing outwards is presented to the left in Figure 5.6. The pie chart shows that 54% are polar, 28% are non-polar, 12% are basic and 6% are acidic. To the right is the distribution of the amino acids facing inwards. 44% are polar, 27% are non-polar, 17% are acidic and 12% are basic. This result is affected by the

previously mentioned exclusion of glycine. Glycine is a non-polar amino acid, which means that the non-polar groups in the diagrams are smaller than they should have been if glycine had been included.

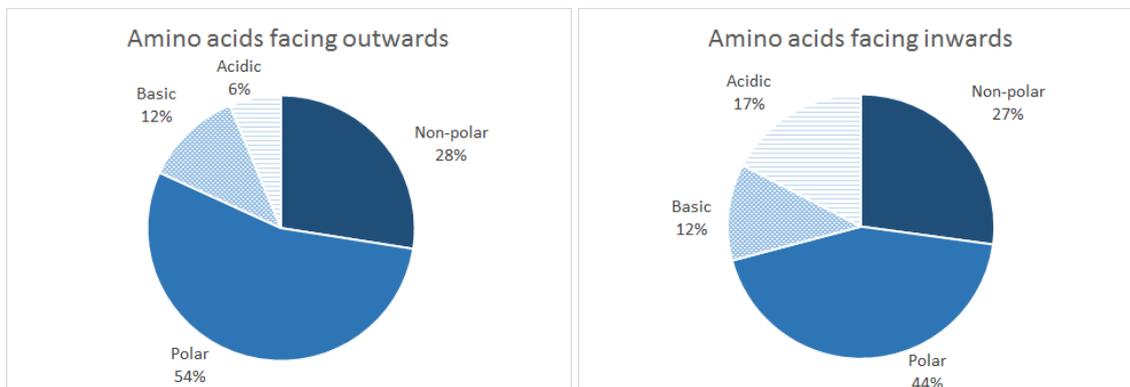


Figure 5.6: Distribution of amino acids within each direction, in bends of β -solenoids. Polar amino acids are most represented in both directions with 54% of those facing outwards and 44% of those facing inwards. The non-polar groups are the second largest but are affected by the exclusion of glycine since glycine is non-polar. The groups with charged amino acids are not of equal sizes. Of the amino acids facing outwards 12% are basic but only 6% are acidic. Of the amino acids facing inwards 12% are basic but 17% are acidic.

5.1.3 Charged amino acids in amyloid fibrils

As with the bends, there were 15 amyloid fibrils included in the analysis of the charged amino acids, see Table 3.1. The same two structures were excluded, 2KIB and 2M5N. Within these 15 structures, 1478 charged amino acids occurred. Figure 5.7 displays a summary of the distribution of the charged amino acids in the 15 amyloid fibrils. The amino acids more present are lysine (K) and glutamic acid (E). The under-represented amino acid is arginine (R), only occurring 71 times compared to lysine and glutamic acid, which appears more than 400 times each.

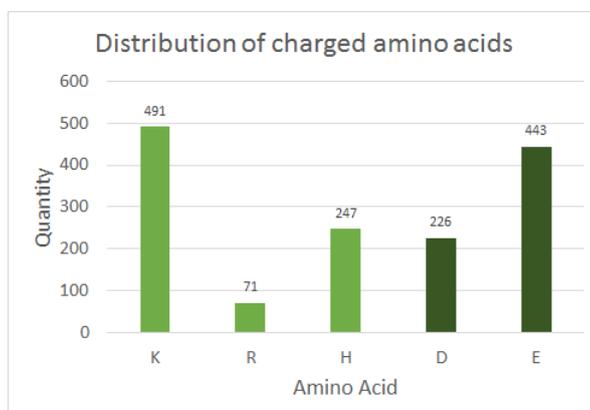


Figure 5.7: Distribution of charged amino acids in amyloid fibrils. The three bars to the left are the basic amino acids and the two bars to the right, dark coloured, are the acidic. Lysine (K) and glutamic acid (E) are over-represented, while arginine (R) is under-represented.

Figure 5.8 shows that 89 % of the charged amino acids are located in an outwards manner relative to the amyloid fibril core and 11% are located inwards. It also indicates that the distribution between acidic and basic amino acids are quite balanced, 45 % and 55 % respectively.

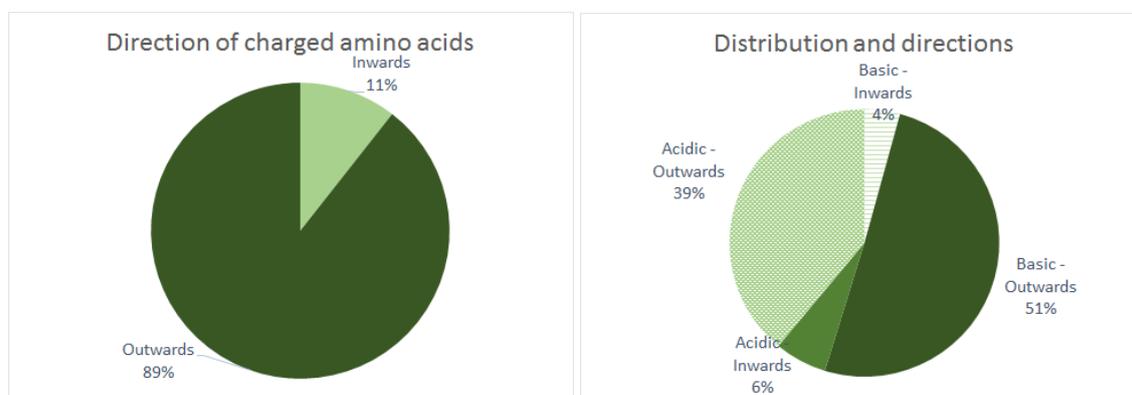


Figure 5.8: Direction and distribution of charged amino acids in amyloid fibrils. The pie chart to the left shows the distribution between the directions of outward and inward. The pie chart to the right indicates the same distribution, but with a further classification according to the charge of the amino acids.

The examples that follow are given to emphasize some general distributions, as well as singular arrangements of the charged amino acids. First, a couple of general arrangements are shown, these do not include any measurements regarding electrostatic attraction. Secondly, some non-general cases are presented. These refer to when the amino acid side chains are located in the center of the amyloid fibril core, when opposite charges are located nearby each other and when there might be unbalances between charges. The figures used to describe the arrangements depict polar and charged amino acids with blue and red colours, due to the presence of nitrogen and oxygen. If the figure only shows, or has highlighted charged amino acids, blue can be regarded as a positive charge, and the red represents a negative charge. The non-polar amino acids lack these elements and are therefore white in all of the figures.

Many of the charged amino acids are positioned in the bends of the fibrils or in the end of the peptide sequences. This occurs in several of the structures related to the A β -protein, such as 2LMN, 2LMP and 2LMQ. Figure 5.9 displays the charged amino acids in 2LMQ.

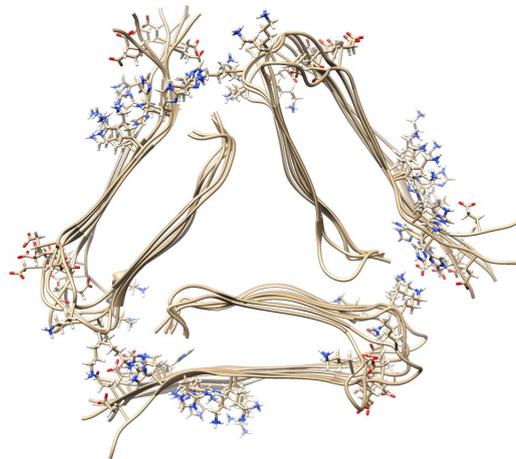


Figure 5.9: The figure shows all the charged amino acids present in the PDB-entry 2LMQ. All of the charged amino acids occur in the bends or the end of the fibril. A similar distribution are present in the related structures of 2LMN and 2LMP.

In several structures, all the charged amino acids are facing the exterior. This recurring distribution is seen in Figure 5.10, which displays the charged amino acids in the structure of 2RNM. 2RNM along with 2KJ3, are derived from the same prion protein and they do have nearly identical structures. This distribution is also seen in 2N1E, with all the charged amino acids perfectly arranged outwards.

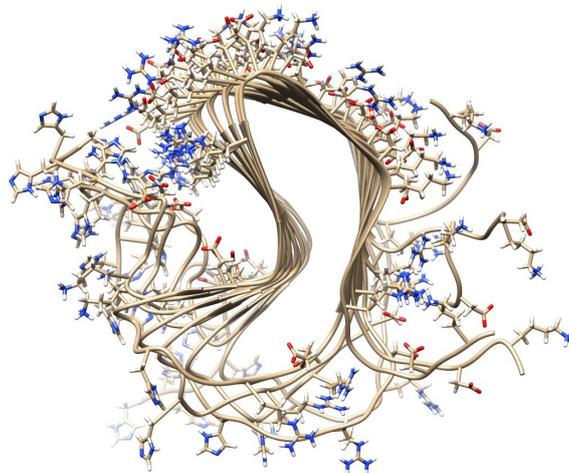


Figure 5.10: The figure shows an example of a structure which has its charged amino acids facing the exterior. PDB-entry: 2RNM.

In the following examples, distances between charged residues have been measured in order to evaluate the presence of electrostatic attractions. The details of these measurements are available in Appendix B.

2NNT has E7² and R24 facing each other on opposite β -sheets at a distance of 2.6-4.4 Å, shown to the left in Figure 5.11, which might give rise to electrostatic attraction and therefore help the stabilization of the interface between the sheets. The surrounding residues in the interface are mainly polar, shown to the right in Figure 5.11. A similar arrangement is also seen in 2BEG, but here the surrounding residues in the core are non-polar. In 2BEG, D23 and K28 are closely arranged, lysine (K28) is present in the bend of the fibril and asparagine (D23) in a β -sheet. The distance between them are approximately 2.5 Å.



Figure 5.11: The structure of 2NNT. In the right figure, the amino acids facing the core are presented, which forms an interface between the two β -sheets in the structure. Two out of ten residues in the interface are polar. In the left figure the charged amino acids, E7 and R24, has been extracted. These charged residues are 2.6-4.4 Å apart.

The structure of 2M4J is strictly ordered, seen in the top of Figure 5.12. Despite the ordered arrangement, it is a complex structure which makes it hard to decide where the core of the fibril is. In this case, K28 and D23, are regarded as being present in a bend, facing the fibril core and therefore being located inwards. These residues are isolated in the lower left structure in Figure 5.12, they are surrounded with non-polar amino acids and separated by a distance of 2.56-2.58 Å. Another couple of charged amino acids, H13 and E11, are present in what is regarded as the core, even though it is neither a turn nor an interface at the position. These amino acids are shown in the lower right structure of Figure 5.12. They are separated by a short distance, 3.70-3.78 Å, and the amino acids nearby are polar compared to the previously mentioned core that were non-polar.

²The letter represent an amino acid, the one letter abbreviations are available in Appendix A, and the number indicates the position of the amino acid in the sequence.

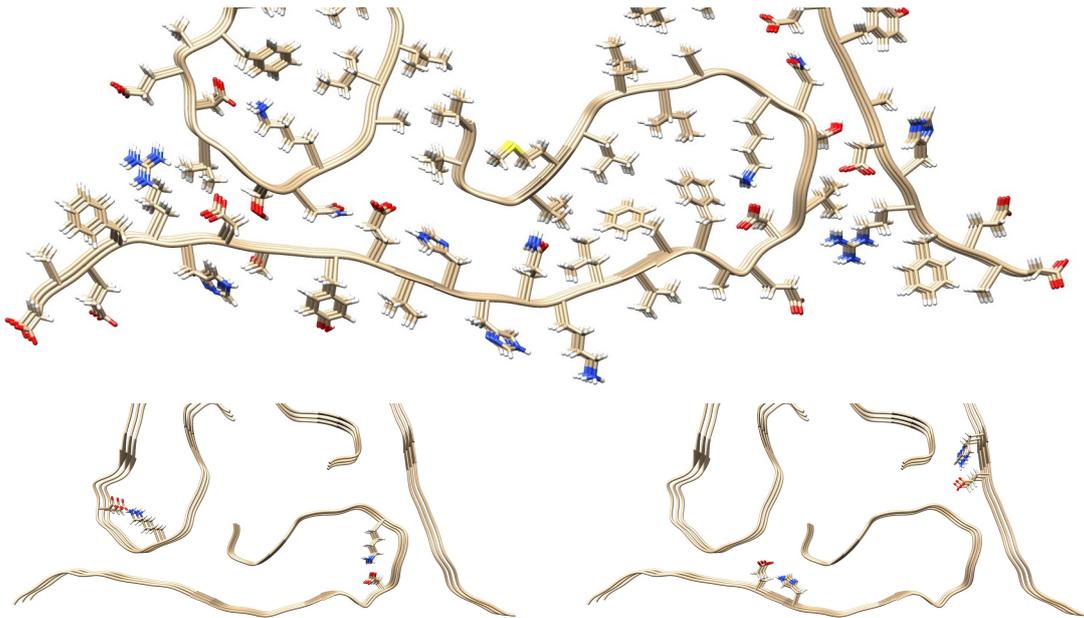


Figure 5.12: The figure displays parts of the structure of 2M4J, with all the amino acids in the structure shown. The lower left shows D23 and K28, and the lower right shows E11 and H13.

The protein structure 2N0A has two interesting parts regarding the charged amino acids, these are shown in Figure 5.13. First, there is a glutamic acid at position 61 (E61) which is facing the core of the fibril, but it is not stabilized by an oppositely charged residue. E61 is highlighted in red in the left figure in Figure 5.13 along with some of the amino acids in the core. It can be seen that the amino acids next to E61 are polar and those thereafter are non-polar. Next, K80 and E46 are oppositely charged and within 2.84-2.85 Å apart, these amino acids are shown in the right figure in Figure 5.13. The interaction between E46 and K80 might stabilize the interface of β -sheets present next to the two charged amino acids.

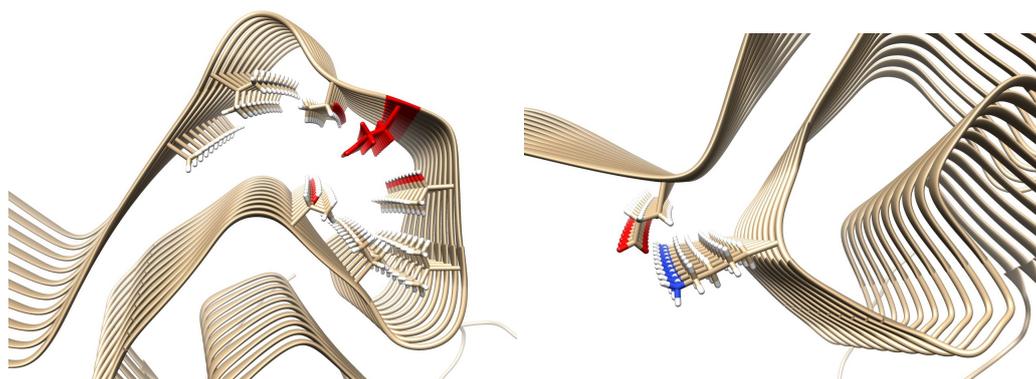


Figure 5.13: The structures display parts of the PDB-entry 2N0A. The left figure shows the amino acids present in a section of the fibril core, with the charged glutamic acid (E61) highlighted in red. The residues next to E61, which has a red atom, are polar and those lacking the colour are non-polar. The right figure shows K80 and E46, two oppositely charged amino acids located just outside the core of the fibril.

5.1.4 Charged amino acids in β -solenoids

The data used for the analysis of the charged amino acids in the β -solenoids were the same as for analysis of the bends and consisted of five different structures, see Table 3.3. The number of charged amino acids in total were 230, Figure 5.14 displays the distribution among these. The figure shows that aspartic acid (D) is a bit over-represented, occurring 72 times. Histidine (H) is a bit under-represented, only occurring 22 times. Lysine (K), arginine (R) and glutamic acid (E), are more evenly distributed with around 50 entries.

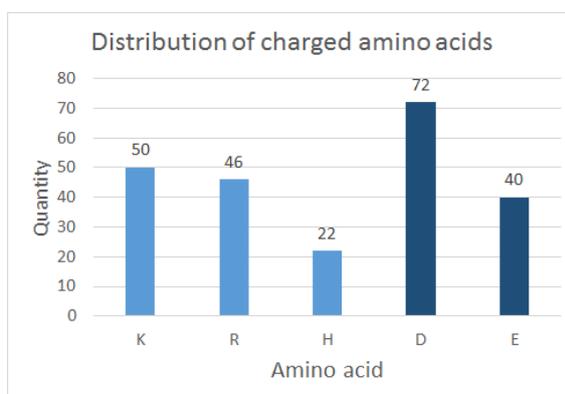


Figure 5.14: Distribution of charged amino acids in β -solenoids. The three bars to the left are the basic amino acids and the two bars to the right, dark coloured, are the acidic.

The result of the charged amino acids shows that, as for the amyloid fibrils, the majority are located in an outwards manner. Figure 5.15 shows that only 7% of the charged amino acids, in the β -solenoids, are located inwards. The distribution of the directions between acidic and basic amino acids are almost evenly balanced, with 49% for the acidic and 51% for the basic, see the chart to the right in Figure 5.15. Since the distribution of charged amino acids which are facing inwards are under-represented, it is worth noticing that a difference in 2% almost doubles the number of amino acids.

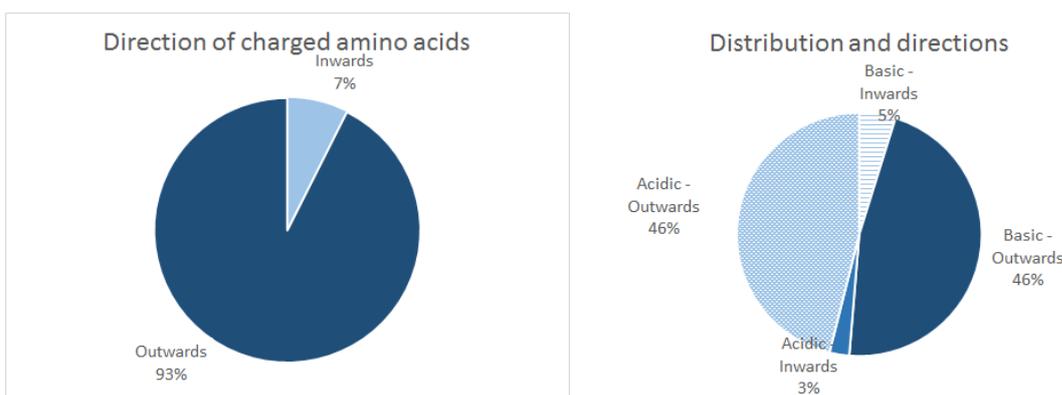


Figure 5.15: Direction and distribution of charged amino acids in β -solenoids. The chart to the left shows that the vast majority of the charged amino acid side chains face outwards. The chart to the right displays that the directions between acidic and basic amino acids are almost even.

The β -solenoids have a more general structure compared to the amyloid fibrils. They all consist of one chain, which is twisted around the axis of the solenoid, giving them a distinct core. The charged amino acids are evenly distributed throughout the structure of the β -solenoids, and there are no stacks of charged amino acids present in the structure. By

looking at the bar chart in Figure 5.4 it can be concluded that 103 of the charged amino acids are located in bends, which is 45% of all the charged amino acids. Figure 5.16 shows the structure of 1EE6, which is displayed to emphasize the general structure of the β -solenoids and the distribution of their charged amino acids. There are no observations of electrostatic interactions which may affect the stability of the β -solenoid.

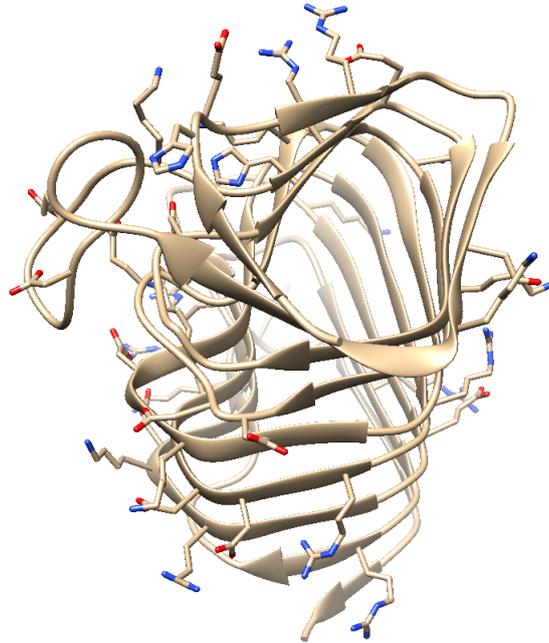


Figure 5.16: The figure shows all the charged amino acids present in the PDB-entry 1EE6. The charged amino acids are evenly distributed throughout the protein structure.

5.2 Tilt and stacking of aromatic rings

In this section, the results from the methods used to determine and calculate the angles between the aromatic rings and the axis of the fibril will be presented. Following this, the data is shown again, adjusted so that the angles lie within the interval $0^\circ - 90^\circ$. Lastly, the ratio between stacked and non-stacked aromatic rings is presented, for some amyloids in the data set.

The analysis of aromatic rings was performed on 12 out of the 17 amyloid fibrils in the data set, resulting in 362 data points. The excluded structures were 2RNM, 2LMN, 2BEG, 2E8D and 2N0A. These were excluded due to limitations in the used method. The 362 data points represent the angles of amino acids with aromatic rings occurring in β -sheets, with a few exceptions. These exceptions are when the structure is perfectly ordered, leading to similarities with β -sheets, while not being classified as such, with one example being 2MVX. For the steric zippers, 15 out of the 17 structures in the data set were analyzed, giving 39 data points. The excluded structures were 3NVG and 3NVH. The data used for the following charts can be found in Appendix C.

Figure 5.17 shows the distribution of measured angles as well as the quantity of the amino acids containing aromatic rings. There is a clear predominance of the amino acid phenylalanine, while the amino acid tryptophan barely exists in any of the proteins. Histidine and tyrosine are present in approximately the same quantity (81 and 66, respectively). The data shows a distribution with two peaks, approximately $40^\circ - 70^\circ$ and $115^\circ - 145^\circ$. In between these, the number of angles are low, not rising above a count of seven, which is the case at $75^\circ-80^\circ$.

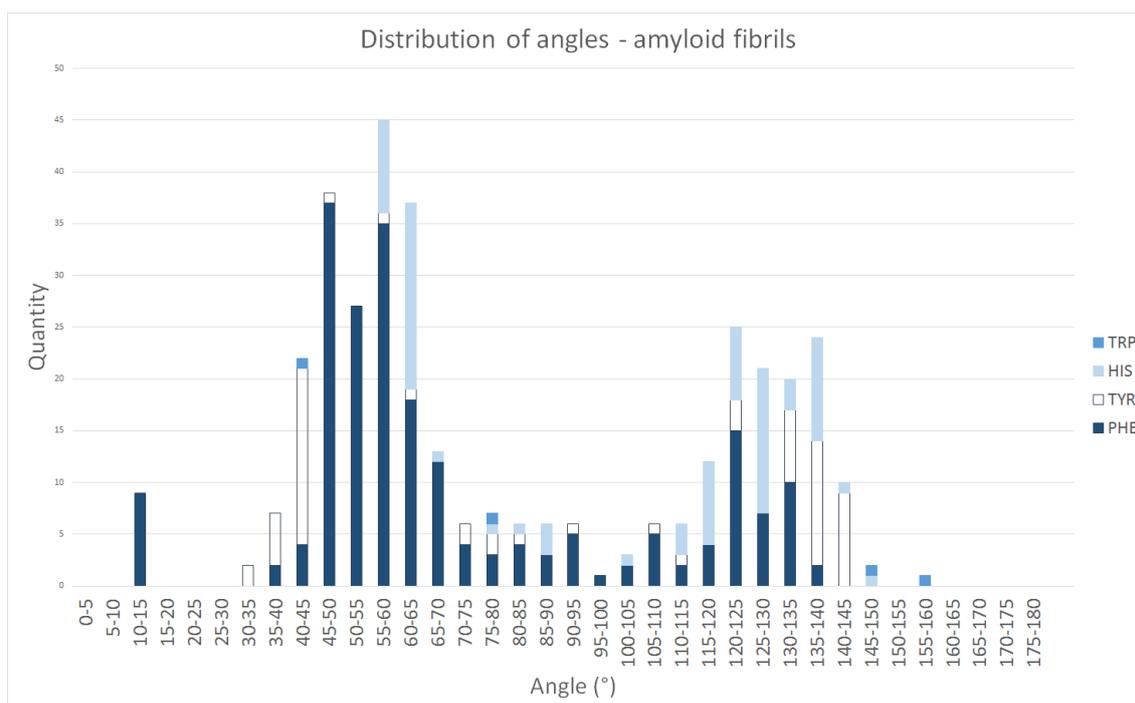


Figure 5.17: A chart showing the distribution of angles of the aromatic rings, in the amyloid fibrils. The distribution shows two peaks, first and strongest around $55^\circ-60^\circ$ and a second, weaker peak at around $120^\circ-125^\circ$.

As mentioned in Section 3.1, steric zippers are short peptides generated into stacks by transformation matrices. Due to this, each layer in the structure is a copy of the original layer, which means that all angles are the same. A single data point for each unique stack is therefore used. Figure 5.18 shows that the distribution of angles for the different amino acids in the set of steric zippers is quite similar to the distribution for the amyloid fibrils. Phenylalanine is once again the dominant amino acid, with tyrosine closely following. There are only four histidine and tryptophan is not present. While the distribution of angles are more spread out in the steric zippers compared to the amyloid fibrils, the majority are still in approximately the same span as for the amyloid fibrils.

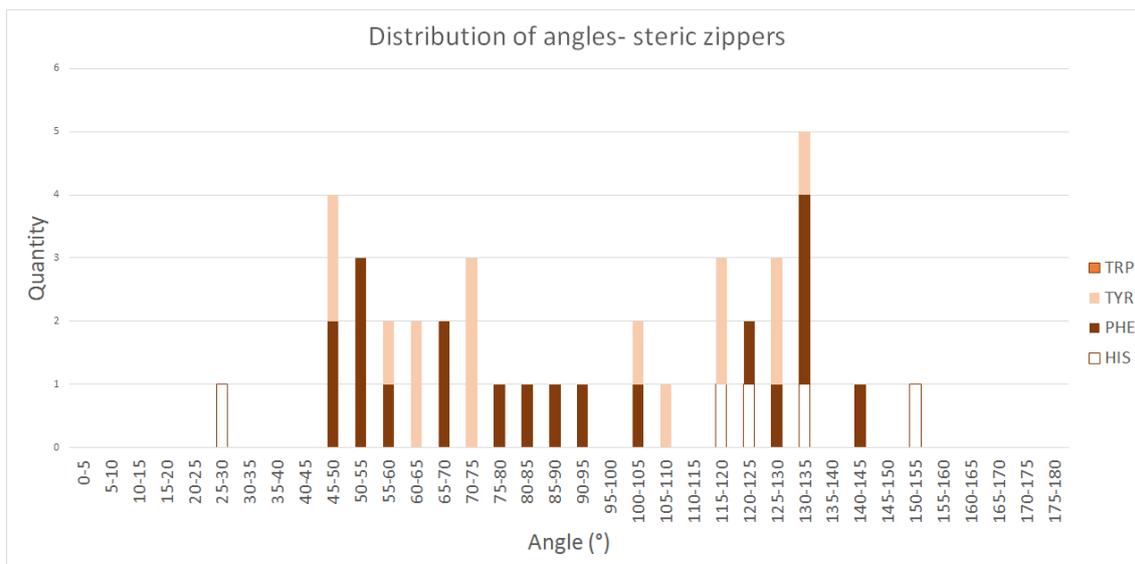


Figure 5.18: A chart showing the distribution of angles of the aromatic rings, in the steric zippers. All data points represents the angle in an entire stack. There are minor peaks around 50° and 130°

It was discovered that the normals to some aromatic rings were reversed, facing the opposite direction than intended. This issue will be further discussed in Section 6.2. To circumvent the issue, each angle above 90° is adjusted by subtracting the angle from 180° and thereby only have angles spanning from 0° to 90°. This generates the data that can be observed in Figure 5.19. It shows an interesting distribution, with angles grouping up between 35°-70°.

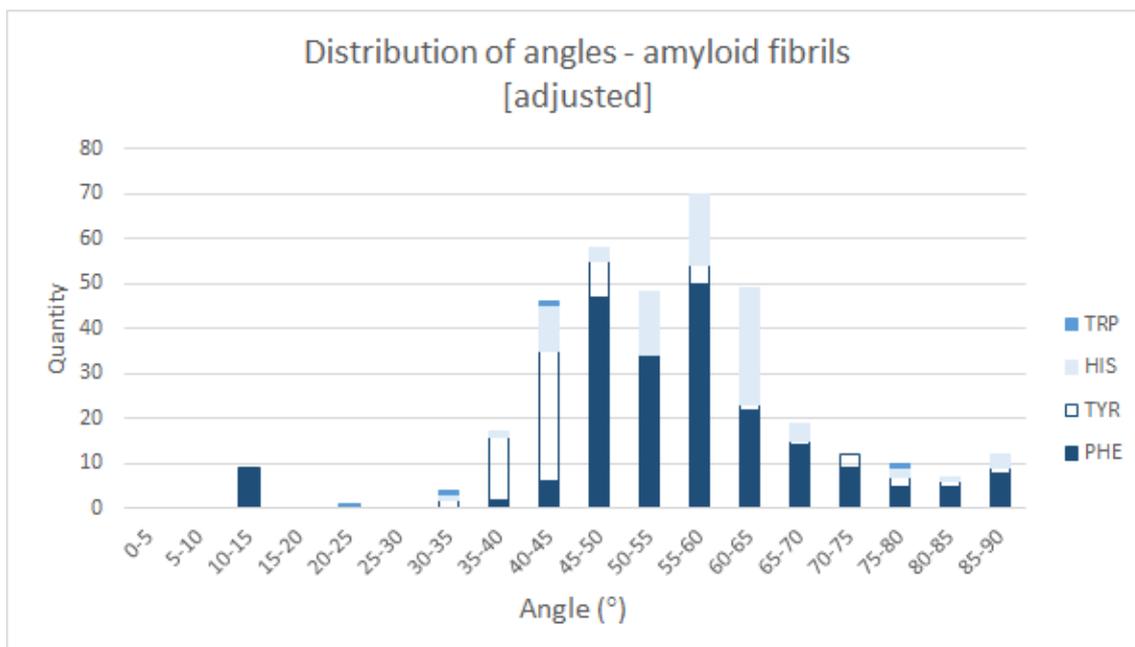


Figure 5.19: A bar chart showing the adjusted distribution of angles of the aromatic rings, in the amyloid fibril.

In comparison, the data of the adjusted steric zippers, see Figure 5.20, shows a more uneven distribution, with a peak around 45°.

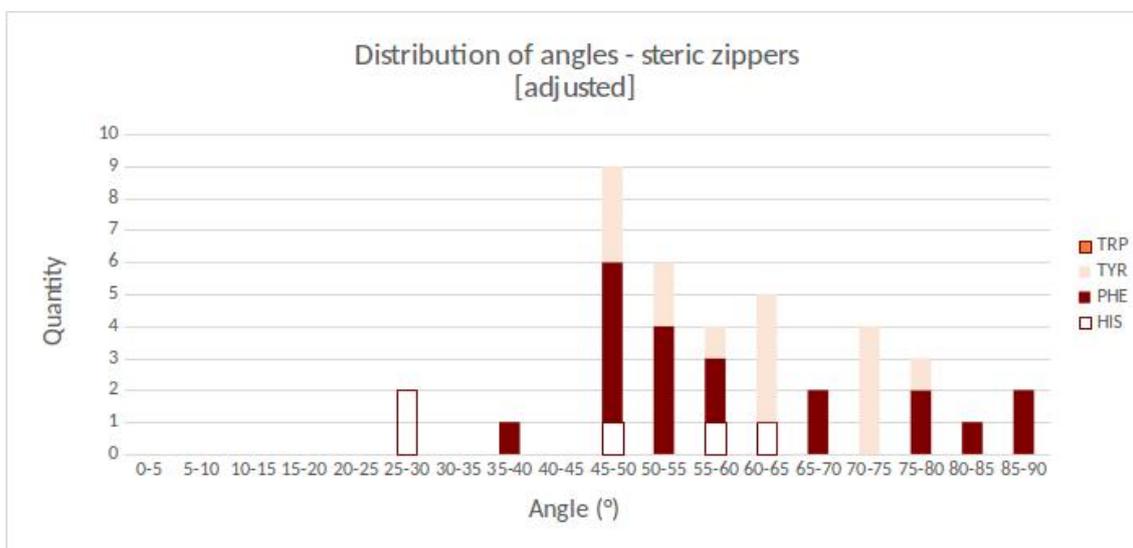


Figure 5.20: A chart showing the adjusted distribution of angles of the aromatic rings, in the steric zippers. Every point of data shows the angle for an entire stack in a steric zipper, due to them all sharing the same orientation in the structure.

11 out of the 17 amyloid fibrils were further analyzed. The same structures were used as in the previous analysis, with the extra exclusion of the PDB-entry 2N1E. The 525 generated data points resulted in a total of 96 stacks. These were distributed among the different amino acids as 44 stacks of phenylalanine, 19 stacks of tyrosine, 30 stacks of histidine and three stacks of tryptophan. The number of stacks for each amyloid can be seen in Table F.1 in Appendix F.

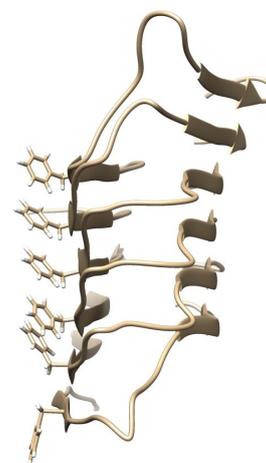


Figure 5.21: A stack of phenylalanine showing tendencies to lose the alignment of the stack in the lower part. PDB-entry: 2LMQ.

The second analysis was performed in order to compare the tilt of the aromatic rings within the same stack. This might indicate whether the stacks are arranged uniformly or not. An example where both these situation occur can be seen in Figure 5.21. The phenylalanines in the middle of the stack align, those in the bottom of the stack do not. The alignment of the aromatic rings in the stack were determined by comparing the angle of an aromatic ring in relation to the rest of rings in the same stack. If two adjacent aromatic rings differ by less than 5° they are classified as stacking. Since this data is composed of all aromatic rings in the amyloid structure, not only from the β -sheets, a more uneven distribution is expected between the stacked and non-stacked rings.

The distribution of stacked and non-stacked occurrences of the benzene ring found in the

side chain of phenylalanine can be seen in Figure 5.22. The data presents a trend that the benzene rings prefers to stack. The non-stacking rings can be attributed to multiple sources. Starting with PDB-entry 2KJ3 which is a small structure, containing only three benzene rings which are outside the β -sheet. Secondly, in a few cases where the rings are located at the top or the bottom of the β -sheet they tend to be less ordered, while inside the sheet the rings stack in a uniform fashion. This situation is present in the PDB-entries 2LMP, 2LMQ, 2MXU and 2NNT.

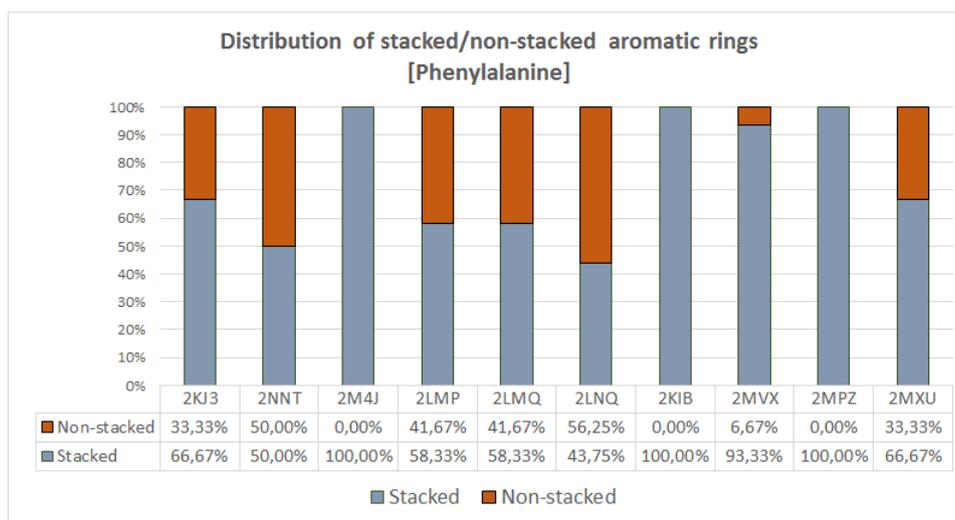


Figure 5.22: A bar chart that shows the ratio of stacked compared to non-stacked benzene rings in the side chain of phenylalanine. The data indicates that the ordered stacks occur inside the β -sheets. Stacks inside β -sheets are the PDB entries 2LMP, 2LMQ, 2MXU and 2MXU.

The data for the imidazole ring found in the side chain of histidine is presented in Figure 5.23. First glance shows a low representation of uniformly stacked imidazole rings in the structures. Visual inspection shows that in the majority of these cases the histidine is present outside of the β -sheets, more specifically in the PDB-entries 2KJ3, 2LMP and 2LMQ. In the rest of the analyzed structures histidine are present in their β -sheets.

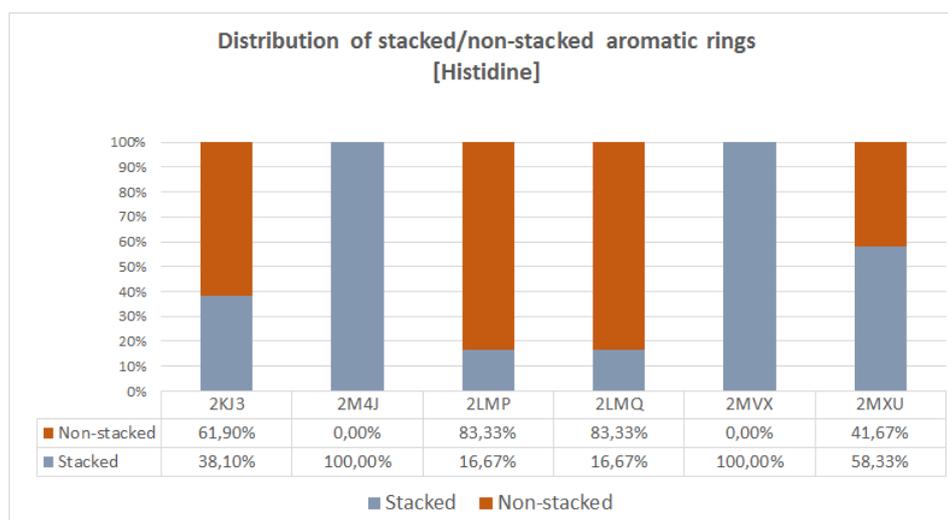


Figure 5.23: A bar chart that shows the ratio of stacked compared to non-stacked imidazole rings in the side chain of histidine. The majority of disordered stacks occur outside the β -sheets, the remaining stacks point towards an ordered stack.

The aromatic ring benzene found as side chain to tyrosine show an inclination to a less uniform structure compared to the amino acids just mentioned. The data, presented in Figure 5.24, shows a higher ratio of unstacked rings. In this case the PDB-entries 2KJ3 and 2NNT has tyrosine in the β -sheet, while 2M4J, 2M5N, 2LMP and 2LMQ do not. This indicates that in the case of tyrosine, the rings do not stack. Note, take into consideration that 2M4J and 2M5N are considered to be perfectly ordered, despite the lack of β -sheets as mentioned in Section 5.1.3.

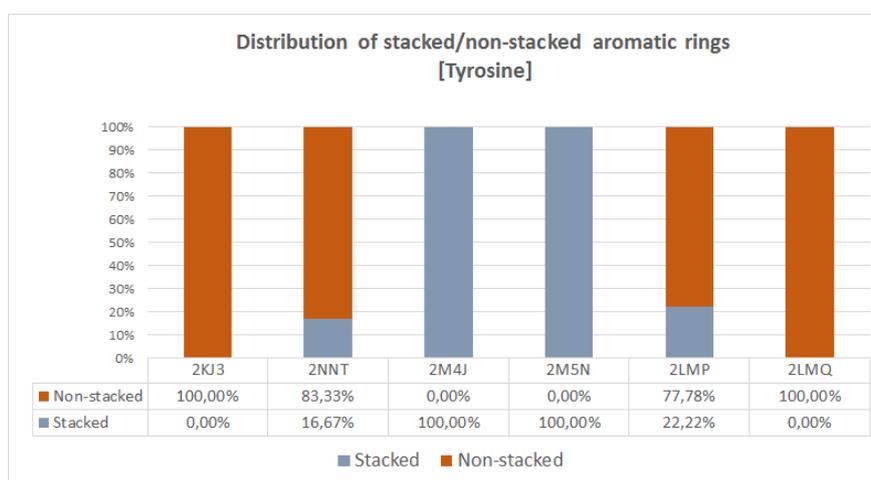


Figure 5.24: A bar chart that shows the distribution of stacked compared to non-stacked benzene rings in the side chain of tyrosine. The data indicate a favour of unordered stacks inside the β -sheets in the case of PDB-entries 2M4J, 2M5N, 2LMP and 2LMQ.

The last analyzed aromatic ring is the benzene ring present in the side chain of tryptophan. As mentioned previously this is an under-represented amino acid in amyloid structures and in effect, only two of the scanned structures contain it, see Figure 5.25. In the case of PDB-entry 2KJ3, tryptophan is found in one stack of three rings, outside the β -sheets.

In 2NNT, tryptophan can be found both in- and outside the β -sheet which can explain the inconsistent stacking of the structure.

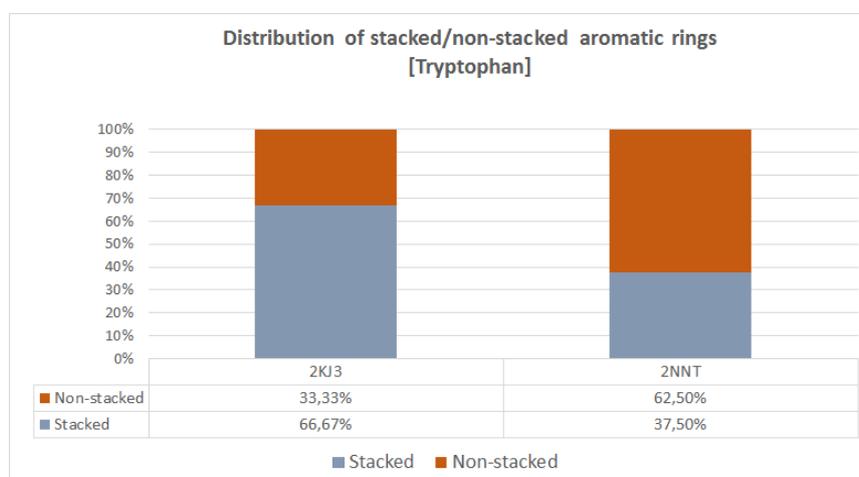


Figure 5.25: A bar chart that shows the ratio of stacked compared to non-stacked benzene rings in the side chain of tryptophan.

6

Discussion

This chapter brings up some discussion points regarding the results from the analyses, as well as reflections related to the chosen methods and possible improvements.

6.1 Direction and distribution of amino acids

The analysis of the direction and distribution of certain amino acid side chains was performed on amyloid fibrils and β -solenoids. The specific aspects studied were the amino acids in the bends and the charged amino acids throughout the structures. Here, a discussion of the results is presented.

6.1.1 Bends of amyloid fibrils and β -solenoids

When examining the distribution of individual amino acids, it appears to be a relation between frequency and size of side chain. It is prominent that the amino acids with smaller side chains are more frequent than those with larger side chains. Glycine is for example significantly more frequent than any other amino acid within the bends of amyloid fibrils. It also happens to be the smallest of the 20 amino acids. Alanine, which has the second highest number of appearances in the bends of amyloid fibrils, is correspondingly the second smallest amino acid. Tryptophan, glutamine and cysteine did not appear at all within the chosen segment, the first two having some of the largest side chains. This suggest that the smaller side chains provide a beneficial conformation in the bends, perhaps offering flexibility. It is also possible that the larger side chains cause too much of a steric hindrance in the segment. Similar patterns were observed in β -solenoids where glycine once again was most common. The β -solenoids have a slightly more evenly distribution, but this could be due to the lack of a representative data set of β -solenoids. An interesting analysis would have been to distribute the amino acids according to the sizes of their side chains, in order to investigate the relation further.

Considering the hypothesis that the core of amyloid fibrils and β -solenoids are hydrophobic and therefore would include more hydrophobic amino acids inwards, there is no significant pattern observed. In bends of amyloid fibrils, only about half of the side chains are hydrophobic. Considering that most of the side chains are hydrophobic no conclusion can be drawn from this. When analyzing the same aspect in β -solenoids, it has to be pointed out that glycine was excluded from this analysis due to difficulties finding its direction. Since glycine is very common, it would have had a significant effect on these results by making the non-polar groups larger. With adjustment of the method it is possible to decide the directions of the glycine, which would have been an interesting complement to

the analysis.

An interesting aspect of the bends in β -solenoids, despite the lack of glycine, is the different sizes of the charged groups within the same direction. In β -solenoids they are of different sizes, as opposed to amyloid fibrils where they are of equal size. The latter is a more expected result since charged amino acids normally interact with each other, thus balancing the electrostatic interaction. The difference is larger among the amino acids facing outwards of the β -solenoids, where the acidic amino acids are only half of the amount of the basic. This indicates that there might be other electrostatic interactions balancing out the charges, in addition to the charged amino acids within the structure. Since environmental conditions such as solvent and pH is not considered, some of the charged amino acids might be interacting with its surroundings, thus having an effect on the charges. This is something to be considered if further analyzed.

An improvement of this analysis would be an adjustment of the definition of the bend. As mentioned in Section 4.1, the bend is defined as the segment connecting two β -strands. This rough definition works better for some structures than others. There are examples of structures included in the data sets which have segments connecting two β -strands that differ from others. Examples of such structures are 2MXU and 2M4J, shown in Figure 6.1. The amino acids assigned as part of the bends in 2MXU are shown to the left in the figure. This structure has a β -strand which continues into the bend and according to the previously mentioned definition, the amino acids in that segment are excluded even though they may be relevant to study when examining patterns characteristic to the bends. The figure to the right shows the amino acids present in the bend of 2M4J, which displays another issue with the definition. There are two parts of this structure that according to the definition would be classified as bends. The lower middle part which is the segment between two β -strands that does not display any amino acids, is according to the definition a bend, but is excluded in the analysis since it does not bend the structure in the same way that the other bend segment which displays the amino acids does. These are examples where, in the case of 2MXU, information may be lost due to the definition, or in the case of 2M4J, too much information is included unless manually corrected. The bends viewed in Figure 6.1 can be compared to a more easily defined bend in Figure 4.1.

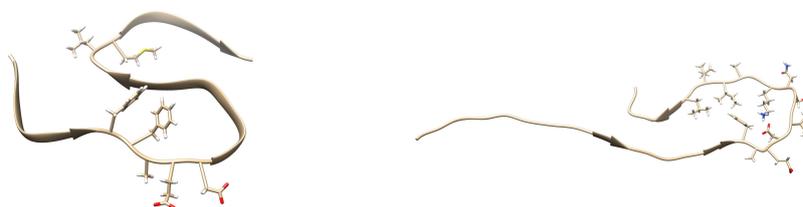


Figure 6.1: The figure exemplifies two kinds of problems related to the definition of the bend. The figure to the left shows that the middle β -strand is present in what otherwise would be considered a bend in PDB-entry 2MXU, forcing an exclusion of potentially important amino acids. The structure to the right displays PDB-entry 2M4J, where the lower middle segment between the two β -strands with no visible amino acids qualifies as a bend even though it is straight. The amino acids visible are all classified as belonging to a bend segment.

6.1.2 Charged amino acids in amyloid fibrils and β -solenoids

There is an increase in the proportion of residues facing outwards in the analysis of the amyloid fibrils, due to amino acids getting an outward direction assigned to them, when not clearly appearing in the core of the amyloid fibril. As a result, two amino acids next to each other, facing different directions, can still both be regarded as facing the exterior. An example of this is shown in Figure 6.2. The amino acids highlighted in blue are all considered to face outwards, although differing by 180 degrees. The issue of assigning a direction is also prominent when charged amino acids occur within the bends and at the end of the sequence of the amyloid fibrils, permitting a less strict arrangement compared to the residues stacking in the β -sheets. This disordered arrangement increases the number of amino acids positioning parallel to the fibril axis, only differing with a few degrees. Combined with the rough definition of the fibril core, assigning the direction of the amino acids gets subjective. Since the outward direction is favoured when the amino acid is not clearly facing the core, there is an increase in the number of residues facing outwards. An example of charged amino acids occurring in the bend and at the end of the fibril, and therefore having a disordered arrangement, is seen in Figure 6.2. Figure 6.2 also show an example of an amino acid with the side chain positioned neither inwards nor outwards.

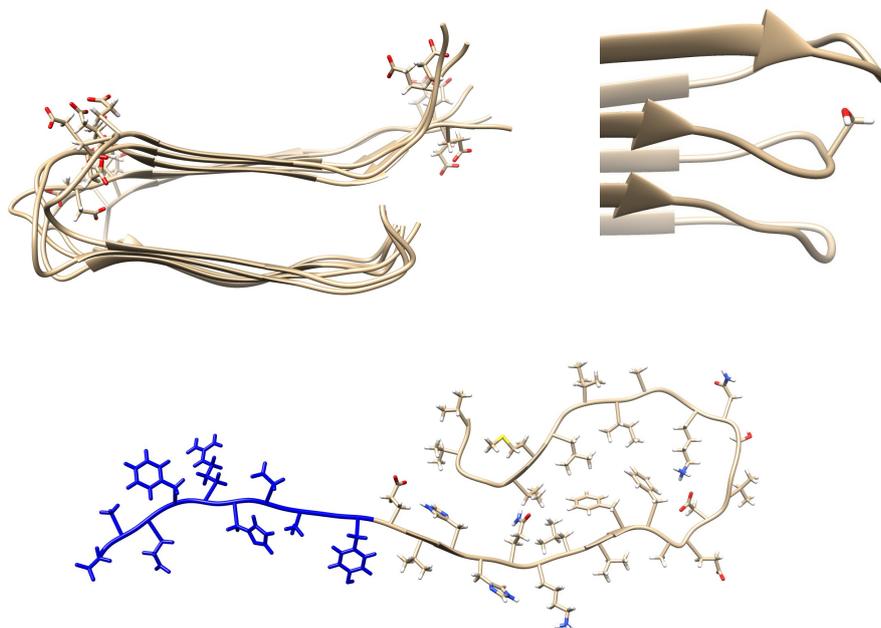


Figure 6.2: At the top left, a single filament of 2LMQ has been isolated. The figure shows the glutamic and aspartic acids, they appear in the bend and at the ends, and are showing an disordered arrangement. A single amino acid occurring in the bend has been isolated in 2LNQ, shown in the top right figure. It illustrates the problem of amino acid residues being positioned neither inwards nor outwards. The lower figure displays a single chain in 2M4J. The amino acids highlighted in blue are all regarded as facing the exterior, although they are clearly facing different directions of the chain.

As mentioned in Section 2.2, amyloid fibrils contain hydrophobic interfaces, resulting in a low dielectric environment. An example of such an interface is seen in the top of Figure 5.12 and in Section 4.1, to the right in Figure 4.2. This property lowers the possibility of a single charged amino acid occurring in the hydrophobic areas, due to it being an unstable conformation. In the analysis of amyloid fibrils there has not been any observations contradicting this information. The only observation of a single charge within the core is seen in the example of 2N0A, Figure 5.13. The amino acids next to the charged E60 is, as previously stated, polar. Polar amino acids have the possibility to form dipoles, a property that can help to balance the charge on the glutamic acid by creating an environment more similar to the aqueous, with a higher dielectric constant. It is also likely that the core, at which E60 occurs, does not form a dry interface, increasing the possibility of contact between the solvent and the residues in the core.

In the amyloid fibrils, the amino acid residues tend to stack upon each other, this is also the case for the charged amino acids. The distance between uniformly charged residues in a stack has not been investigated, but it could possibly be of interest to take the electrostatic repulsions into account. If electrostatic repulsions influence the stacks, a possible orientation of the charged amino acids would be in an alternating manner, see Figure 6.3, but such an orientation might be prevented due to the crowded space and tight interfaces of amino acid residues in the amyloid fibrils. Another aspect which has been disregarded in the analysis of the charged amino acids, are the possibility of the residues being neither protonated or deprotonated, causing them to be neutral. If the charged amino acids would be neutral, the uniform stacking of the residues gives a good opportunity to form hydrogen bonds. In further analyses it could be of interest to investigate whether the charged amino acids located inwards are neutral or not, along with the possibility of repulsive forces and hydrogen bonding.

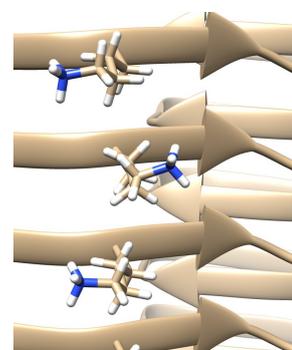


Figure 6.3: The figure displays three residues in a stack, which are oriented in an alternating manner. The stack displayed is K16 in 2MXU.

The analysis of the charged amino acids in β -solenoids was less extensive than the analysis of amyloid fibrils. However, the results show that the vast majority of the charged amino acids face outwards. This indicates that the β -solenoids might have a hydrophobic core, which is consistent with the theory mentioned in Section 2.3. However, before drawing any conclusions, the directions of the polar amino acid side chains would have to be examined, since their side chains would give more information about the internal environment.

6.2 Angles and tilt of aromatic rings

The presented result from the analysis of the aromatic rings shows, as seen in Section 5.2, that the adjusted data are more evenly distributed, compared to the unadjusted. The reason for the adjustment was the problem where the normal to the aromatic rings were reversed compared to what was expected. The problem arises from the cross product operation,

where $\vec{V} \times \vec{U}$ creates a normal which is reversed in relation to the normal given by $\vec{U} \times \vec{V}$. The situation can be seen in Figure 6.4, the bottom two rings facing opposite directions, compared to the top one. One possible reason could be that researchers submitting structures to PDB do not follow the convention of indexing atoms in aromatic rings, defined by the IUPAC-IUB [33]. By not following the convention, b_3 and b_5 , seen in Figure 4.9, may be interchanged resulting in a reversed normal. However, it is more likely that the researchers do indeed follow the convention and designate the atoms as intended, since most rings face a common direction.

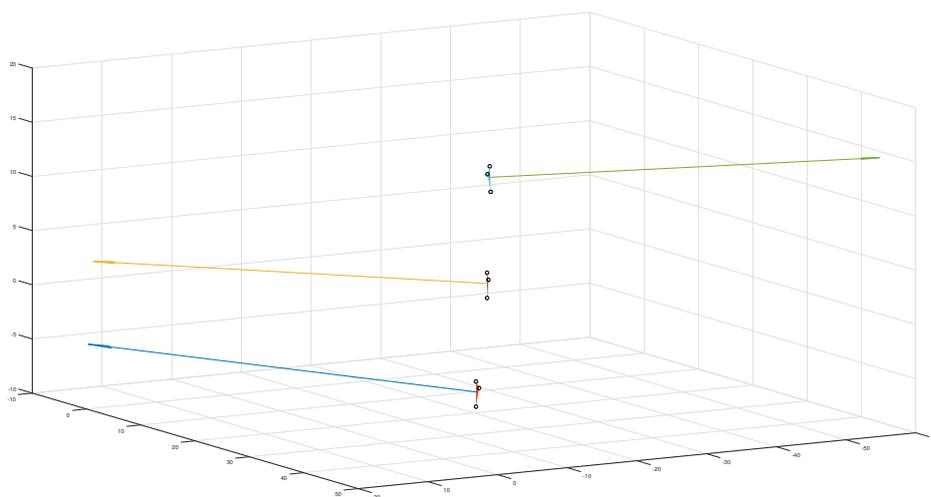


Figure 6.4: Graphical representation of the reversed angle issue with the aromatic rings shown as triangles. As can be seen, the aromatic rings are in a stack, but the directions of their normals are different. PDB-entry-2KJ3

As mentioned in Section 3.1, amyloid fibrils have varying numbers of β -strands in their β -sheets, ranging between 3 and 11. This might influence the results of the angles, since a stack of aromatic rings generally have similar angles. The combination of the two properties means that a stack of aromatic rings in the structure of 2MXU, which has 11 β -strands, influences the results more than a stack in 2M4J, which only has three. A possible improvement to the analysis of the aromatic rings would therefore be to use an average angle for each stack in the fibrils, along with the standard deviation within each stack. This might also be relevant for the analyses of direction and distribution of the bends and charged amino acids.

Regarding the alignment of the aromatic rings within a stack, the initial hypothesis was that they would align in a uniform manner when present in β -sheets. This idea is supported by the data presented for phenylalanine and histidine, but not for tryptophan and tyrosine. No further analysis has been conducted into the matter. The data can also be slightly misleading due to the restriction which determines whether two aromatic rings are defined as stacking or not. In this analysis a threshold of 5° was used. Future analysis might attempt to find a more accurate restriction which could be used as a definition as for whether or not aromatic rings are stacking.

6.3 Reflections and future development

Computational methods, such as those described in Sections 4.2 and 4.3, are efficient when performing calculations on large data sets and investigating specific aspects of a protein structure. The analyses of directions, described in Section 4.1, is on the other hand based on a molecular graphics program which allows for a more qualitative analysis. The methods have different strengths and weaknesses, which provides multiple ways of attacking the problem at hand. As experienced during the course of this project, a combination of the two is to prefer. A visualization program is a good aid in the process of understanding the complexity of protein structures, giving ideas of possible computations, as well as for debugging. Computational methods can, on the contrary, process much larger sets of data and perform mathematical analyses crucial for many investigations.

Future work on the analysis of direction and distributions of amino acid side chains would, as mentioned earlier in this chapter, be improved with more specific definitions of the bend and the directions inward and outward relative the fibril core. Instead of a manual analysis using a molecular graphics program, a computational method using mathematical boundaries could be developed to increase the efficiency and quality of the analysis.

Considering the analysis of the aromatic rings, it would also be of interest to further analyze the reason for the interval seen in the distribution of angles, Figure 5.19 and 5.20. Two possible approaches regarding further analysis have been discussed. The first involves the investigation of torsion angles influencing the aromatic rings. Due to steric hindrance, there are only certain intervals in the conformation which are possible. This approach would also be useful to include when improving the definition of the bend. The second approach would be to take into account that aromatic rings are conjugated systems. This has previously been done by Thornton and Singh [34], [35], as well as by Burley and Petsko [36], on globular proteins. The same approach is applicable on amyloid fibrils [37]. Another interesting aspect to study would be the tilt of the aromatic rings in β -solenoids. This would require further development of the methods presented in this report.

7

Conclusion

The results from the analyses presented in this report give some indications of patterns within the structure of amyloid fibrils. It was hypothesized that amino acids in bends having an inwards direction would be predominantly hydrophobic, but the results from the analysis did not show a significant over-representation of hydrophobic amino acids. On the other hand, glycine was as expected overrepresented in the bends, indicating that size of side chain is a contributing factor. Charged amino acids were more commonly facing outwards, which corresponds with the assumption that the fibril cores are more hydrophobic. The analysis of the aromatic rings shows that they tend to be positioned with an angle of about 55° relative the axis of the fibril. Aromatic rings in β -sheets, located in the same stack, are in most cases uniformly arranged, which is consistent with the expectations stated in the purpose.

When the results from the two data sets of amyloid fibrils and β -solenoids are compared, they show similarities regarding the distribution of charged amino acids and amino acids in the bends. This corresponds with the previously mentioned structural similarities between the two classes of proteins and implies that both classes could be further analyzed to gain insight into the structure of amyloid fibrils.

As mentioned in the introduction, research on protein structure is important in order to give insights that can help in generating better three-dimensional protein models. The results from the analyses presented in this report could help with future attempts to build three-dimensional models of amyloid structures.

Bibliography

- [1] J. D. Sipe and a. S. Cohen, “Review: history of the amyloid fibril.”, *Journal of structural biology*, vol. 130, no. 2-3, pp. 88–98, 2000, ISSN: 1047-8477. DOI: 10 . 1006/jsbi.2000.4221.
- [2] Amyloidosis Foundation. (2016-04-01). Facts - Amyloidosis Foundation, [Online]. Available: <http://amyloidosis.org/facts>.
- [3] D. Hall and H. Edskes, “Computational modeling of the relationship between amyloid and disease”, *Biophysical Reviews*, vol. 4, no. 3, pp. 205–22, 2012, ISSN: 18672450. DOI: 10 . 1007/s12551-012-0091-x.
- [4] C. K. Mathews *et al.*, *Biochemistry*, 4th ed. Toronto: Prentice Hall, 2013.
- [5] G. Merlini and V. Bellotti, “030807 molecular mechanisms of amyloidosis”, *Amyloid International Journal Of Experimental And Clinical Investigation*, vol. 349, no. 6, pp. 583–596, 2003, ISSN: 15334406. DOI: 10 . 1056/NEJMr023144.
- [6] L. C. Serpell, “Alzheimer’s amyloid fibrils : structure and assembly”, *Biochimica et Biophysica Acta*, vol. 1502, pp. 16–30, 2000. DOI: 10 . 1016/S0925-4439(00)00029-6.
- [7] C. P. Ferri *et al.*, “World alzheimer report 2009”, Tech. Rep., 2009, pp. 1–96.
- [8] A. Wimo *et al.*, “The worldwide economic impact of dementia 2010”, *Alzheimer’s and Dementia*, vol. 9, no. 1, pp. 1–11, 2013, ISSN: 15525260. DOI: 10 . 1016/j. jalz.2012.11.006.
- [9] J. M. Berg *et al.*, *Biochemistry*, 7th ed. Basingstoke: Palgrave Macmillan, 2012.
- [10] J. Clayden *et al.*, *Organic chemistry*, 2nd ed. Oxford: Oxford University Press, 2012.
- [11] B. Alberts, *Molecular biology of the cell*, 5th ed. New York: Taylor & Francis, 2008.
- [12] C. A. Ross and M. A. Poirier, “What is the role of protein aggregation in neurodegeneration?”, *Nat Rev Mol Cell Biol*, vol. 6, no. 11, pp. 891–898, Nov. 2005, ISSN: 1471-0072. DOI: 10 . 1038/nrm1742.
- [13] Y. Miyazaki *et al.*, “A method to rapidly create protein aggregates in living cells”, *Nat Commun*, vol. 7, May 2016. DOI: 10 . 1038/ncomms11689.
- [14] R. Tycko, “Molecular structure of amyloid fibrils: insights from solid-state NMR”, *Quarterly Reviews of Biophysics*, vol. 39, no. 1, pp. 1–55, 2006. DOI: 10 . 1017/S0033583506004173.
- [15] T. R. Jahn *et al.*, “The common architecture of cross- β amyloid”, *Journal of Molecular Biology*, vol. 395, no. 4, pp. 717–27, 2010. DOI: 10 . 1016/j. jmb.2009.09.039.

- [16] T. Eichner and S. E. Radford, “A diversity of assembly mechanisms of a generic amyloid fold”, *Molecular Cell*, vol. 43, no. 1, pp. 8–18, 2011. DOI: 10.1016/j.molcel.2011.05.012.
- [17] M. Sunde and C. C. F. Blake, “From the globular to the fibrous state - protein structure and structural conversion in amyloid formation”, *Quarterly Reviews of Biophysics*, vol. 31, no. 1, pp. 1–39, 1998.
- [18] M. Margittai and R. Langen, “Fibrils with parallel in-register structure constitute a major class of amyloid fibrils: molecular insights from electron paramagnetic resonance spectroscopy”, *Quarterly Reviews of Biophysics*, vol. 41, no. 3/4, pp. 265–297, 2008. DOI: 10.1017/S0033583508004733.
- [19] M. Sunde *et al.*, “Common core structure of amyloid fibrils by synchrotron X-ray diffraction”, *Journal of Molecular Biology*, vol. 273, no. 3, pp. 729–739, 1997, ISSN: 0022-2836. DOI: 10.1006/jmbi.1997.1348.
- [20] A. W. P. Fitzpatrick *et al.*, “Atomic structure and hierarchical assembly of a cross- β amyloid fibril”, *Proceedings of the National Academy of Sciences*, vol. 110, no. 14, pp. 5468–5473, Apr. 2013. DOI: 10.1073/pnas.1219476110.
- [21] D. J. Barlow and J. M. Thornton, “Ion-pairs in proteins”, *Journal of Molecular Biology*, vol. 168, pp. 867–885, 1983. DOI: 10.1016/S0022-2836(83)80079-5.
- [22] L. Yu *et al.*, “Crystal structures of polymorphic prion protein β 1 peptides reveal variable steric zipper conformations”, *Biochemistry*, vol. 54, pp. 3640–3648, 2015. DOI: 10.1021/acs.biochem.5b00425.
- [23] M. R. Sawaya *et al.*, “Atomic structures of amyloid cross- β spines reveal varied steric zippers”, *Nature*, vol. 447, no. 7143, pp. 453–7, 2007. DOI: 10.1038/nature05695.
- [24] R. Nelson *et al.*, “Structure of the cross- β spine of amyloid-like fibrils.”, *Nature*, vol. 435, no. 7043, pp. 773–8, Jun. 2005, ISSN: 1476-4687. DOI: 10.1038/nature03680.
- [25] J.-P. Colletier *et al.*, “Molecular basis for amyloid- β polymorphism”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 41, pp. 16938–43, 2011. DOI: 10.1073/pnas.1112600108.
- [26] A. V. Kajava *et al.*, “ β -arcades: recurring motifs in naturally occurring and disease-related amyloid fibrils.”, *The FASEB journal : Official publication of the Federation of American Societies for Experimental Biology*, vol. 24, no. 5, pp. 1311–1319, 2010, ISSN: 0892-6638. DOI: 10.1096/fj.09-145979.
- [27] J. Hennetin *et al.*, “Standard conformations of β -arches in β -solenoid proteins”, *Journal of Molecular Biology*, vol. 358, no. 4, pp. 1094–1105, 2006, ISSN: 00222836. DOI: 10.1016/j.jmb.2006.02.039.
- [28] A. V. Kajava and A. C. Steven, “ β -rolls, β -helices, and other β -solenoid proteins”, *Advances in Protein Chemistry*, vol. 73, no. 06, pp. 55–96, 2006, ISSN: 00653233. DOI: 10.1016/S0065-3233(06)73003-0.
- [29] H. Berman *et al.*, “Announcing the worldwide protein data bank”, *Nature Structural & Molecular Biology*, vol. 10, no. 12, pp. 980–980, 2003. DOI: 10.1038/nsb1203-980.

-
- [30] F. Sievers *et al.*, “Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega”, *Molecular Systems Biology*, vol. 7, no. 1, 2011, ISSN: 1744-4292. DOI: 10.1038/msb.2011.75.
- [31] E. F. Pettersen *et al.*, “UCSF Chimera — a visualization system for exploratory research and analysis”, *Journal of computational chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004. DOI: 10.1002/jcc.20084.
- [32] W. Kabsch and C. Sander, “Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features”, *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, 1983. DOI: 10.1002/bip.360221211.
- [33] IUPAC-IUB Comm. on Biochem. Nomenclature, “IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and symbols for the description of the conformation of polypeptide chains. Tentative rules (1969)”, *Biochemistry*, vol. 9, no. 18, pp. 3471–3479, 1970. DOI: 10.1021/bi00820a001.
- [34] C. A. Hunter *et al.*, “ π - π interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins”, *Journal of Molecular Biology*, vol. 218, no. 4, pp. 837–846, 1991, ISSN: 00222836. DOI: 10.1016/0022-2836(91)90271-7.
- [35] J. Singh and J. M. Thornton, “The interaction between phenylalanine rings in proteins”, *FEBS Letters*, vol. 191, no. 1, pp. 1–6, 1985, ISSN: 00145793. DOI: 10.1016/0014-5793(85)80982-0.
- [36] S. K. Burley and G. A. Petsko, “Aromatic-aromatic interaction: a mechanism of protein structure stabilization”, *Science*, vol. 229, no. 4708, pp. 23–28, 1985. DOI: 10.1126/science.3892686.
- [37] E. Gazit, “A possible role for π -stacking in the self-assembly of amyloid fibrils”, *The FASEB Journal*, vol. 16, pp. 77–83, 2002. DOI: 10.1096/fj.01-0442hyp.

A

List of the 20 amino acids and their abbreviations

Table A.1: The 20 amino acids, their abbreviations and the group they belong to according to their chemical properties.

Group	Amino acid	Three letter abbreviation	One letter abbreviation
Non-polar	Tryptophan	Trp	W
	Phenylalanine	Phe	F
	Glycine	Gly	G
	Alanine	Ala	A
	Valine	Val	V
	Isoleucine	Ile	I
	Leucine	Leu	L
	Methionine	Met	M
	Proline	Pro	P
Polar	Tyrosine	Tyr	Y
	Serine	Ser	S
	Threonine	Thr	T
	Asparagine	Asn	N
	Glutamine	Gln	Q
	Cysteine	Cys	C
Basic	Lysine	Lys	K
	Arginine	Arg	R
	Histidine	His	H
Acidic	Aspartic acid	Asp	D
	Glutamic acid	Glu	E

B

Distances between charged amino acids in amyloid fibrils

Table B.1: Distances between some of the oppositely charged amino acids in the amyloid fibrils. The distances shown are less than or close to 4 Å.

PDB-code	Basic(+) amino acid	Acidic(-) amino acid	Distance (Å)
2NNT	ARG524	GLU507	2.770
	ARG424	GLU407	2.640
	ARG324	GLU307	4.412
	ARG224	GLU207	2.660
2BEG	LYS28.B	ASP23.A	2.580
	LYS28.C	ASP23.B	2.499
	LYS28.D	ASP23.C	2.494
	LYS28.E	ASP23.D	2.498
2LNQ	LYS16.C	GLU22.B	2.639
	LYS16.D	GLU22.C	2.810
	LYS16.G	GLU22.H	3.186
2M4J	LYS28.A	ASP23.A	2.559
	LYS28.B	ASP23.B	2.563
	LYS28.C	ASP23.C	2.560
	LYS28.D	ASP23.D	2.562
	LYS28.E	ASP23.E	2.569
	LYS28.F	ASP23.F	2.565
	LYS28.G	ASP23.G	2.565
	LYS28.H	ASP23.H	2.569
	LYS28.I	ASP23.I	2.562
2LMN	LYS28.C	ASP23.A	4.221
	LYS28.C	ASP23.B	2.409
	LYS28.D	ASP23.B	3.558
	LYS28.E	ASP23.C	3.368
	LYS28.F	ASP23.D	3.286
	LYS28.I	ASP23.G	3.371
	LYS28.J	ASP23.H	3.705
	LYS28.K	ASP23.I	3.656
	LYS28.L	ASP23.J	3.167
Continued on next page			

B. Distances between charged amino acids in amyloid fibrils

Continuation of Table B.1. Distances between some of the oppositely charged amino acids in the amyloid fibrils. The distances shown are less than or close to 4 Å.

PDB-code	Basic(+) amino acid	Acidic(-) amino acid	Distance (Å)
2LMP	HIS13.C	GLU11.C	2.493
	HIS13.H	GLU11.H	2.587
	HIS13.I	GLU11.I	3.618
	HIS13.K	GLU11.K	2.958
	HIS13.M	GLU11.M	3.371
	HIS13.O	GLU11.O	3.484
2LMQ	HIS14.D	GLU11.E	2.481
	HIS13.F	GLU11.F	3.564
	HIS13.J	GLU11.J	3.994
	LYS28.J	GLU11.D	3.349
2N0A	GLU46.B	LYS80.A	2.838
	GLU46.C	LYS80.B	2.842
	GLU46.D	LYS80.C	2.841
	GLU46.E	LYS80.D	2.838
	GLU46.F	LYS80.E	2.843
	GLU46.G	LYS80.F	2.847
	GLU46.H	LYS80.G	2.848
	GLU46.I	LYS80.H	2.850
	GLU46.J	LYS80.I	2.848
2MVX	GLU11.A	HIS13.A	3.206
	GLU11.B	HIS13.B	3.218
	GLU11.C	HIS13.C	3.209
	GLU11.D	HIS13.D	3.210
	GLU11.E	HIS13.E	3.206
	GLU11.F	HIS13.F	3.206
	GLU11.G	HIS13.G	3.226
	GLU11.H	HIS13.H	3.236
	GLU11.I	HIS13.I	3.225
	GLU11.J	HIS13.J	3.200

C

Angles of aromatic rings in amyloid fibrils and steric zippers

Table C.1: The raw data used to create the charts over angular distribution for amyloid fibrils

Angle (°)	Phe	Tyr	His	Trp	Total
0-5	0	0	0	0	0
5-10	0	0	0	0	0
10-15	9	0	0	0	9
15-20	0	0	0	0	0
20-25	0	0	0	0	0
25-30	0	0	0	0	0
30-35	0	2	0	0	2
35-40	2	5	0	0	7
40-45	4	17	0	1	22
45-50	37	1	0	0	38
50-55	27	0	0	0	27
55-60	35	1	9	0	45
60-65	18	1	18	0	37
65-70	12	0	1	0	13
70-75	4	2	0	0	6
75-80	3	2	1	1	7
80-85	4	1	1	0	6
85-90	3	0	3	0	6
90-95	5	1	0	0	6
95-100	1	0	0	0	1
100-105	2	0	1	0	3
105-110	5	1	0	0	6
110-115	2	1	3	0	6
115-120	4	0	8	0	12
120-125	15	3	7	0	25
125-130	7	0	14	0	21
130-135	10	7	3	0	20
135-140	2	12	10	0	24
140-145	0	9	1	0	10
145-150	0	0	1	1	2
150-155	0	0	0	0	0
155-160	0	0	0	1	1
160-165	0	0	0	0	0
165-170	0	0	0	0	0
170-175	0	0	0	0	0
175-180	0	0	0	0	0

Table C.2: The raw data used to create the charts over angular distribution for steric zippers

Angle (°)	His	Phe	Tyr	Trp	Total
0-5	0	0	0	0	0
5-10	0	0	0	0	0
10-15	0	0	0	0	0
15-20	0	0	0	0	0
20-25	0	0	0	0	0
25-30	1	0	0	0	1
30-35	0	0	0	0	0
35-40	0	0	0	0	0
40-45	0	0	0	0	0
45-50	0	2	2	0	4
50-55	0	3	0	0	3
55-60	0	1	1	0	2
60-65	0	0	2	0	2
65-70	0	2	0	0	2
70-75	0	0	3	0	3
75-80	0	1	0	0	1
80-85	0	1	0	0	1
85-90	0	1	0	0	1
90-95	0	1	0	0	1
95-100	0	0	0	0	0
100-105	0	1	1	0	2
105-110	0	0	1	0	1
110-115	0	0	0	0	0
115-120	1	0	2	0	3
120-125	1	1	0	0	2
125-130	0	1	2	0	3
130-135	1	3	1	0	5
135-140	0	0	0	0	0
140-145	0	1	0	0	1
145-150	0	0	0	0	0
150-155	1	0	0	0	1
155-160	0	0	0	0	0
160-165	0	0	0	0	0
165-170	0	0	0	0	0
170-175	0	0	0	0	0
175-180	0	0	0	0	0

D

Adjusted angles of aromatic rings in amyloid fibrils and steric zippers

Table D.1: The adjusted raw data used to create the charts over angular distribution for amyloid fibrils

Angle (°)	Phe	Tyr	His	Trp	Total
0-5	0	0	0	0	0
5-10	0	0	0	0	0
10-15	9	0	0	0	9
15-20	0	0	0	0	0
20-25	0	0	0	1	1
25-30	0	0	0	0	0
30-35	0	2	1	1	4
35-40	2	14	1	0	17
40-45	6	29	10	1	46
45-50	47	8	3	0	58
50-55	34	0	14	0	48
55-60	50	4	16	0	70
60-65	22	1	26	0	49
65-70	14	1	4	0	19
70-75	9	3	0	0	12
75-80	5	2	2	1	10
80-85	5	1	1	0	7
85-90	8	1	3	0	12

Table D.2: The adjusted raw data used to create the charts over angular distribution for steric zippers

Angle (°)	His	Phe	Tyr	Trp	Total
0-5	0	0	0	0	0
5-10	0	0	0	0	0
10-15	0	0	0	0	0
15-20	0	0	0	0	0
20-25	0	0	0	0	0
25-30	2	0	0	0	2
30-35	0	0	0	0	0
35-40	0	1	0	0	1
40-45	0	0	0	0	0
45-50	1	5	3	0	9
50-55	0	4	2	0	6
55-60	1	2	1	0	4
60-65	1	0	4	0	5
65-70	0	2	0	0	2
70-75	0	0	4	0	4
75-80	0	2	1	0	3
80-85	0	1	0	0	1
85-90	0	2	0	0	2

E

Data from the direction and distribution analysis

Amount of amino acids in bends of amyloid fibrils				
Chemical property	Amino acid	Total amount	Amount inward	Amount outward
Non-polar	W	0	0	0
	F	47	47	0
	G	306	52	254
	A	164	76	88
	V	76	25	51
	I	18	9	9
	L	31	21	10
	M	12	12	0
	P	20	0	20
	Total	674	242	432
Polar	Y	10	0	10
	S	137	82	55
	T	16	16	0
	N	114	36	78
	Q	0	0	0
	C	0	0	0
	Total	277	134	143
Basic(+)	K	115	27	88
	R	8	0	8
	H	4	0	4
	Total	127	27	100
Acidic(-)	D	77	38	39
	E	96	0	96
	Total	173	38	135
Total:		1251	441	810

Figure E.1: Number of amino acids present in bends of amyloid fibrils sorted between specific amino acid as well as chemical groups.

Amino acids in bends of beta-solenoids					
Chemical property	Amino acid	Total amount (with G)	Total amount (without G)	Amonut inward	Amount outward
Non-polar	W	5	5	4	1
	F	11	11	8	3
	G	80	0	0	0
	A	41	41	27	14
	V	13	13	5	8
	I	5	5	2	3
	L	9	9	6	3
	M	5	5	1	4
	P	24	24	16	8
	Total	193	113	69	44
Polar	Y	12	12	3	9
	S	58	58	28	30
	T	30	30	21	9
	N	69	69	46	23
	Q	20	20	11	9
	C	9	9	2	7
	Total	198	198	111	87
Basic(+)	K	21	21	11	10
	R	15	15	9	6
	H	13	13	10	3
	Total	49	49	30	19
Acidic(-)	D	41	41	33	8
	E	13	13	11	2
	Total	54	54	44	10
Total:		494	414	254	160

Figure E.2: Number of amino acids present in bends of β -solenoids sorted between specific amino acid as well as chemical group. Note that the columns "Amount inward" and "Amount outward" are based on the column "Total amount (without G)".

E. Data from the direction and distribution analysis

Amount of charged amino acids in amyloid fibrils				
Chemical properties	Amino acid	Total amount	Amount inwards	Amount outwards
Basic(+)	K	491	29	462
	R	71	4	67
	H	247	30	217
	Total	809	63	746
Acidic(-)	D	226	59	167
	E	443	34	409
	Total	669	93	576
Total:		1478	156	1322

Figure E.3: Number of charged amino acids present in amyloid fibrils.

Amount of charged amino acids in beta-solenoids				
Chemical properties	Amino acid	Total amount	Amount inwards	Amount outwards
Basic(+)	K	50	4	46
	R	46	3	43
	H	22	4	18
	Total	118	11	107
Acidic(-)	D	72	3	69
	E	40	3	37
	Total	112	6	106
Total:		230	17	213

Figure E.4: Number of charged amino acids present in β -solenoids.

F

Number of stacks present in a subset of the amyloid fibrils

Table F.1: Number of stacks for each amino acid with an aromatic ring in its side chain and which PDB-entry they are located in.

PDB-entry	His	Phe	Tyr	Trp
2KJ3	1	1	1	1
2NNT	1	3	2	2
2M4J	9	9	3	-
2M5N	4	-	-	-
2LMP	6	6	3	-
2LMQ	6	6	3	-
2LNQ	3	-	-	-
2KIB	4	-	-	-
2MVX	6	6	2	-
2MPZ	6	-	-	-
2MXU	2	2	-	-